SCUOLA
NORMALE
SUPERIORE

Classe di Lettere e Filosofia
Tesi di Perfezionamento

Inter-gestural Coordination in Temporal and Spatial Domains in Italian: Synchronous EPG + UTI Data

Candidato: Cheng Chen

Relatore: Prof. Pier Marco Bertinetto
Co-relatrice: Dott.ssa Chiara Celata

Table of Contents

Chapter I Introduction

This dissertation explores the temporal coordination of articulatory gestures in various segmental conditions in Italian, by comparing onset and coda singletons as well as word-final and intervocalic consonant clusters in a Tuscan variety of Italian. Articulatory models of syllable structure assume that the coordination between the vocalic gesture and the consonantal gesture may differ in onset vs. coda and in singletons vs. clusters. Based on previous literature on different languages, we expect to find differences in the temporal coordination of singletons and clusters in Italian too. In addition, recent literature suggests that the articulatory and coarticulatory properties of the segments play an important role in determining the details of the coordination patterns, and that not all segments or segmental sequences behave in the same way as far as their gestural coordination relations are concerned. Thus, an additional aim of this work is to compare consonants with different coarticulatory properties (in the sense of modifications of C articulation in varying vocalic contexts) and seek for possible relations between coarticulation and coordination patterns.

The methodology used is new. We used an original system for the acquisition, real-time synchronization and analysis of acoustic, electropalatographic (EPG) and ultrasound tongue imaging (UTI) data, called *SynchroLing*. EPG and UTI instrumental techniques provide complementary information on, respectively, linguo-palatal contact patterns in the anterior vocal tract and midsagittal profiles of the whole tongue, including postdorsum and root. *SynchroLing* allows real-time inspection of contacts in the artificial palate and tongue midsagittal movements, coupled with acoustics. A preceding version of the multi-level system has been used for the analysis of rhotic variation in Tuscan (Spreafico, Celata, Vietti, Bertini and Ricci, 2015; Celata, Vietti and Spreafico, 2017). One of the goals of this paper is that of testing the validity of this new instrumental environment for the analysis of coordination among segments and coarticulation.

In articulatory models of syllable structure, syllable structures are viewed as the results of different coordinative structures between/among the articulatory gestures (Browman and Goldstein, 1988, 1992; Ladd, 2014). In particular, syllable-onset positions are associated with tighter articulatory constrictions and greater stability than syllable-coda positions (since Krakow, 1989, 1999). Additionally, onset clusters are said to be globally aligned to the following nuclear vowel along their temporal midpoint in such a way that an increase in onset complexity results in an increase of CV overlap ("c-centre effect"); by contrast, coda clusters are sequentially organised (since Browman, Goldstein, 2000; De Jong, 2003; Hermes, Mücke and Grice, 2013).

However, it has been recently shown that the articulatory and coarticulatory properties of the segments play an important role in determining the details of the coordination patterns. For instance, onset-vowel timing interacts with coarticulatory resistance, since the articulatory composition of the cluster predicts the degree of overlap between the final consonant of an onset cluster and the following vowel (*e.g.,* Pastätter and Pouplier, 2015 on Polish). Similarly, not all coda clusters are sequentially organized with respect to the preceding vocalic nucleus, since V shortening and an increased overlap between the vocalic nucleus and the following consonant have been observed for some complex coda clusters (*e.g.,* Marin, Pouplier, 2010 for American English laterals; Marin, Pouplier, 2014 for Romanian rhotics). This evidence suggests that the articulatory properties of segments (and primarily their coarticulatory resistance, *e.g.,* Recasens, Espinosa, 2009) influence the timing of syllables and motivates a closer investigation of language-specific and consonant-specific coordination patterns.

This dissertation is structured in the following way.

Chapter 2 is devoted to the theoretical background. A review of the putative similarities existing between the speech system and other physical/physiological systems is provided. The characteristics of a self-organised physical system are then outlined, with particular emphasis on the macroscopic *vs.* microscopic phenomena of variation. The basic tenets of Articulatory Phonology that are useful to our investigation are also shortly presented.

Chapter 3 and Chapter 5 illustrate, from an articulatory perspective, what the macroscopic and microscopic properties of the human speech system consist of. The canonical coordination patterns found in human speech – similar to those holding for limb movements – are intended as the macroscopic property of the system. The variations in the inter-gestural coupling strength that depends on the articulatory properties of the constriction gestures are meant as the microscopic properties of the system.

Chapter 4 gives an illustration on how the traditional tune-text association in the autosegmental-metrical framework is understood as the intra-level inter-gestural coordination between the articulatory gestures at the segmental level and at the suprasegmental level.

Chapter 5 also introduces the approach of degree of articulatory constraints (known as DAC model), which is able to explain how the inter-gestural timing can be influenced in the spatial dimension by the different articulatory properties of individual constriction gestures.

Chapter 6 contains the information relative to the empirical study, including the speech material, the experimental setting based on the *SynchroLing* system, the data acquisition and analysis procedures and how the necessary articulatory landmarks are defined and established on the basis of EPG and UTI evidence.

The presentation of the results of the experiment is divided into two parts. The coordinative patterns (*i.e.*, the temporal aspects of inter-gestural coordination) are presented in Chapter 7. The coarticulatory phenomena (*i.e.*, the spatial aspects of inter-gestural coordination) are reported in Chapter 8.

The discussion of the findings, alongside some implications for future research, is contained in the concluding Chapter 9.

Chapter II Modelling the Speech System through Detail: from Strict Dichotomy to Articulatory Phonology

The empirical and traditional study of human speech has been oriented in two different domains: one regards speech as the physical activities changing continuously in time and the other concerns its underlying linguistic (cognitive) structure that could be further decomposed into a sequence of basic units. By doing this, the physical aspect of speech is parallel to its abstract and structural properties. From this perspective, the phonological-phonetic mapping rules are always required in order to get the complete picture of speech system.

The framework of Articulatory Phonology, found in the mid-eighties of last century, has led to a considerable advance in our understanding of speech by assuming that the two different domains are in fact the physical (high) and the linguistic/cognitive (low) dimensional descriptions of a single system. Like any other physical system, the speech system is intrinsically endowed with both (categorically-defined) macroscopic and (contextually-varied) microscopic properties.

In this chapter, we will provide an overview of the development of the speech models from treating phonological units and phonetic units as the units in two distinct domains to viewing them as a single unit – dynamic articulatory gesture. Besides, we will also examine the nature of the (self-organised) physical system and how macroscopic and microscopic properties shape the speech system (the section 2.1). Finally, the basic principles of the Articulatory Phonology are reviewed (the section 2.2).

2.1 Phonological structure and physical systems

2.1.1 Phonology and phonetics

Linguistic theories in the last century developed various approaches or models that favour the idea of the existence of discrete cognitive units which are assumed as the fundamentals in the structure of natural languages. Under such approaches, the words in human languages are not merely the combination of a set of sounds without an internal structure. Rather, they are composed of elements belonging to an inventory of limited elements which are denominated discrete phonological primitives. These discrete units are not intrinsically meaningful *per se*, however they are able to combine with one another in various ways to form the lexical elements of a certain language.

Following this idea, phonology is then seen as the discipline which describes spoken languages as a combinatorial system typically composed of 30 – 40 basic units (Goldstein and Fowler, 2003; Pouplier, 2011). The smallest inventory known makes use of about 10 contrastive units, the largest of up to 150 (Maddieson, 1984). In such a system, the atomic, minimal units are claimed to be distinct in a discrete way from one another. Moreover, these units are supposed to maintain their discreteness even if they combine together to form larger units. Thus, phonological theories have two basic functions: in the first place, they are able to capture the nature of these primitives which serve as the atoms of speech production and perception, and consequently, they describe the rules through which these primitives are organised to combine themselves into lager structures according to language-specific restrictions.

The postulated units (or entities) are stored at the cognitive level of the language users, where the sound-to-meaning association takes place. Thus, phonological representation is abstract, coarse-grained or discrete, cognitive and context-independent. However, the combinatorial property of the phonological system appears to be at odds with the physical observation of speech because only non-discrete (continuous), contextually dependent changes can actually be obtained from speech analysis. The discreteness of the phonological units contrast, for instance, with the observation that in the Italian word [pasta] 'noodles, or the products made of the flour', the first [a] and the second [a] are clearly two different [a]s. From the physical point of view, the phonetic representation of a word is concrete, continuous, physical and context-dependent, that is, specified at the level of fine-grained physical distinctions.

The relation between phonology and phonetics still is the hot topic in the study of linguistic systems (see *e.g.,* Ohala, 1997, 1999; Scobbie, 2007; Romero and Riera, 2015 for the most recent developments). Articulatory Phonology was elaborated within a theoretical milieu that emphasised the need of integrating the methodologies of experimental phonetics and the research questions of phonology. As argued by Pierrehumbert (1990), even though the gap between the linguistic and physical structure of speech has always been difficult for phonological theory to bridge (Browman and Goldstein, 1986), a theory encompassing phonology, phonetics and their relation to each other is needed as a foundation for a theory of language processing and language acquisition.

2.1.2 Is physical system gradient?

It has been shown in the previous section that the discrepancy between phonology and phonetics arises from the contrast between the abstract phonological units and observable phonetic entities, that is, between the cognitive representation and the physical reality. The cognitive representation is claimed to be based on qualitative and categorical distinctions, whereas the physical reality is gradient and time-conditioned. According to this view, the cognitive and the physical domain are necessarily different. As a result, any theory of speech production and perception that tries to integrate these two dimensions must entail a series of mapping rules, *i.e.,* how these two different dimensions are linked (Pierrehumbert 1990).

However, early studies on coordination and movement, especially of fingers and legs (*e.g.,* Kelso, 1984; Kelso and Tuller, 1984; Turvey, 1990), have challenged the dichotomy between cognitive and physical systems. As shown below, the physical world is both gradient and categorical (qualitative). In order to achieve coordinated movements, the interaction between cognitive/neural activities and the principles of physical self-organisation is very important. A classic example comes from the investigation by Kelso (1984) on the oscillation of two index fingers. Two phasing modes are possible for the finger movements - the symmetric phasing mode and the antisymmetric phasing mode. The former refers to the simultaneous oscillation of two index fingers, whereas the latter refers to the sequential movements of two fingers: only when the movement of the first finger reaches its maximum displacement with respect to the departure point, the second one begins its oscillation. An abrupt phase transition to the symmetric mode is found at a certain time point when the frequency of the antisymmetric phasing movements is increased (but never the reversed transition)[1]. Then, Haken, Kelso and Bunz (1985, discussed in Turvey, 1990) suggested that physical stability can account for such phase-transition phenomenon observed in human finger movements. They used potential function to predict the stability of a particular phase relations that reflects the frequency of movement, and the resulting abrupt phase change is due to the increased instability of the antisymmetrically-coordinated movements as the frequency of the movement increases (Haken *et al.*, 1985).

Since two index fingers belong to a single human body and the human body can be seen as a physical system, they are somehow mechanically (at least neurally) connected and coupled to one another. It is therefore not surprising that the fingers' movement is subjected to physical stability.

The coupling movements of two different subjects found by Schmidt, Carello and Turvey (1990) seem to additionally suggest that there might be a higher level of information which the physical system is subjected to. The authors demonstrated that when two subjects speed up the swinging of their legs, at a critical point, the antisymmetric coupling mode is suddenly shifted to the symmetric one. Such interaction of limb movements of two individuals is, according to the authors, informational in the

---

[1] The details on the phase transition between two alternative coordination modes will be discussed in the following chapter.

sense that it is the optical information about the leg of another subject that links two sub-physical systems. Even more importantly, there is no mechanical linkage between the two individuals whose swing of legs form a larger physical system (Schmidt *et al.,* 1990).[2]

In the linguistic domain, Browman and Goldstein (1990b) further interpreted this kind of information as 'cognitive' because it has to do with the visual information processing. They subsequently provided a clarification for those who might doubt whether the meaning of the term 'cognitive' was the same as the one by which we usually refer to phonological units. According to the authors, the term 'cognitive' as applied to phonological units refers to a kind of introspectively available cognition that is not specifically perceptual or motoric. And the term 'cognitive' they employ for the informational linkage (in broad sense, the visual information processing is traditionally considered) is not substantively different from the introspectively available cognition (the phonological use of 'cognitive', see above), since both of them could be intentional (Browman and Goldstein, 1990b). In such a physical system that is not tightly assembled, two people can watch the swing of legs of one another without the intention to synchronise their motions. However, without the intentionality or the mechanical linkage (certain physical connection between the two individuals, for example), the complex system can never be actually assembled from pieces. Thus, the informational linkage constitutes a boundary condition under which a self-organisation physical system takes place (Schmidt *et al.,* 1990; Turvey, 1990).

The principle of physical bifurcation (*i.e.,* the one applying to the phase transition phenomenon reviewed above for limb oscillation) is a common property of complex systems. Being a physical system, speech is expected to exhibit the same properties - involving simultaneously gradient behaviours which vary over time and qualitatively distinct forms. During speech production, the articulators are continuously in motion to create constrictions within the vocal tract. Since it is easy to model these spatial variations of the articulators as a function of time, thus, such continuous motion can be described in a gradient fashion. However, the motion of the articulators over time can

---

[2] Actually, in the case in which the visual information about other person's leg is eliminated, neither phase locking nor, relatedly, the bifurcation from the alternate to symmetric mode can be observed (Schmidt, Carello and Turvey 1990: 246).

also be characterised in terms of invariant dynamical specifications, for example, a series of continuous motion constitutes a constriction in a certain place within the vocal tract, whereas another series of motion constitutes the constriction in a different place within in the vocal tract, and these two constrictions are qualitatively distinct (Fowler *et al.*, 1980; Browman and Goldstein, 1986, 1989, 1990a, 1990b, 1992, 1993). These observations suggest that although the speech system, like all other complex physical systems, exhibits the gradient properties characterised by time-varying motion of articulators, it is still a system that can be modelled in a discrete way, in the sense that the dynamical specifications underlying the motion of the articulators do not vary over time, *i.e.*, they are invariant. These temporally invariant specifications for a given series of continuous motion constitute a temporal interval which in turn contributes to a discrete and qualitatively distinct description to the continuous motion[3]. In terms of a phonology that is articulatorily defined, such invariant temporal interval refers to the *stiffness* of a given articulatory movement, which in turn provides the practical ground for the dynamical description regarding of intrinsic activation of the articulatory gesture in question (see also the section 2.2 for a further discussion).

2.1.3 Macroscopic and microscopic properties of human speech

Since the cognitive-physical and the qualitative-gradient dichotomies are unwarranted, the traditional two-way representational approach - phonological and phonetic - appears to be no longer satisfactory for a fully-fledged description of the speech system. Browman and Goldstein (1990b) proposed that the gradient and qualitative behaviours observed in the physical systems described in the previous section could be interpreted as the macroscopic and microscopic properties of a single complex system. The coarse-grained properties employed for the differentiation of the contrastive units in the speech

---

[3] A similar approach can be found in Quantal Theory of speech put forward by Stevens (1972, 1989) from the acoustic and auditory point of view. The theory holds the idea that the acoustic and auditory properties of the constrictions within a discrete region of the vocal tract is rather stable, although the distribution of the constrictions within the vocal tract can be described in a gradient fashion. For example, a vowel like /i/ in English has a rather stable palatal constriction from the acoustic point of view, and a little forward or backward shift of the constriction location (the articulatory perturbation) will not result in the perception of a categorically different sound. These discrete regions are qualitatively distinct from one another and they are further served as the basis for the phonological contrasts. Thus, the Quantal Theory provides a ground for the speech system as simultaneously gradient and qualitative.

system are the properties at the macroscopic level (also called low dimensional[4], and corresponding to what is traditionally known as phonological), whereas the microscopic properties (also called high dimensional, traditionally known as phonetic) refer to those fined-grained acoustic and articulatory descriptions.

The interaction between macroscopic and microscopic properties is one of the most important characteristics of complex physical system (Kugler and Turvey, 1987). According to Kugler and Turvey (1987), the nature of microscopic behaviours influences the stability of the macroscopic organisation of the entire system, whereas the macroscopic structure feeds the specific microscopic behaviours back. This mutual intraction between macro- and microscopic perspectives implies that a physical system is self-organised.

How does this reciprocity work in the phonological structure? The application of the macro-microscopic perspectives to the phonological structure means that, on the one hand, the macroscopic properties of phonological structure such as contrastive, combinatorial and categorical properties emerge from the microscopic properties (e.g., the detailed articulatory and acoustic behaviours) of an individual phonological unit along the span of a particular temporal interval, and on the other hand, the macroscopic properties feed the microscopic domain back and constrain the details of the microscopic units (Browman and Goldstein, 1990b, 1993). Such interaction appropriately reflects the Pierrehumbert's outlook that a theory encompassing both dimensions as well as their relation to each other is necessary to a full understanding of speech phenomena (Pierrehumbert, 1990). According to this view, the speech system is a self-organised physical system, which additionally overcomes the need for mapping rules between phonetics and phonology, indispensable in traditional phonological theories (see above the section 2.1.1).

In the laboratory phonology literature, much evidence has been found to support both microscopic-oriented constraints and macroscopic-oriented details. For example, observed cross-linguistic differences in the articulatory properties of a given

---

[4] Low dimensionality is a definitional property of a combinatoric system. Browman and Goldstein call the macroscopic forms low dimension because the number of atoms is always small with respect to the number of combinations (Browman and Goldstein, 2002).

phonological unit are hypothesized to be regulated by the macroscopic property of the contrasts present in the phonological system involved (Fant, 1960; Stevens, 1972, 1989; Wood, 1979, 1982; Goldstein, 1983; Pierrehumbert, 1990). Fant (1960) showed that the formant frequencies of a horn-shaped vocal tract are a function of constriction location, constriction size (i.e., narrowness or constriction degree) and lip aperture. Stevens (1972) found that for the high front unrounded vowel /i/, the subtle modification in the constriction location does not affect the acoustic resonance and thus the perception remains unchanged, while the perturbation of the narrowness could result in substantively different acoustic output. A recent ultrasound study on the articulation of vowel height in English front vowels by Whalen, Noiray, and Bolaños (2010) confirmed the observation that constriction degree does matter in the articulation of high front vowel. In a similar vein, Wood (1979, 1982) showed that the degree of frontedness of /i/ varies from language to language. In languages in which /i/ and /y/ contrast, a constriction location in the pre-palatal region is preferred, whereas a mid-palatal constriction is found when the roundness contrast is absent. These findings can be taken as demonstrating that the macroscopic property of contrast is constraining the detailed articulatory properties, i.e., the microscopic properties, of the individual units.

The macro-micro interaction in the human speech system can also account for contextual variations of phonetic units - the phenomenon usually referred to as coarticulation. The output of phonological/macroscopic forms are represented in space and time by continuous phonetic/microscopic implementation in both articulatory and acoustic domains. Keating (1988) modelled the possible range of contextual variations of a given phonetic unit using the metaphor of a ‘window’. Windows have their own duration and a width representing all the possible physical values that a target can be characterised, *i.e.*, the range of variability within a target. The width of the window depends primarily on whether the relevant features are associated with the output phonetic unit of the phonological component, that is, if the features are precisely specified then the width of the associated window will be narrow and in consequence, little amount of variations are allowed; if the features are left unspecified, then a larger amount of contextual variations is allowed by a relatively wide window. The exact width of a window can be derived for each speech system from information on the maximum amount of contextual variability observed in speech: all intermediate degrees between maximally narrow and maximally wide windows are then possible. By allowing windows to vary continuously in their width, the model can represent the phonologically unspecified segments that are associated with articulatory targets, which might further contribute to the different degrees of coarticulation between the phonetic segments. Therefore, the width of a window, *i.e.*, the macroscopic property, is a language-specific issue and it is constrained by the contextual variations (*i.e.*, microscopic information) on the maximum amount of possible contextual variations observed in the real speech data for each language (Keating, 1988; Farnetani and Recasens, 2006). Similarly, Manuel and Krakow (1984) defined the macroscopic property as 'target area' in the acoustic space. They further found that the size of target area for vowels is a function of the number of contrastive vowels in the phonological system (the more vowels has a system, the smaller is the size of the target area for each vowel in such system). This observation, again, suggests the existence of a relationship of mutual constraint between micro- and macroscopic properties in the speech system. The DAC (Degree of Articulatory Constraint) model developed by Recasens (1996, 1999, 2006, 2014), which will be discussed in chapter 4, also derives from a similar need of postulating the existence of reciprocal interactions between these two domains.

Scholars have pointed out that reciprocal constraints might be responsible for historical sound change as well. Labov, Yaeger and Steiner (1972) identified a series of chain shifts in the sound change of English vowels. Three main tendencies in such vowel shifts were found: 1) front or back tense vowels (peripheral) tend to raise, 2) front or back lax vowels (less peripheral) tend to lower and, 3) back vowels tend to be fronted. Briefly speaking, the main tendency regulating these shifts lies in the modification of vowel quality in the vertical dimension. Goldstein (1983) examined the articulatory-acoustic relations in English 'chain shifts' using the simulation of articulatory variabilities. He showed that acoustic variations resulting from 360° random articulatory perturbations of the articulatory location of the target vowel can be represented along the same dimensions of English chain shift. This ground-breaking finding led the author to suggest that instead of thinking of vowel shift (and sound change in general) as the product of either articulatory or acoustic modifications, we should consider vowel shift "as exploiting variability patterns that emerge from the characteristics of the vocal tract resonator" (Goldstein, 1983: 272). In other words, the microscopic properties of vowels can give rise to specific macroscopic tracks of sound changes (Goldstein, 1983; Browman and Goldstein, 1990b).

The microscopic properties of speech actually allow a large number of 'mechanically' available forms to emerge from a small, restricted set of primitive forms. Hence, microscopic properties provide the macroscopic forms with a potentially higher number of degrees of freedom. The multiplicity of degrees of freedom, in turn, contributes to the diversity of phonologies across languages (Kelso, Saltzman and Tuller, 1986; Browman and Goldstein, 1993).

2.2 Articulatory Phonology

2.2.1 Articulatory gestures as phonological units

The theoretical proposal which best elaborated the ideas and suggestions highlighted above is Articulatory Phonology. Articulatory Phonology (henceforth, AP) differs from previous feature-based or segment-based approaches in assuming that articulatory movements can serve as the basis of phonological representations. Speech is said to be structured in terms of articulatory gestures, while the segment itself is only epiphenomenal (Ladd, 2014). Consistently with the view of the mutual relationship between micro and macroscopic properties of phonological systems, the articulatory gestures are considered the basic phonological units, inasmuch as describing speech in terms of articulatory gestures can capture both physical and cognitive properties (Browman and Goldstein, 1986, 1989, 1990, 1992).

Articulatory gestures, as described by Browman and Goldstein (1992), are characterisations of discrete, physically real events that unfold during the speech production process. Articulatory Phonology attempts to describe lexical items in terms of these discrete physical events and their interrelations. Such an approach implies that gestures are not only articulatory actions, but also primitive units of contrast among lexical items. In this sense, these physical events have a simultaneous phonological meaning. Thus, articulatory gestures are characterised as both units of action (concrete and continuous) and units of information (abstract and discrete).

Gestures are units of action in the sense that they could serve as bricks in building the architecture of coordinative structure. Each articulatory gesture represents a dynamic neuromotor system that guides the spatial-temporal coordination of a series of independent articulators and muscles involved in the action of the formation of a specific constriction in the vocal tract. It can be captured by observing the movements of articulators during the speech. A gesture of labial closure, for instance, is an action which requires a coordinative structure are composed of three different articulators: upper lip, lower lip and mandible, together with some fifteen distinct muscles. Once the motor command for labial closure is issued, the action is immediately activated in the vocal tract, all the articulators and muscles potentially involved in the formation of such constriction are recalled in the way that they cooperate among themselves to fulfil the given 'task'[5] – the lower lip approaching the upper lip and closing the lips.

Gestures are units of information in the sense that individual gestures may be used to distinguish, or contrast, utterance or lexical items from one another, by the presence vs. absence of a particular gesture, or by the properties of a constriction action in the vocal tract in relation to its location and degree (see also the following section). In simple terms, even though articulatory gestures can be ascribed to a dynamic system that guides the multi-articulator and muscle coordination, they are discrete units which are able to characterise the abstract phonological contrasts at the cognitive level (low-dimensional/macroscopic properties), nevertheless they are defined intrinsically by their

---

5 In the self-organised systems like speech production, the term 'task' refers to the actions with a precisely defined goal. There are lots of ways to get to Rome, for example, one can take train, or take a flight, or even arrive on foot, depending on the conditions (distance, safety, the moods and so on) in the real world. While arriving in Rome is the goal of the task, the action that one arrives in Rome is a task. The task and the goal are invariant, since they are not subjected to the conditions in the physical world. Thus, the task and the goal are contextually independent and somewhat abstract. As to the specific ways that one employs to achieve the goal, they enjoy high degrees of freedom, therefore they are contextually dependent.

physical properties (high-dimensional/microscopic properties). Such a view of the relation between phonological structure and linguistic production does not postulate anymore the existence of mapping functions between phonological representation and phonetic implementation because gestures make them unnecessary. The unit of information (phonological entity) and the unit of action (phonetic entity) constitute one and the same linguistic unit – the articulatory gesture –, which includes all necessary information for its execution.

Supporting evidence for the assumption that articulatory gestures, as units of action, are simultaneously discrete units of information came from studies on speech errors. Speech errors were usually investigated within the generative frameworks as anticipation and/or the exchange of particular distinctive features. Before AP, speech errors were typically modelled in terms of a two-stage process composed of planning (when the feature substitution takes place) and execution (when the new string is produced, according to an 'interpolation' function between the new features; Shattuck-Hufnagel, 1983; Levelt, 1989).

An articulatory study using EMG[6] carried by Pouplier, Chen, Goldstein and Byrd (1999) suggested a totally different view. The execution mechanism was suggested to be the source of speech errors. For instance, in a phrase such as “cop top” (/kɒp tɒp/), the /k/ of the first word involves a dorsum gesture with a post-palatal constriction location, and the onset of the second word /t/ contains a tongue tip gesture with the constriction in the dento-alveolar region. If the sequence is repeated over and over by a speaker, an error takes place. EMG analysis showed that, in some repetitions, the tongue dorsum closure for /k/ occurs simultaneously with the tongue tip closure for the gesture of /t/. Sometimes this gestural modification is inaudible; other times the magnitude of dorsum raising is increased so that a dorsal stop rather than a coronal stop is perceived (Pouplier *et al.*, 1999; Goldsetin, Pouplier, Chen, Saltzman and Byrd, 2006).

The two-stage early model opposing planning and execution thus fails to interpret such ‘erroneous’ production for the reason that English phonotactics does not allow the

---

[6] Electromyography (EMG) is a technique that measures the electrical activity (EA) of several muscles originally in order to diagnose movement disorders. It is further extended to the investigation on speech production, in particular, tracking the articulatory trajectories during the articulation.

simultaneous production of gestures for /k/ and /t/; the output would be ill-formed and could not be generated at the stage of planning. Nonetheless, if the execution mechanism involves the dynamic coupling relations between units of action (much as in the example of human limb motion reviewed above), this kind of error could be accounted for at the stage of execution in terms of the mismatch of the gestural phasing relations. The phenomena involved in speech errors can therefore support the notion that articulatory gestures are discrete units of action.

2.2.2 Dynamical specifications of articulatory gestures

Based on these and other similar findings from articulatory investigation, Goldstein and Fowler (2003: 161) further put forward three key hypotheses in the framework of Articulatory Phonology about the nature of phonological units:

i) vocal tract activities can be analysed into constriction actions of distinct vocal organs;

ii) constriction actions are organised into temporally overlapping structures;

iii) constriction formation is appropriately modelled by dynamic systems.

Six distinct vocal organs are employed in the formation of constriction in the vocal tract. There are, from anterior to posterior, lips, tongue tip, tongue body, tongue root, velum and larynx. A constriction action executed by one of these organs is known as an articulatory gesture (Goldstein, 1993).

Vocal organs involved in the actions of constriction and release in the vocal tract are intrinsically DISTINCT and INDEPENDENT, in spite of the fact that the constricting organs may share the articulators and muscles with one another. It has been mentioned in the previous section that each gesture can be seen as a dynamical system. For instance, the constriction gestures operated by the lips, the tongue tip and the tongue body can share the mechanical degrees of freedom of mandible movements.

The vocal organs are distinct in the sense that the constrictions produced by different vocal organs are different; they are independent in the sense that with each of these vocal organs, a constriction or a release in the vocal tract could be formed without necessarily causing a constriction or a release by any one of other organs. Since vocal organs are distinct and independent, the constriction actions cause by such organs, also known as gestures, embody the nature of discreteness, which can further create phonological contrasts. That these gestures act independently in order to produce different constriction actions makes them function as the atoms in a combinatory system – the phonological system. A tongue body gesture can be contrasted to a tongue tip gesture because different vocal organs are employed to constrict the vocal tract. Given that lexical items can be treated as sequences of constriction gestures (Browman and Goldstein, 1992), two lexical items can be distinguished based on the nature of discreteness of the constriction gestures. In other words, two sequences of gestures could be defined as contrasting with one another as long as they contain at least one distinct constriction gesture. The words pen and ten, for example, contrast with one another in that in the onset position of both words, the former contains a constriction gesture by the lips while the latter by the tongue tip. Nevertheless, it is true that not all the phonological contrasts present in the languages of the world involve gestures realised by different vocal organs. The lexical items such as *pan* and *fan* could be differentiated by the nature of the constriction actions, i.e., the articulatory goals, of the labial gesture found at the onset position of both words. It is also argued by Goldstein and Fowler (2003) that between-organ contrasts are the primary ones within phonological systems, while within-organ contrasts (such as [p] vs [f]) tend to be language-specific.

A contradiction arises when describing physical speech in terms of discrete phonological units due to the fact that the articulators move continuously along the temporal dimension. In fact, in most phonological and speech production theories the continuous development of speech signals over time is simply modelled by the “interpolation” at the level of phonetic realisation between the consecutive “phonological targets," such as the speech error model described before, the window-model of coarticulation by Keating (1985), the autosegmental-metric approaches of intonation (Beckman and Pierrehumbert, 1989; Gussenhoven, 2004; Ladd, 2008) and so on. From such perspectives, phonological targets and production are two distinct notions of the theory. As commented by Goldstein and Fowler (2003: 166), "[i]n this view, the phonological information (the target) is not a part of production itself – it is hidden within the mind of the producer, and is only communicated through the resulting interpolation." Thus, how the continuous sound flow can be characterised by a finite number of discrete primitives becomes a crucial theoretical issue in Articulatory Phonology, rather than a problem on the mapping between phonological representation and phonetic realisation.

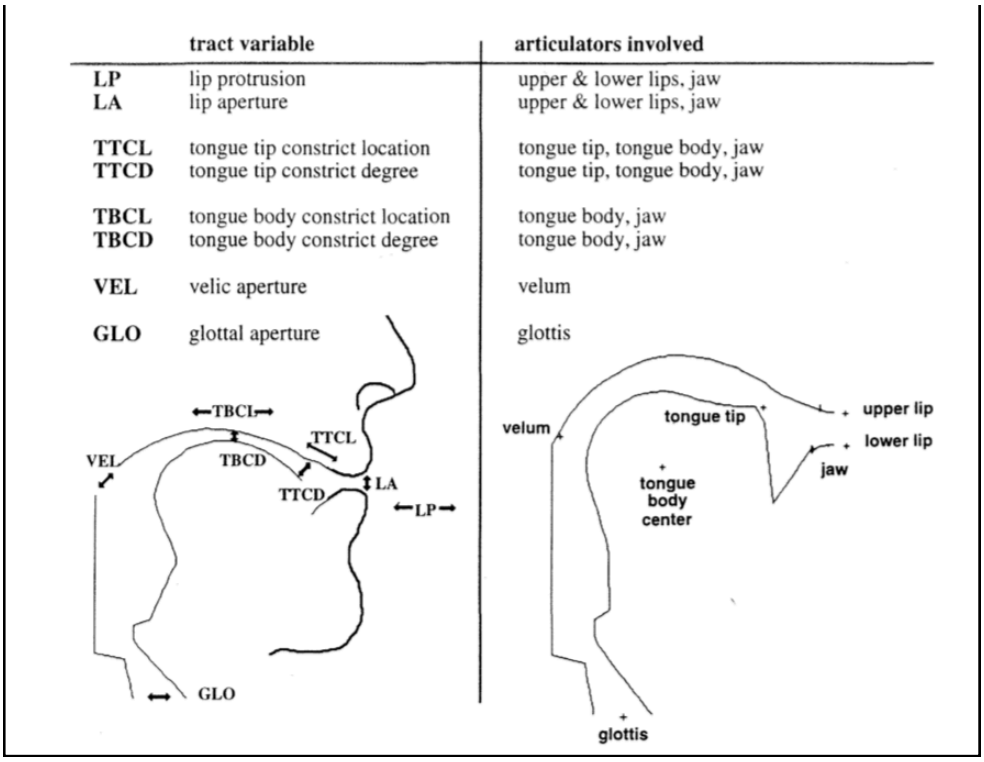
AP incorporates articulatory gestures into a quantitative computational model – a task dynamic approach based on a critically damped mass-spring model (Saltzman 1986, 1991, Saltzman and Kelso 1987, Saltzman and Munhall 1989). As the term task dynamic implies, such model describes the articulatory movement in terms of the 'task' to be executed, and such task is dynamic *per se*. It should be clarified immediately that it is not the articulators involved in the execution of a certain task that are dynamic, but the task itself. The task can be thought of as a goal of articulatory movements, such as creating a constriction somewhere in the vocal tract in order to produce a consonant (e.g., the tongue tip moves towards the alveolar ridge to form an oral fricative /s/), which corresponds precisely to what defines an articulatory gesture. Thus, such an articulatory movement pattern or constriction action is linguistically meaningful. The dimensions along which a linguistically meaningful task fulfils its goal are called '(vocal) tract variables', which corresponds exactly to the vocal organs.

The tasks – articulatory gestures – are typically differentiated along two dimensions: constriction location (CL, specific for tongue tip and tongue body) [7] and constriction degree (CD). The motion of tract variables, not the movement of the individual articulators, plays an important role in the framework incorporated with AP and the Task Dynamics. The six tract variables are further divided into three subsystems: oral, velic and glottal. They differ from one another in parametric settings: the velum and the glottis have default settings (e.g., for /p/ the tract variable glottis is specified with [wide] while for /b/ the glottal activity remains without specification) with respect to the oral

---

[7] The specification of constriction location is not excluded by the theory for lips either; for example, /f/ and /v/ might be understood as a labial gesture with the constriction location at the super teeth.

tract variables. Figure 2.1 gives the inventories of the vocal tract variables involved in the approach of AP.



	tract variable	articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

__Figure 2.1:__ Adapted from Browman and Goldstein (1992, 1993). Inventories of (vocal) tract variables and their associated articulators, along with the specifications of parameters of Constriction Degree (CD) and Constriction Location (CL).

The possible values (also called 'descriptors' in Browman and Goldstein, 1989) of CL and CD are shown in (a-b):

(a) Constriction degree: [closed], [critical], [narrow], [mid] and [wide];

(b) Constriction location: [protruded], [labial], [dental], [alveolar], [postalveolar], [palatal], [velar], [uvular] and [pharyngeal].

Given that the dynamically defined articulatory gestures are posited to be atoms of phonological structure in AP, it is important to note that articulatory gestures are relatively abstract as well. That is, the physically continuous movement trajectories could be further analysed as resulting from a set of discrete, concurrently active gestural regimes. These gestural control regimes are discrete in the following two dimensions:

i) the dynamical parameters characterising a given tract variable remain constant throughout the discrete temporal interval of gestural activation;

ii) the modification of the values of dynamical parameters associated to a given gesture might give rise to another gesture, in a given language the latter may differ from the former one in discrete ways, as represented by different dynamical specifications.

The articulatory gestures, thus, have a dual function (Browman and Goldstein, 1986, 1989, 1992; Goldstein and Pouplier, 2014; Pouplier, 2011): that of characterising the actual physical movements of articulators observed (as units of action) and that of capturing the phonological contrasts (as units of information). It should be noted that the discrete nature of articulatory gestures implies an intrinsic temporal dimension (an invariant temporal interval, see also the section 2.1.2) further characterizing them. In other words, in addition to the specification of dynamical values, articulatory gestures are also endowed with an intrinsic activation interval. Such activation interval is then referred to as *stiffness* in the Task Dynamic model. It is precisely the temporal specification of gestures that makes AP different from most phonological theories in which the minimal units represent only timeless phonological events as phonological targets, while in Articulatory Phonology, phonological primitives are also specified by their intrinsic temporal activation interval.

Different perspectives on phonological primitives within various theoretical approaches will be further discussed in detail in the section 2.3. In the following section the focus will be put on how the discrete and concurrently activated gestures combine together to form lager units.

2.2.3 Gestural scores: articulatory tiers and temporal organisation of gestures

2.2.3.1 A preliminary understanding on the inter-gestural coupling relation

It is introduced above that articulatory gestures can be represented in terms of Task Dynamics and gestures are defined with a series of dynamical specifications which further enable a description of the articulatory movements at the macroscopic level from the perspectives of both inherent space and intrinsic temporal interval. Task Dynamics, thus, could serve as a theory of how continuous motion of articulators can be divided into a limited number of discrete and qualitative underlying units of actions, *i.e.*, gestures, required for phonological contrasts (Goldstein *et al*., 2006, 2009; Saltzman *et al.*, 2008; Nam *et al.,* 2009; Pouplier, 2011). Now the crucial questions arise: in which way are the combinatoric articulatory gestures assembled and what is the 'glue' (mechanism) that binds the discrete units of action together? In other words, how is the phonological structure organised in terms of dynamic gestures?

Since the human speech system is a self-organised system (recall the parallelism with limb motion in the section 2.1.2), articulatory gestures are similar to the gestures of limb motion, *e.g.,* they are both gradient and qualitative; their intrinsic interval to achieve the goals are dynamically specified underlyingly in both macroscopic and microscopic dimensions, *etc*. Thus, as the coordination patterns make sense in the limb motion, phasing relations - temporal activation sequence - stand for a similar meaning in gathering the articulatory gestures together. Based on the experimental studies on limb motion, it is inferred that the modes along which an action can be coordinated with another are rather limited - basically, two: symmetric and antisymmetric. Moreover, stable modes (low dimensional macroscopic properties of the system) may arise and an uni-directional switch-of-mode (from less stable to more stable) can take place as the response to the local variations (*e.g.*, the frequency of the motion) (Turvey, 1990; Schmidt *et al*., 1990; Browman and Goldstein, 1990b, 2002; Hermes, 2014).

Phasing relationships are therefore those that serve as the glue which further assemble the articulatory gestures together. The exact coupling modes that emerge from such constellation of gestures make the system structured in an organised way. The question, now, becomes the following one: what are the functions of the coordination patterns in the phonological structure of human speech? The patterns of coordination modes are independent of the speech system and of the gestures involved in the sense that phasing relations are potentially available in any self-organised physical system. Therefore, the way in which the particular gestures participate in the coupling relation is informational (Browman and Goldstein, 2002). For example, consider the English words “dab” and “bad”; for both of them three gestures are basically involved: a closure gesture of tongue tip (TT) at the alveolar area, a lip closure gesture and a pharyngeal constriction exerted by a tongue body (TB) gesture. In the word “dab”, the TT gesture and the TB gesture, which characterise the onset consonant and the nuclear vowel respectively, coordinate in a synchronous way, whereas the TB gesture and the lip gesture for the coda show an asynchronous or sequential) way of coordination. By contrast, in the word “bad”, both the articulatory gestures and the coupling relations involved are identical to the “dab”, but they are applied to different set of gestures, i.e., the lip closure gesture and the TB gesture assume the (synchronous) relation previously observed between the TT and the TB gestures, while the relation found between the TB and the lip closure gestures is the asynchronous one. Thus, the coordination modes of gestures are able to contrast the phonological structures of different lexical items, and they are part of the macroscopic properties of the speech system. Details on inter-gestural coordination will be further discussed in Chapter 3.

The coordination modes are not consistently implemented in all languages. A direct supporting evidence comes from the fact that, for instance, in some languages coda consonants are not allowed, which means that the speech system of these languages does not employ a particular type of coordination mode to signal phonological and lexical contrast. Furthermore, the coordination modes are not applied to all types of gestures within a language. Some languages only allow coronals in coda position, but not labials or dorsals. Thus, in these languages, a TT gesture exhibits one further coordination mode than tongue tip gestures, for example, when coupling with the TB gesture for the adjacent vowel gesture. In addition, the implementation of the coordination modes can vary from language to language in the sense that some languages allow the inter-gestural coupling of the gestures belonging to different levels - segmental and suprasegmental - in relation to the phonological and lexical contrast. A detailed discussion in this respect will be addressed in the section 4.4 on the intra-level gestural coordination in Mandarin Chinese lexical tones. Finally, one type of coordination mode or another can also exhibit a number of language-specific microscopic variations. These superficial[8] variations have usually been described in terms of coarticulation, as it will be discussed below in Chapter 5.

2.2.3.2 Gestural scores

In AP, the utterance is produced via the temporal coordination of articulatory gestures. Then, the gestural structure, i.e., the abstract representation of gestural planning with respect to the (temporal) relation among the individual gestures' activation interval, could be represented in terms of the diagram called gestural score (Browman and Goldstein, 1992, 1993, 2002; Goldstein *et al.*, 2006; Pouplier, 2011). On one hand, the gestural score provides a visual impression on the actual parametric settings of the gestures and their phasing relations; on the other hand, it gives a direct visualisation on how atomic units (articulatory gestures) are bound together into larger units. The gestural score is supposed to be a universal phonetic tool which could capture not only within-language contrasts (*e.g.*, "dab" *vs*. "bad") and microscopic variations, but also cross-linguistic differences (Browman and Goldstein, 1993, 2002).

Within a gestural score, gestures are temporally organised in several tiers, which correspond to distinct tracts variables. The dynamical parameters such as constriction degree (CD) and constriction location (CL) of each gesture are further specified through distinct values. In addition, the activation interval of individual gestures is plotted in the gestural score as dark boxes. Figure 2.2 shows the gestural score for the English word "pawn". There are five articulatory tiers, corresponding to five distinct vocal tracts used in the articulation of the word. They are arranged in parallel along the vertical dimension in terms of dark boxes. The horizontal span of the box represents the activation interval of gestural control of each tract variable. The labels on the box show the actual dynamic specifications of the gestures. For instance, the lips gesture at the beginning has a value [labial] for CL and the value [clo] for CD; the tongue tip (TT) gesture in the word final has a TTCL at the alveolar region (specified as [alv]) and a TTCd of [clo] indicating the

---

[8] According to Browman and Goldstein (2002), these language-specific microscopic variations are referred to superficial consequences resulted from the use of one or another coordination mode in that particular language.

hit of the tongue tip against the alveolar ridge; the tongue body (TB) gesture for the vowel /ɑ/ has a TBCL at pharyngeal area and a TBCD of [narrow]; the velum and the glottis,[9] as mentioned above, have the default setting [wide].

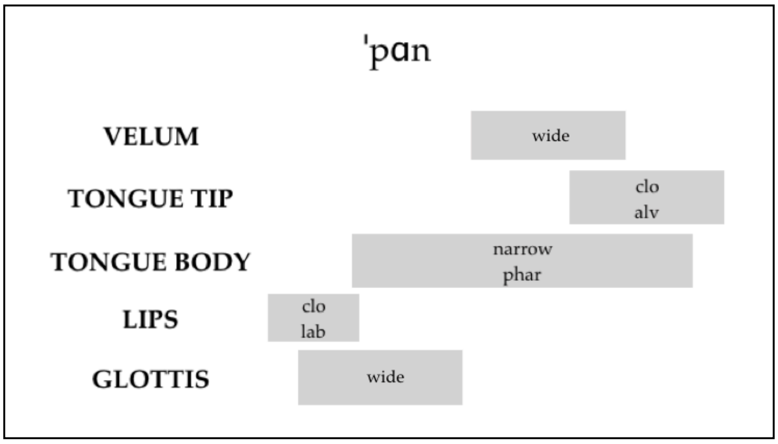


Figure 2.2*: The gestural score for the English word "pawn", based on Browman and Goldstein (1995).*

Although the gestural score shows the overall timing relations, it does not tell us the particular phasing relations between the particular gestures, in other words, which gesture is coupled to one another in which mode of coordination. In order to show the particular phasing relations, one thus needs another representation at a more abstract level - the gestural structure. By hypothesis, the gestural structure is the most abstract level of representation of the gestural planning, as it captures the tasks of the gestures, the underlying phasing relations between gestures, and the articulatory properties of the gestures from which the resulting gestural score can be generated (Goldstein *et al.*, 2006; Pouplier, 2011).[10]

Figure 2.3 gives an illustration on the gestural structure and the gestural score of the English word "bad". The gestural structure provides the macroscopic phasing relation between the gestures of different prosodic constituents (*e.g.,* LIPS closure for syllable onset, the TB gesture for the nucleus and the TT closure for the coda) without specifying

---

[9] The glottal aperture gesture (a wide abduction of GLO) at the beginning of the word is still activated after the release of the labial closure, the lag between the right edges of two gestures respectively indicates the period of aspiration (Browman and Goldstein, 2002; Goldstein and Fowler, 2003).

[10] It should be emphasised that the overall timing pattern visualised in a given gestural score results from the inter-gestural phasing relations under the circumstance of speech rate or prosodic context: the following chapters will show that speech rate and prosodic context may cause some important variations on timing patterns (e.g., Stetson 1951).

the exact activation duration, while the gestural score shows the temporal activation interval for each tract variable derived from the analysis of articulatory data[11] under particular circumstances (contextual and speech rate, for instance). In the diagram of the gestural structure, the solid line and the dotted line indicate respectively two different coordination patterns observed for the timing between different syllabic constituents, that is, onsets and codas (see also the section 3.2).

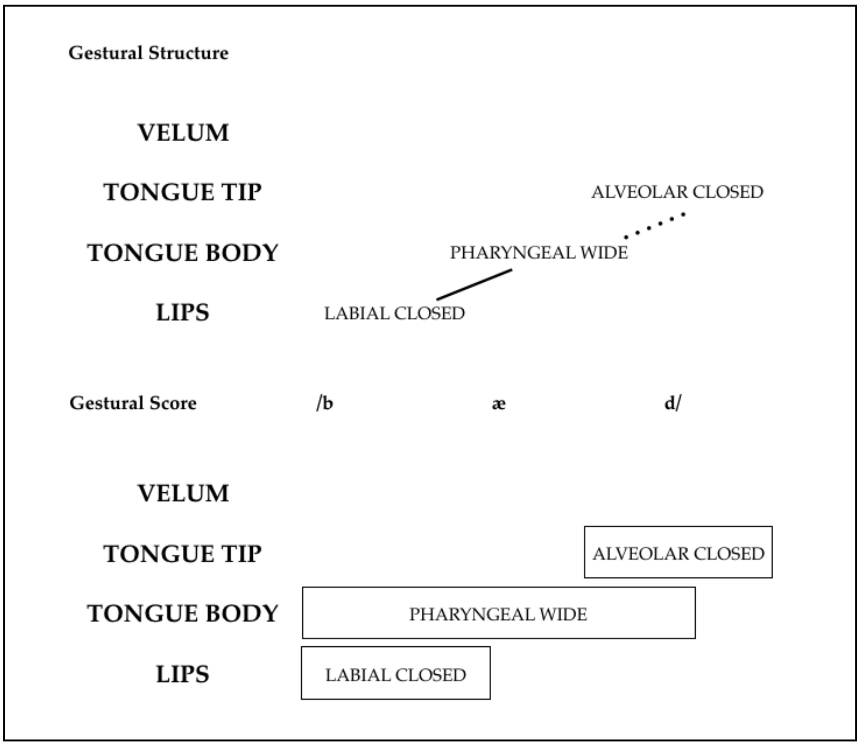


Figure 2.3: *Adapted from Pouplier (2011). The gestural structure and the gestural score for the English word "bad". The upper panel illustrates the abstract representation - gestural structure - of the lexical item; the gestural score is shown in the lower panel. In the upper panel, the solid line indicates one coordination mode (synchronous) usually observed for onset and nuclear vowel; whereas the dotted line represents another phasing mode (asynchronous) typically characterising nuclear-coda timing.*

Once the gestural score is introduced to the AP for a representation of the gradient timing properties of the categorical gestural actions, the richness of the phonological

---

[11] Although the gestural score is able to show the overall timing among the gestures involved in an utterance, yet the gestural score is still a theoretical analysis in the sense that there is no general procedure or algorithm that can generate automatically the gestural score from an arbitrary utterance as input in every circumstance. In particular, there are some gestural overlapping cases in which it is even harder to determine which one of the overlapping gesture is responsible for a certain observed articulatory movement (Browman and Goldstein, 2002). Usually in AP, the analysis-by-synthesis procedure (Browman, 1994) incorporating the task dynamic model could be used to determine the gestural score for these tough cases (see also Browman and Goldstein, 2002 for a detailed description on such four-step procedure).

structures could be explained in terms of AP through the patterns of how the articulatory gestures are organised in time with respect to one another. In particular, microscopic contextual variations (usually referred to “coarticulation”) and allophonic variations might be attributed to differences in gestural overlap (inter-gestural timing). In other words, the same set of gestures could result in different gestural constellations, and thus give rise to different superficial phonetic realisations. Therefore, articulatory gestures and gestural overlap in human speech are able to do more than merely capturing the phonological and/or lexical contrasts. In the next Chapter, we will introduce how the gestures coordinate with each other in the temporal dimension and form the coordinative structure in the domain of syllable.

Chapter III Internal Temporal Organisation of Syllables: A Macroscopic Perspective

3.1 Coupled oscillator model of inter-gestural timing

In non-linear phonologies, prosodic structure is defined by prosodic constituents in terms of hierarchically-organised structure (*e.g.*, Nespor and Vogel 1982, 1986 and Selkirk 1982, 1984 for Prosodic Phonology; Hayes 1982, 1985 and Hogg and McCully 1987 for Metrical Phonology). Gestural theories, from a partly different perspective, model syllable organisation on the basis of different inter-gestural timing. That is, the syllable is viewed as result of typical patterns of temporal coordination among multiple articulatory gestures. Although that the patterns of gestural constellations vary systematically, two preferred patterns (in-phase and anti-phase) are sufficient to reflect syllable structure (Browman and Goldstein 1988, 1995, 2000 and Krakow 1993, 1999). The internal organisation of syllable is assumed to be the projection of different patterns of the inter-gestural timing.

Krakow (1993), in her study on nasals and laterals in English, reported that inter-gestural timing can vary systematically as a function of syllable position. More specifically, she recorded the gesture of velum lowering for the nasal /m/ in the sequences such as see#me and seem#E and found that the inter-gestural timing between velum lowering gesture and the labial gesture differs according to the different segmental conditions where they are found. For a word-final nasal consonant /m/, the velum lowers during the interval of the preceding vowel; when the velum gesture achieves its target, the lips start the closing movement. For the word-initial nasal, in contrast, velum lowering is roughly synchronised with labial closure. Browman and Goldstein (1995) then argued that syllable structure is epiphenomenal with respect to different relative timing relations in the activation interval of the articulatory gestures involved. Indeed, such a viewpoint does not shed the light over the status of the syllable as a phonological unit, instead, it casts the shadow over the hierarchically-organised structure of the relevant prosodic constituents, for example, the syllable.

As anticipated in 2.1.2, gestural timing in AP is viewed as an intrinsic property of a dynamic system, similar to the timing relationships holding in other dynamic systems of the human body. The inter-limb coordination, for example, has two intrinsic modes - in-phase and anti-phase. Interestingly, Turvey (1990) shows that it is possible to shift from one mode to another when the frequency of the movement is increased. More specifically, the anti-phase coordination could switch to the in-phase coordination with the increase of the frequency, but not vice versa. The upper panel of the Figure 3.1 illustrates the in-phase and anti-phase coordination for the finger movement. Starting from an anti-phase coordination, that is, when the index finger on the left begins to move downwards and arrives its target, the right one begins to move upwards, an increase in the frequency of such sequential activation of the movements will result in an in-phase coordination between two index fingers, thus, two fingers move toward the same direction simultaneously, as shown in the right side of the upper panel. The limb coordination presented in the lower panel shows the same result. Moreover, if a pair of coupling movements begins in the in-phase mode, as the frequency of the movement increases, the original coupling pattern - in-phase mode - remains unshifted. These results suggest that the in-phase coordination is the most stable one and it is acquired automatically without learning.

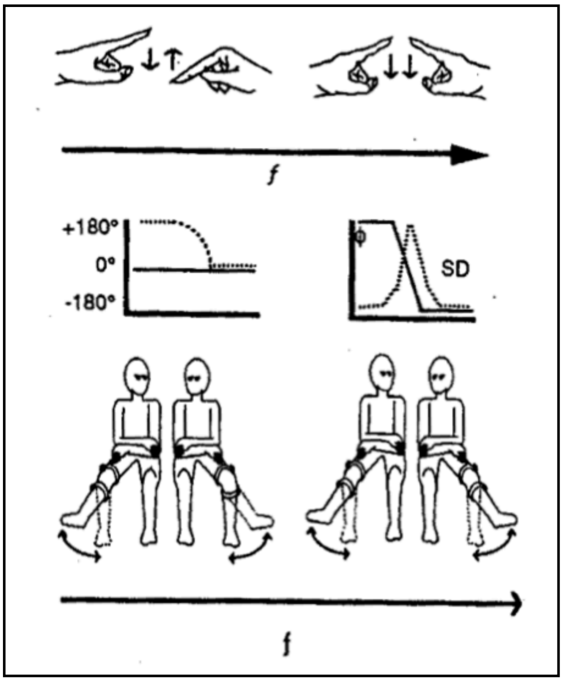


Figure 3.1: *Two intrinsic coupling modes for the finger movements and in limb coordination: in-phase and anti-phase, adapte from Tuvey (1990).*

The coordination patterns observed from body gestures are then also applied on the coordination of articulatory gestures within the framework of AP. For the purpose of modelling speech production planning in terms of integration of dynamic gestures and relative coupling, each articulatory gesture, as unit of actions, is then associated with a planning oscillator, or a clock. Within a given domain, the relative timing among/between gestural activations can be manipulated (planned) by the coupled, non-linear oscillators. The phonological information is thus stored at both levels, *i.e.,* at the level of gestures and at the level of temporal glue - relative timing between/among gestures. The activation of a given gesture is trigged at a particular phase of its oscillator. A pair of gestures can be coordinated to one another in time by coupling their oscillators respectively. Once a coupling pattern of relative phase between two oscillators is stabilised during their internal planning, the activation of each gesture can be triggered by its corresponding oscillator. As a result, stable relative timing between two gestures is achieved. Two basic inter-gestural coupling modes thus emerge: simultaneous activation (in-phase mode) and sequential activation (anti-phase mode).

Again, similar to the coordination of body gestures, the two inter-gestural timing patterns are intrinsic, therefore no learning is required. The in-phase coordination is more stable than the anti-phase coordination. And with the increase of the frequency of the gesture, in particular, the speech rate, at a critical stage, there is an abrupt pattern shift from the anti-phase, to the in-phase mode (Goldstein *et al.*, 2007, Hermes *et al.*, 2017).

3.2 Coupling oscillator modes within syllables

3.2.1. Onset-nucleus coordination and nucleus-coda coordination

Goldstein, Byrd and Saltzman (2006) proposed to extend the coupling hypothesis to account for syllable structure. According to Goldstein and colleagues, there are two basic types of gesture - consonant gesture and vowel gesture. Consonant gestures are broadly referred to as the closure gestures of lips, tongue tip and tongue body. They can be triggered simultaneously (in-phase inter-gestural timing) with velum gesture and/or glottis gesture, giving rise to a nasal or a voiceless consonant. Two important differences are found for consonant gestures (henceforth C gestures) and vowel gestures (henceforth V gestures):

i) V gestures are less constricted than (stop or fricative) C gestures;

ii) V gestures are generally more slowly and thus the gestural activation of vowel span longer in time than that of consonant. Therefore, the intrinsic stiffness of the vowel gestures has a lower parametric value.

The temporal activation of consonant and vowel gestures can thus be made reference to the coupling (non-linear) oscillator model of inter-gestural timing, assuming a coordination in either of two intrinsically stable modes observed from the movements of body gestures: in-phase (the most stable one) and anti-phase. It is hypothesised that the syllable initial consonant coordinates in an in-phase mode with the following nuclear vowel (Goldstein *et al.*, 2006), which means the onset of the C gesture is synchronised with the onset of the V gesture. The evidence is shown in Figure 3.2 in which the time function of the tract variables for the phrase *pea pots* is represented. Based on the plotted data of the velocity over time of LA and TB (the tract variables of /p/ and /a/ in pots, respectively), the onset of the lip gesture - the C gesture - is roughly synchronous (Δt is less than 25 ms) with the tongue body gesture for the vowel (Goldstein et al. 2006, Nam et al. 2009), which suggests an in-phase coordination between C and V gestures.

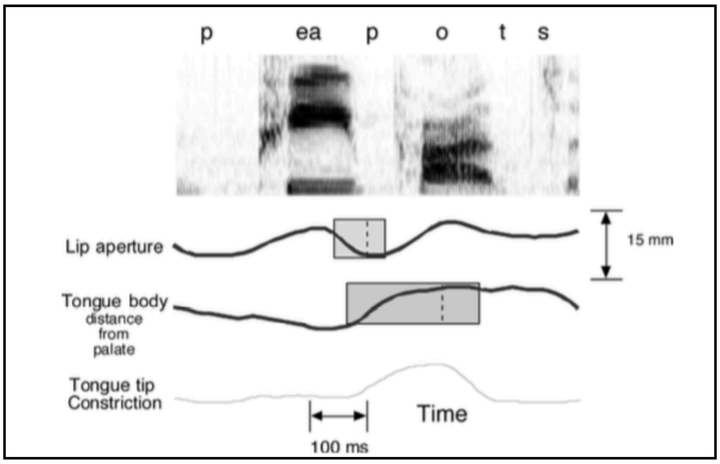


Figure 3.2: *Adapted from Goldstein et al. (2006). The boxes delimit the temporal activation of the constriction gestures for LA (roughly coincides with the onset C gesture in pots) and TT (the following nuclear V gesture). The left edge of the boxes represents the onset of the gesture.*

Since the in-phase mode is the most stable one, the simultaneous coordination can emerge spontaneously without being acquired. This also explains why CV syllable is preferred in the early stage of language development. In contrast, the relation between nuclear V gesture and the coda C gesture is supposed to be coordinated in the anti-phase mode. The inter-gestural timing of the English word spot in Figure 3.3 reveal that the tongue tip gesture - the C gesture for /t/ in coda - occurs at a point which corresponds almost to the final phase in the control of the vowel gesture - the onset of the closure of the tongue tip gesture approximately coincides with the end of the stability of movement of tongue body gesture.

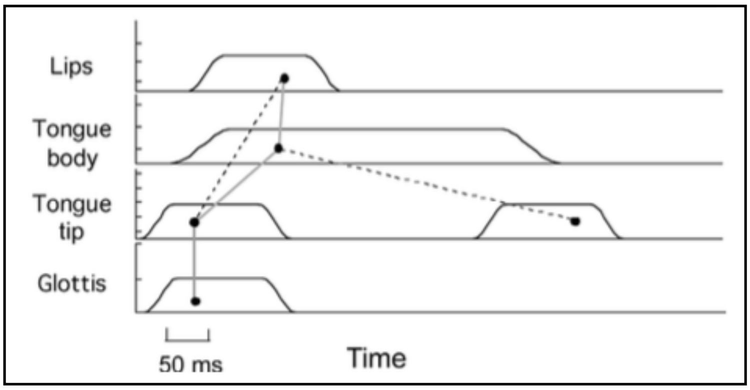


Figure 3.3: *Adapted from Goldstein et al. (2006). The coupling graph of the inter-gestural timing of the articulatory gestures employed in the production of the English word* spot. *The vertical dashed line on the right side indicates that the end of the stable status of tongue body approximately coincides with the onset of the closure of tongue tip, implying a sequential order between the gesturers involved.*

3.2.2 Coordination within consonant clusters

3.2.2.1 Competitive coupling mode in onset consonant clusters

Both the empirical studies and the theoretical interpretation based on articulatory data have shown that the onset consonant is associated with the vowel differently than the coda consonant. These articulatory studies concern only the cases in which the syllable onset and coda are singleton. Given a segment sequence /spa/, for instance, the gestural timing model of syllable structure predicts that there is an in-phase relationship between onset and the following /a/, thus, both the C gesture of /s/ and the C gesture of /p/ are supposed to be in-phase with the nuclear V gesture, which implies that the two C gestures are activated simultaneously. Such synchronisation of multiple C gestures might result in a perceptually unrecoverable structure[12] (Nam *et al.*, 2009, Goldstein and Pouplier 2014, Bruner *et al.*, 2014). Since the gestural timing within

---

[12] In some multi-gestural consonants like nasals or voiceless stops, synchronising oral constriction gesture with the velum or glottis gesture still produce a perceptually recoverable structure, for example, a glottal aperture gesture does not block the perception of a sumultaneous tongue tip closure gesture at dento-alveolar aera, and thus /t/ and /d/ can easily be distinguished by the speakers. Similarly, the velum raising gesture can occur simultaneously with the tongue tip closure at the dento-alveolar position without necessarily obscuring the tongue tip closure. Such inter-gestural synchronisation could further contribute to the contrastive characteristics of the multi-gestural consonants like /t/ and /n/.

consonant clusters cannot be compatible with the current coupling modes, there must be other coordination relations than in-phase and anti-phase mode.

Browman and Goldstein (1988) first reported the C-centre effect for the complex onsets in their empirical studies on English C(C)(C)VC(C) sequence. They find that the centre, termed as the temporal midpoint, of the onset consonant maintains a rather stable relation to the following vowel anchor point regardless of the number of the consonants found in the syllable onset. The anchor point is defined as the articulatory offset of the vowel gestural activation - the acoustic onset of closure of the postvocalic consonant is then chosen as the anchor point in their original studies (Browman and Goldstein 1988, 1990a). Therefore, in a syllable with the complex onset, the individual C gestures are, in a strict sense, no longer aligned synchronously with the following V gesture. Instead, the complex onset is coupled, as a whole, with the vowel.

The coordination relation between the complex onset and the nucleus is further modelled as a competitive mode by Browman and Goldstein (2000). The complex onset, as a unity, is supposed to be coupled with the nucleus in the in-phase fashion, *i.e.*, the midpoint of the onset of the C gestures is approximately synchronous with the onset of the V gesture, and the 'competition' results from the hypothesised simultaneous activation of individual consonant gesture and the nuclear V gesture. In addition, the consonant gestures in a complex onset are assumed to be triggered in an overlapping fashion (not sequentially ordered in a strict way) in order to result in a perceptually recoverable structure. For instance, in the /spa/ sequence, the upward movement of the tongue tip gesture for /s/ occurs before the tongue body gesture for the /a/, and the labial closure for /p/ is activated after the onset of the tongue body gesture, while the midpoint of tongue tip and lip gestures synchronises with the onset of the tongue body gesture. Figure 3.4 gives a schematisation of the gestural timing within the syllables with the complex onsets[13].

---

[13] In the model of split-gestural dynamics proposed by Nam (2007), the author finds a similar pattern in singleton onsets which he nominated single consonant C-centre effect, as opposed to the coordination pattern observed with complex onsets. Basically, he decomposes a single C gesture into two gestures - each of them is associated to a certain phase during the articulation of a stop - closure (CLO) and release (REL). The onset of the V gestural movement for the V gesture occurs after the beginning of the CLO gesture and before the beginning of the REL gesture. The midpoint of the onsets of CLO and REL gestures roughly coincides with the beginning of the V gesture. Such an approach, however, has two limitations: first, split-dynamics can be applied only to oral C gestures; for glottal, velic and V gestures the movement cycle needs to be modelled on the basis of a single gestural task (Nam *et al.*, 2004;

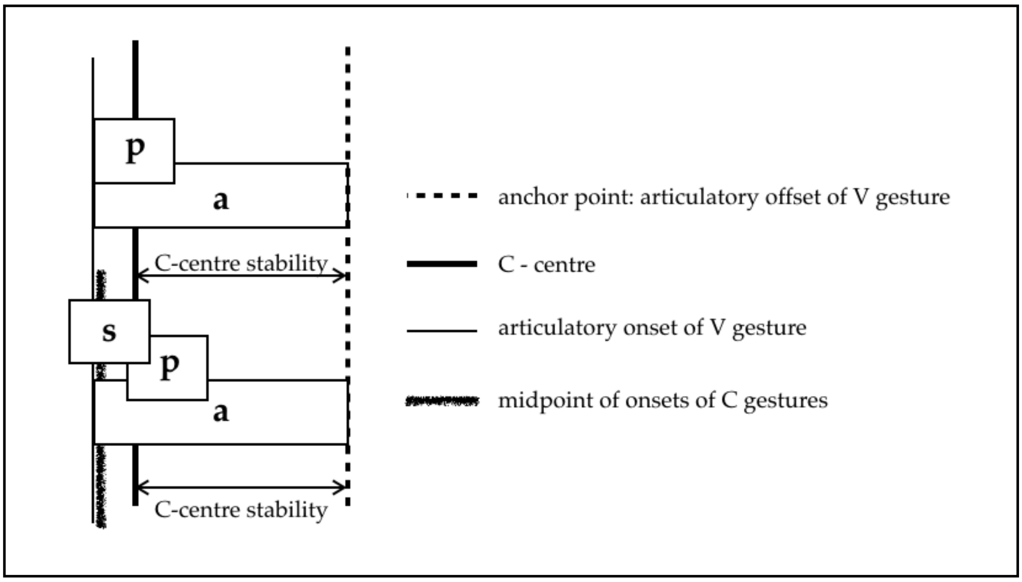


Figure 3.4. *Schematisation of C-centre stability and the gestural timing between nucleus and the complex onset: the C-centre, represented by a bold solid line, maintains a stable timing relation to the following vowel anchor point regardless of the number of the onset consonants; similar to the in-phase coordination found in CV syllables, the midpoint of the activation of the both C gestures in complex onset aligns approximately in an in-phase mode with the onset of the V gesture.*

Adding a consonant to the syllable onset does not lead to a modification of the relative timing of the C-centre to the following V anchor point. As shown in figure 3.4, the C-centre adjustment leads to a rightward shift of the consonant adjacent to the vowel when a consonant is added to the left-edge of the onset. If a consonant is added to right-edge of the onset, then, the leftmost consonant is expected to be shifted leftwards in order to make room for the added one. It should be noted that these shifted movements relative to the V anchor are measured in relation to the case in which the onset is a singleton. Figure 3.5 below displays an idealised pattern of the consonantal shift resulted from the C-centre effect. Such a kind of shift measurement was initially taken as an indicator for the identification of a legal complex onset in languages

---

Poiplier, 2011); second, the decomposition of the C gesture into CLO and REL gestures obscrues to some extent the distinction between C gesture and V gesture, which are supposed to be different in nature by hypothesis. Moreover, such an approach seems not to be consistent with the previous studies on the inter-gestural timing model. No matter what a C gesture is decomposed into two sequentially ordered phases or considered as a single gesture, in both cases the onset of the CLO gesture should be seen as the onset of the C gesture involved. Thus, as a logical consequence, the onset of the V gesture is more or less synchronised with the beginning of CLO, instead of being activated after it. Even the REL gesture adopted in the split-gesture dynamics is not able to push the CLO leftward because the REL gesture here is not a gesture added to the left of the V but a part of the C gesture. Therefore, as to the Conset-Vonset timing, there is a discrepancy between the two approaches (gestural dynamics *vs.* split-gesture dynamics).

(Browman and Goldstein 2000, Goldstein *et al.*, 2006, 2007, 2009). Several studies by Pouplier and colleagues (Marin and Pouplier 2010, Pouplier 2012) have subsequently established the shift analysis - *i.e.*, the measurement of the shift of the rightmost consonant in a consonant cluster in relation to the following V anchor - as a diagnostic of syllabic status of word-initial consonant clusters. These measurements provide further evidences for the hypothesis that syllable structure is reflected in the gestural coordination.

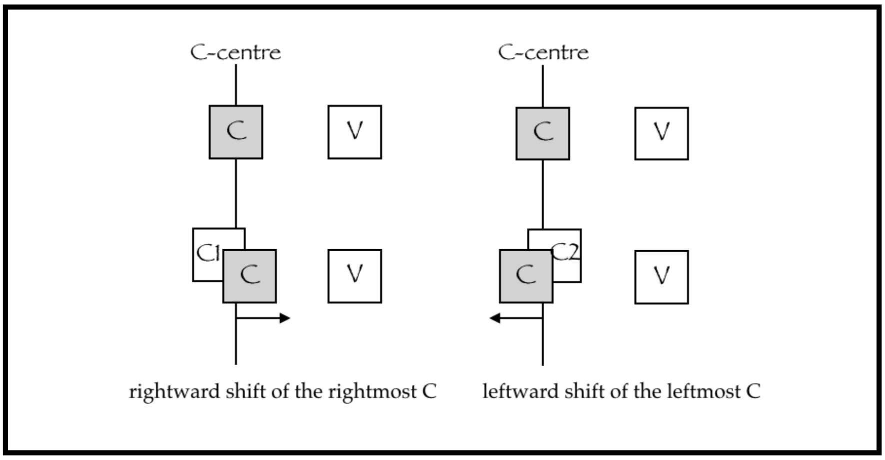


Figure 3.5. *Idealised patterns of consonant shift when another C is added to the syllable onset. When a C is add to the left of the onset (the left panel), the rightmost C in the onset undergoes a rightward shift relative to the following V anchor. Whereas a C is added to the right of the syllable onset, the leftmost C in the onset is expected to undergo a leftward shift with respect to the following V anchor. In addition, it can be observed in the figure that wherever a C is added, the C-centre remains to be unaffected.*

Actually, the gestural timing relation within the complex onset is not sequentially coordinated in a strict sense. Rather, C gestures in onset are partially sequential, or partially overlapping. Recent works (*e.g.*, Marin and Pouplier, 2010; Goldstein, 2011 and Pouplier, 2012) suggest an eccentric coupling mode rather than the anti-phase mode for the gestural timing with the onset C gestures. That is, each C gesture in a complex onset coordinates in-phase with the following V gesture, which results in a competition between/among the C gestures and gives rise to a global coordination between onset and nucleus. Beside this, the competition of the C gestures also leads to a partial overlap between C gestures and in this sense their activations can no longer be seen as strictly anti-phased. A relation between gestural overlap and the C-centre effect is then observed: C-centre stability is mostly found for consonant clusters with much gestural overlap (Pouplier, 2012; Brunner *et al*., 2014). The eccentric pattern is said to be not intrinsic, but rather it needs to be acquired. This would also explain why consonant clusters are acquired later than simple onsets in language acquisition (Nam *et al.,* 2009). Actually, as shown later in the present dissertation (the section 3.2.2.2 and the section 5.1.2), the eccentric coupling mode manifested by so-called 'C-centre effect' is not universally plausible for the gestural organisation within the complex onsets.

3.2.2.2 Cross-linguistic evidences on C-centre effect

The C-centre effect in word-inital consonant clusters has generally been observed in a variety of languages (*e.g.,* Goldstein et al. 2009 and Marin and Pouplier 2010 for English; Kühnert *et al*. 2006 for French; Goldstein *et al*. 2007 for Georgian; Hermes *et al.* 2008 for Italian[14]; Marin 2011 for Romanian[15]; Pouplier 2012 for German; and so on). Browman and Goldstein (1988) reported the first evidence for American English revealing a stable global organisation between the word-initial complex onset and the nucleus regardless of the number of the consonants found in word-initial position, thus supporting the view that syllabic structure is epiphenomenal with respect to gestural coordination patterns. Further evidence provided by Marin and Pouplier (2010) for S+C clusters is consistent with the results obtained almost two decades earlier. Their study confirms that although the /sC/ sequences in syllable onset violate the Sonority Sequencing Principle, they are still tautosyllabic in English, in other words, they form a complex onset.

Another study on American English was carried on by Waltl and Marin in 2010. They tested the gestural coordination pattern in the three-consonant-onset (/sCC/ cluster) syllables with an initial sibilant - /spl/ and /skl/. Their results are not in line with the previous empirical studies on the complex onsets with only two consonants (*e.g*., Browman and Goldstein 1988, Marin and Pouplier 2010). The results from the shift analysis show that the /l/ is not shifted rightwards in a consistent way compared

---

[14] In Italian, the syllable affiliation of word initial /sC/ clusters is always problematic. The detail will be reported in the following section.

[15] The Romanian data are actually not valid for all types of syllable/word initial consonant clusters (Marin 2011).

to the CC clusters and/or /sC/ clusters. Additionally, concerning the measure of stability, no evidence of C-centre stability is found between the triconsonantal sequence and the following V anchor. These findings lead the authors to conclude that unlike word-initial CC clusters, there is no consistent C-centre organisation holding for more complex onsets[16]. They further suggest that the relative infrequency of the CCC-sequences in American English might be responsible for the absence of C-centre effect and for the fact that the CCC-sequence appear to behave more similar to the cross-boundary (*e.g.,* across words) clusters.

In addition to English, word initial consonant clusters in Romanian have attract much interest. Particularly, a recent study by Marin (2011) investigates Romanian complex onsets with regard to the gestural coordination pattern in terms of C-centre organisation. Romanian is a language that shows a large tolerance for consonants clusters in syllable- or word-initial position. The possible cluster types include stop-liquid sequences (*e.g.*, /kl/, /pl/), stop-stop sequences (*e.g.*, /kn/, /kt/), /s/-stop sequences (*e.g.*, /sm/, /sp/, /sk/) and stop-/s/ sequences (*e.g.,* /ks/, /ps/). Hence, Marin (2011) divides the possible clusters into three series (/s/-series, /l/-series and stop series) and explores whether Romanian allows the complex onsets in all three series. The /s/-series refers to /s/-stop sequence, the /l/-series refers to stop-/l/ sequences and the stop series contains all the other consonant clusters in which the leftmost consonant is a stop, *e.g.*, /ps/ and /kn. The results indicate discrepancies among the three series. First, for the /s/-series clusters, both the shift analysis and the stability analysis confirm the C-centre organisation (see figure 3.6, top), which suggests that the /s/-stop consonant clusters in Romanian form a complex onset. For the /l/-series clusters, the scenario is less clear. While both the rightward shift of /l/ towards the following vowel anchor and leftward shift of the stop away from the vowel anchor are observed with regard to their timing as singletons (*e.g.*, [pl-] vs. [l-] and [pl-] vs. [p-]), the C-centres in two conditions (*e.g.,* the midpoint of [p-] and the midpoint of [pl-]) fail to line up with each other (see figure 2.6, middle). However, the C-centre is still the most stable landmark for the gestural

---

[16] The results of Waltl and Marin (2010) also show, however, the variations among speakers and the types of clusters, *i.e.*, segmental property of the consonant in the CCC cluster. For the details on the issue of segmental properties, see also the Chapter 4.

coordination in the stop-/l/ clusters, compared to the other temporal landmarks, *e.g.*, articulatory edges. Romanian syllable-initial /l/-series clusters can thus be treated as a complex onset as well. As to the stop-series, no C-centre stability is found[17]. In addition, surprisingly, the rightmost consonants in the clusters do not show the rightward shift as expected. Instead, both consonants shift away from the following vowel anchor (see figure 3.6, bottom). Moreover, the high variability of the right-edge timing rules out the hypothesis of a heterosyllabic constituency for stop-series clusters. As a result, the syllabic affiliation of stop-series clusters remains unclear. Marin (2011) assumes that this variability is probably due to the low frequency of the occurrence of stop-series clusters in syllable initial position in Romanian; as a consequence, the production system finds difficulty in elaborating them, since it is unlikely that consonant sequences are syllabified differently based on the different articulatory properties of the consonants involved in the complex onset (Hermes 2012).

---

[17] In a recent review, Hermes, Mücke and Auris (2017) point out that the choice of the anchor point in the study by Marin (2011) may also contribute to such tripartite patterns in Romanian onset consonant clusters. Sometimes, the coda consonant is defined as the anchor point, bot other times the onset consonant of the following syllable is chosen as the anchor, *e.g.*, /spa.l̲ə/ ('washes') *vs.* /psal̲m/ ('psalm'). According to Hermes *et al.* (2017), the articulatory anchors are required to be within the syllables involved, for the reason that the anchors could directly participate in the syllable internal coupling relation. Choosing an anchor outside the domain of the investigated syllable, such as the onset consonant of the following syllable, runs the risk of introducing more variations into the data and results in a higher variability. Such a viewpoint seems to challenge a crucial baseline of Articulatory Phonology, that is, what is syllable in articulatory terms? Browman and Goldstein (1992) conclude that syllable structures are reflected in the inter-gestural coupling relations and thus, syllable can be seen as a result of the gestural timing. Ladd (2014: 18) claims that the theory (of Articulatory Phonology) unambiguously treats articulatory gestures as the primitives and unambiguously considers segments to be epiphenomenal. Therefore, it is possible that in some cases the anchor is tautosyllabic while in other cases it is not, pertaining, for example, to the onset of the following syllable. In a study of German onset and crossword clusters, Brunner *et al.* (2014) choose an anchor outside the investigated syllable (the onset of the following one) and found the unexpected coupling patterns for /pl/ and /kv/ word initial clusters.

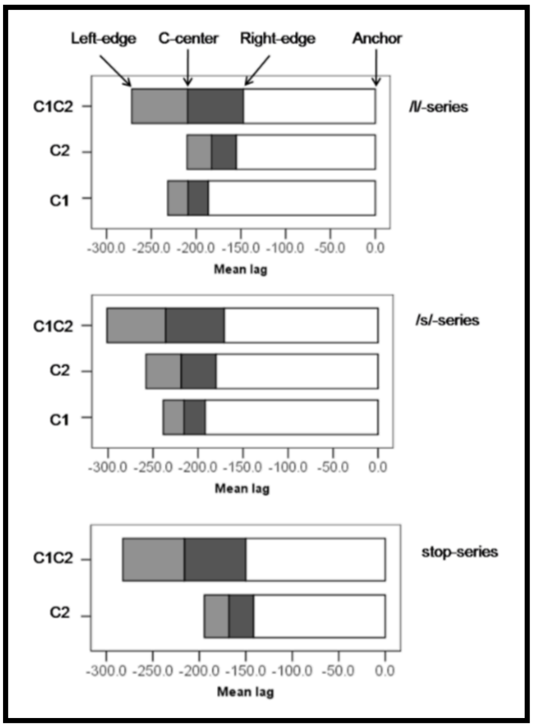


Figure 3.6: *Adapted from Marin (2011). Mean temporal latency from the vowel anchor to the temporal landmarks: left-edge, right-edge and C-centre for all three series of the consonant clusters*

Although several traditional phonological approaches (*e.g.*, metrical phonology) suggest that the 's' in the /sC/ or /sCC/ clusters is not a part of the onset in many languages, being treated as an extra-syllabic element (*e.g.*, in American English) because it violates the Sonority Sequencing Principle, articulatory evidences appear to lead to a different analysis. In both American English and Romanian, the /sC/ clusters are confirmed to be complex onsets. Then one might wonder whether different languages can share the same gestural coordination pattern - *i.e.*, a global, competitive coordination with C-centre organisation - for word- or syllable-initial /sC/ clusters, or rather languages have specific temporal organisation. The next section deals specifically with this issue, discussing the case of Italian /sC/ clusters.

3.2.2.3 Articulatory coordination of word initial /sC/ clusters in Italian

Word-initial /sC/ clusters in Italian have a very special status. Their syllabification is an issue of much debate in the literature. Allomorphic alternation of definite and indefinite articles, in the first place, suggests that word-initial /sC/ clusters differ from other consonant clusters in Italian (see table 3.1) [18]. The /s/ in the word-initial cluster of the masculinenouns which require the allomorphic forms is referred to in Italian as the "S impura" ('impure S'). Impure S does not undergo Raddoppiamento Sinttatico (henceforth RS) differently from onset /s/ (*e.g.*, è [s:]alvo ' he is safe') or other consonant clusters (e.g. è [b:]ianco 'it is white', è [b:]ravo 'he is good'). (Nespor, 1993; Bertinetto, 1999). Acoustic and perceptual studies fail to give a clear answer to the syllabic affiliation of the S impura, and the results of psycholinguistic experiments lead Bertinetto (2004) to conclude that the syllabification of /sC/ clusters is undecidable even by native speakers themselves.

definite articles (masculine)		**indefinite articles (m.)**
singular	**plural**	**singular**
il libro (Eng. *the book*)	**i** libri (Eng. *the books*)	**un** libro (Eng. *a book*)
il saggio (Eng. *the essay*)	**i** saggi (Eng. *the essays*)	**un** saggio (Eng. *an essay*)
lo studente (Eng. *the student*)	**gli** studenti (Eng. *the students*)	**uno** studente (Eng. *a student*)

Table 3.1: *Allomorphic variation of the masculine definite and indefinite articles before different word initial consonants in Italian, in both singular and plural forms.*

Hermes, Grice, Mücke and Niemann (2008, 2012) provide a preliminary articulatory study of the coordination pattern of word-initial consonant clusters in Italian. In the shift analysis, they test the rightward shift of the prevocalic consonant (the rightmost one in the cluster) for singletons, CC clusters, /sC/ and /sCC/ clusters. The results draw a clear distinction between CC clusters and /sC/ (including also /sCC/) clusters. In CC clusters, the consonant adjacent to the vowel shows a rightward shift towards the following vowel anchor in order to make room for the added consonant on the left, compared to the context with singleton onsets. The C-centre stability is observed as well. These results confirm that in Italian complex onsets are allowed, at least for CC clusters. In the case of /sC/ and /sCC/ word-initial clusters, on

---

[18] In the case of female nouns, the female forms of both definite and indefinite articles end with an *a/e,* i.e. *la, le* and *una*, respectively. Even if a given female noun begins with an *'impure S'*, the resyllabification process can still assign the *impure S* to the coda of the article, forming a well-formed syllable in Italian. Thus, no extra allomorph is needed for the female articles.

the contrary, no rightward shift of the rightmost consonant is attested and no C-centre effect is found. Even if an /s/ is added to the left of a singleton or a complex onset, the articulatory timing between the prevocalic consonant and the vowel anchor remains unchanged.

Based on these results, the authors claim that in Italian the /sC/ clusters constitute a special case in which the /s/ does not exhibit the articulatory timing required for the membership of the syllable onset (Hermes *et al.*, 2012). They further model the gestural coordination pattern for Italian consonant clusters according to which an eccentric coupling mode is suggested for the CC clusters due to the partial overlap of the two consonants. For the word-initial /sC/ clusters, since no evidence for an underlying competitive coupling relation is observed[19], they hypothesise that instead of being coordinated with the following vowel, the Italian 'S impura' joins the coupling relation with the vowel of the previous syllable, assuming an underlying anti-phase coordination (Hermes, 2012; Hermes *et al.,* 2008, 2012). As it will be shown in the Chpater 5 that C-centre effect is no longer viewed as the universal diagnostic of the complex onsets, the heterogeneaous behaviour of Italian word-initial /sC/ clusters could be interpreted in an alternative way, that is, the languages with a restricted phonotatics are endowed with less articulatory flexibility.

3.2.2.4 Complex coda consonants

While the complex onset shows a complicated gestural timing relation both internally and with respect to the following V gesture, such a global organisation is not found for complex codas, at least based on the limited data almost exclusively from English (*e.g.*, Browman and Goldstein 1988, 2000, Byrd 1995, Marin and Pouplier 2010, *etc*.). In these studies, it is proposed that the coupling relationship in VCC sequence is rather simple, the same as in VC sequences: the leftmost C in a complex coda coordinates with the previous V gesture in the anti-phase mode, and other C gestures couple only to the immediately preceding C gesture in a sequential pattern and not to the nuclear V

---

[19] In fact, there are phonological evidence that is contradictory to the results of Hermes et al. (2008, 2012) regarding on the /sC/ clusters. In particular, the word-internal intervocalic /sC/ clusters might be possibly parsed as tautosyllabic clusters.

gesture (Browman and Goldstein 2000, Goldstein and Pouplier 2014). The articulatory evidences provided by Marin and Poulier (2010) also confirm that the coordination of complex codas with the preceding vowel differs significantly from the coordination of complex onsets. If more consonants are added in the coda, only the leftmost consonant of the coda is in a stable relationship to the previous V gesture, irrespective of coda complexity.

To sum up, consonants in onset tend to be coupled with their following vowel in an in-phase mode but with their preceding vowel in an anti-phase mode when in coda; consonants in sequences, whether in onset or in coda, are organised sequentially. Then the question arises, why does the gestural internal organisation allow multi-mode/competitive coordination only in syllable-initial position (*i.e.,* in onsets) and not in syllable-final position (*i.e.,* in codas)? Browman and Goldstein (2000) suggest that the combinatorial strength between nucleus and coda is not sufficient to pull the more distant coda Cs into a coupling relation with the preceding vowel, which corresponds to the results of studies conducted on the articulatory timing of consonants within syllables - less gestural overlap is found for coda Cs as opposed to the onset C gestures with respect to the vocalic nucleus. In addition, the intrinsically strong combinatorial force induced by the participation in an in-phase coordination allows less variation of inter-gestural timing. As a consequence, gestural timing within complex onsets is generally more stable than within complex codas[20].

Such topological differences on gestural articulatory timing might account for the reason why coda consonants contribute to syllable weight in many languages whereas the onset consonants seldom do so (Browman and Goldstein 2000, Nam 2007, Nam *et al.,* 2009). It is proposed that the temporal interval between/among articulatory gestures can be further interpreted as the surface manifestation of the mora, one of the prosodic units hierarchically organised into the prosodic structure. Cross-linguistic evidences on competitive coupling mode for complex onsets as opposed to non-competitive coupling mode for complex codas suggest that coda consonants are

---

[20] Pouplier (2012) reports, however, that the opposite results are obtained for German. Coda clusters show less difference in their inter-gestural timing compared to onset clusters.

moraic, in that the coda consonants assume a strictly local, sequential coordination with both the previous nuclear vowel and the adjacent consonant(s); consequently, adding a consonant to the coda, rather than causing the decrease of the acoustic duration of the vowel, would increase the duration (or the weight, in metrical terms) of the entire syllable (Nam *et al.,* 2009). On the contrary, onset consonants are claimed to be weightless irrespective of the number of the consonants found in the onset position in almost all the languages investigated[21]. Based on this reasoning, an important implication can be put forth on the relation between gestural coordination patterns and syllable weight: in languages in which coda consonants do not bear syllable weight (*e.g.,* Malayalam), a competitive coupling mode will hold for the gestural coordination of coda consonants; this would result in a compensatory shortening of the tautosyllabic vowel, that is, vowel duration would be compressed as a function of the consonants added to the coda (Nam, 2007). Such hypothesis still remains an open issue nowadays, further cross-linguistic studies are needed to confirm whether the inter-gestural coupling relation reflects the syllable weight as well.

3.2.2.4 Syllables without vowels: so-called ‘syllabic consonants’

The coupling modes found for the syllable constituents timing so far are based on a common assumption, that is, vowel provides the basis for the inter-gestural coordination within syllables. Recall that in the original coupling hypothesis proposed by Goldstein *et al*. (2006), vowel gesture is treated as one of the two basic types of gesture which, in turn, enters into the non-linear oscillators coupling relations between/among articulatory gestures. The slower movement in time and the longer temporal activation of the vowel gestures make them essentially different from the consonant gestures. As a result, it is easy for the language user to recover both gestures participating in a preferred synchronous onset-nucleus coordination. There are, however, languages in which syllabic consonants are allowed, *e.g.*, Slovak and Tashlhiyt Berber. The onset-nucleus and nucleus-coda coordination can thus be seen as a subset of coordination options within consonant clusters, since there is no

---

[21] Ratak, a dialect of Marshallese, spoken in Philippine, might be an exception (Bender,1999; Nam *et al.*, 2009).

evidence that the syllabic consonants are more ‘vowel-like’ than their consonantal counterparts (Pouplier and Beňuš, 2011). In short, although syllabic consonants occupy the position of the nucleus in the syllables, their gestures are still treated as consonantal gestures. As a consequence, the canonic onset-nucleus in-phase mode might be no longer valid here probably due to a problematic perceptual recoverability resulted from the synchronisation of onset and a consonantal nucleus (see also the section 3.2.2) (Goldstein and Pouplier, 2014).

Slovak is a language allowing liquids (/l/ and /r/) as syllable nucleus. Pouplier and Beňuš (2011) report that in Slovak the inter-gestural timing in syllables with consonantal nuclei differs significantly from the timing in syllables with vocalic nuclei. Their results unfold an asymmetrical gestural overlap as a function of syllable position. In particular, for the consonant sequence /mrk/ with a traditionally believed consonant nucleus /r/, both /mr/ and /rk/ sequences show less inter-gestural overlap, thus greater latency between the gestural activations, as compared to the cases in which the same sequences are found in onset (*e.g.,* /mrak/) and in coda (*e.g.,* /park/). Moreover, even in the vowel-less syllables, a significant distinction between onset-nucleus coordination and nucleus-coda coordination is drawn, which again confirms that internal organisation of the syllable is expressed in terms of timing relations between its constituents, although the exact coupling mode within vowel-less syllables is still need to be further studied because on the one hand, onset-nucleus coordination relationship within a syllable varies systematically as a function of whether the syllable nucleus is a consonant or a vowel (Pouplier and Beňuš 2011); on the other hand, syllabic consonants constitute a situation in which the segmental identity of the gestures involved in a sequence (be it consonantal in nature or vocalic in nature) might play a larger role than the pure simplest forms of the gestural syllable model reported so far. Thus, onset-nucleus coordination depends on whether the nucleus is a consonant gesture or a vowel gesture (Pouplier 2012). Their study then reveals another interesting aspect regarding to the inter-gestural timing within the consonant clusters, namely, no C-centre effect is found in Slovak, neither for the consonant sequences containing a syllabic consonant nor for onset consonant clusters within canonical syllables, which means as more consonants are added to the onset, the C gestures are coupled in an anti-phase fashion, irrespective of the segmental identity of the gesture in the nucleus.

The syllabic consonants in Tashlhiyt Berber, a language spoken in Morocco, grabs lots of interests in recent years. In Tashlhiyt Berber, basically, any consonant can be syllable nucleus. Goldstein et al. (2006) show a pair of words - /tu.da/ versus /tb.da/ - in which both of the nuclei of the first syllable involve a lip gesture, but in one case it is a V gesture and a C gesture in the other. As shown in Figure 3.7, the C gesture specified for the tongue tip is activated synchronously with the vowel lip gesture. In contrast, an anti-phase relationship can be observed for the coordination between the onset tongue tip gesture and the nuclear C gesture. This is consistent with the results obtained from Slovak, which implies that the onset-nucleus coupling mode must depend on whether the nucleus is a C gesture or a V gesture. Since a cross-linguistic preference on the in-phase onset-nucleus coordination fails to provide a valid interpretation of the results from the Slovak and Tashlhiyt, it is possible that languages may phonologise a greater range of coordination patterns than has been thought possible so far (Pouplier and Beňuš 2011, Pouplier 2012).

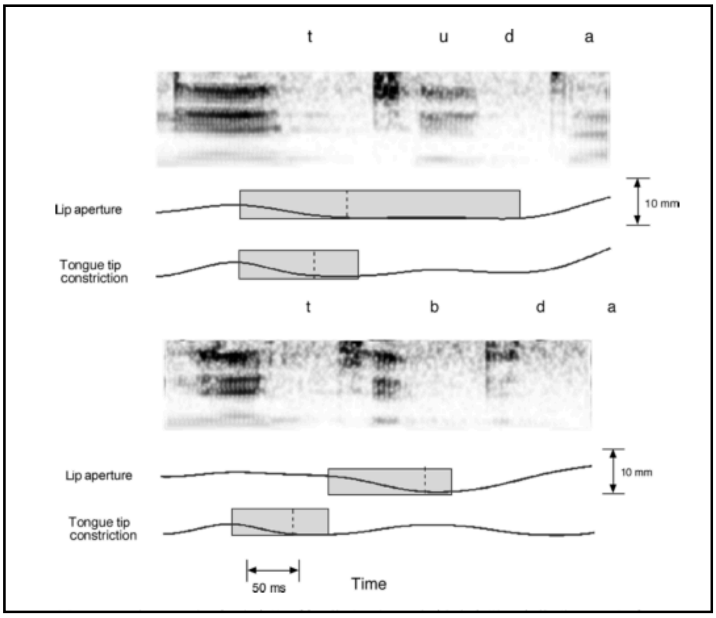


Figure 3.7: *Adapted from Goldstein et al. (2006). The upper panel shows an in-phase coordination between tongue tip gesture for /t/ in /tu.da/; while in the lower panel, in /tb.da/ where /b/ is a syllable nucleus in Berber Tashlhiyt, the tongue tip gesture and the lip gesture are triggered sequentially.*

3.3 Syllable structure, syllable and coupling hypothesis

The coupling hypothesis of syllable structure can account in articulatory terms for the differences between CV and VC structures generally captured by the traditional hierarchical syllable structure. First, it has been observed that CV syllables are preferred with respect to VC syllables, as shown by the fact that CV syllables occur in all human languages while VC syllables do not. Such distributional difference can be explained by the higher stability holding in the in-phase coordination pattern between onset and nucleus compared to the anti-phase relation between nucleus and coda. Second, in a V1CV2 sequence, the intervocalic consonant may be re-syllabified as the onset of the second syllable in running speech even if there is a syllable boundary between C and V2 ([V1C]σ[V2]σ), especially when speech rate increases (De Jong, 2001). This can be explained by the fact that the more stable in-phase mode tends to be automatically selected by the speakers, under perturbation conditions. Third, the asymmetric combinatorial strength between onset-nucleus relation and nucleus-coda relation is present even in languages in which the VC syllable structure is allowed. Onset consonant and nuclear vowel usually show a high degree of combinatorial freedom, whereas combinations of nuclear vowel and coda consonant are somehow restricted, that is, not all the consonants in a speech system could appear in syllable coda. Moreover, the number of the consonant clusters allowed in onset or in coda are even more restricted[22]. Goldstein *et al.* (2006: 231) propose that "coupling and phonological combinatoriality are related". The idea is that any two gestures can freely form a combination as long as they are coordinated synchronously "because of the [intrinsic] stability and availability of the in-phase mode" (Goldstein *et al.,* 2006: 231). Consequently, onset C gestures are able to combine freely with the following nuclear V gestures, which in turn, results in a stronger combinatorial strength for onset consonant clusters than the coda ones, in spite of the fact that in both cases the consonant gestures are claimed to be coordinated sequentially. Nam (2007) proposes two economic principles for understanding the development of different underlying gestural timing models observed in human languages:

i) maxmising the synchronous coupling;

ii) minimising the asynchronous coupling.

The in-phase mode (synchronous) has been shown to be the most stable pattern. Therefore, the phonological system of languages tends to maximise it. Since the synchronous pattern can always be accessed immediately without being acquired and with the least cognitive cost, it is easily to be maximised. For this reason, the asynchronous patterns, *e.g.*, the anti-phase mode and the competitive mode, are minimised by the system as a trade-off.

Summing up the observations from cross-linguistic findings on gestural timing model of syllable structures, it seems to suggest that the syllable, itself, may function as a phonological domain in Articulatory Phonology. The supporting evidences come

---

[22] For example, /l/ and /r/, when being the onset, could combine freely with the following nuclear vowel in Italian, but the combination of /l/ and /r/ is not a valid complex onset.

from two levels. First, the inter-gestural timing within a multiple-gesture consonant varies according to the consonant's position in a given syllable, as shown in Krakow (1993, 1999) for the English laterals and nasals. Second, to a large extent, languages may phonologise the relative activation timing between/among gestures (C gestures and V gestures) according to their position in a given syllable. In other words, it is the timing relation between C gestures and V gestures and between/among C gestures that shapes the syllable and its constituents in articulatory terms, which also corresponds to the view of Browman and Goldstein (1995), together with others (Vennemann, 1988; Dressler and Dziubalska-Kolaczyk, 1994; Ohala and Kawasaki-Fukumori, 1997; Ladd, 2014), that syllable structure is epiphenomenal with respect to gestural timing.

Chapter IV Coupling of Prosodic and Constriction Gestures

4.1 Traditional view in autosegmental-metrical theories on the basis of acoustics

In the autosegmental-metrical model of intonation, suprasegmental features are treated as timeless tonal events - H(igh) and L(ow), represented by relative target values of F0. The pitch movement, phonologically, is seen as the interpolation between tonal targets. A rising pitch movement (F0 excursion) can thus be thought of as the combination of two independent tonal events: L+H. Following the same reasoning, a falling pitch movement is represented in terms of the interpolation between an initial H and an ending L, *i.e.*, H+L. Then, on the basis of the acoustic evidence, a star symbol (*) can be assigned to either one of two tonal events to indicate that such tonal event is phonologically associated with the vowel of the accented syllable. Thus, the inventory of pitch accents is composed of six basic tonal morphemes: H*, L*, L+H*, L*+H, H+L* and H*+L.

Suprasegmental features represented by F0, such as tones or accents, belong to specific phonological domains in the segmental string. For instance, Chinese lexical tones go with syllables (or syllable rhymes), English pitch accents go with stressed syllables and Japanese word accents go with moras, etc. These relationships between suprasegmental and segmental levels are known as 'phonological association' in autosegmental-metrical (hereinafter AM) models of tones and intonations. However, the precise 'phonetic alignment' of F0 events with segmental events can be very complex, both within and across languages. Within the last two decades, the discrepancy between phonological association and phonetic alignment has become a hot issue in the AM framework of intonation. Such issue has been inspired by Ladd *et al.* (1999) to a large extent, because of his definition of the notion of 'segmental anchoring'. The term 'segmental anchoring' refers to a phenomenon in which both the beginning and the end of a given pitch movement seem to be independently aligned with identifiable points in the segmental string.

The first report of segmental anchoring came from a study on the alignment of the prenuclear rising pitch accent in Modern Greek carried out by Arvaniti, Ladd and Mennen (1998). They found that both the beginning (L) and the end (H) of the pitch rise are consistently aligned in time relative to specifiable landmarks in the segmental string. Specifically, the rising movement begins at the beginning of the stressed syllable and ends at the beginning of the post-tonic vowel. Such an alignment pattern remains surprisingly stable irrespective of the temporal distance between two alignment points. That is, the duration of the rising movement is a function of the temporal interval between the two segmental landmarks - the beginning of the stressed syllable and the beginning of the post-tonic vowel. Importantly, since the 'scaling' of F0 rise is a consequence of its duration, the 'slope' of the pitch excursion is determined by the distance between the two segmental landmarks. The following Figure gives an idealised diagram of the findings reported by Arvaniti *et al.* (1998).

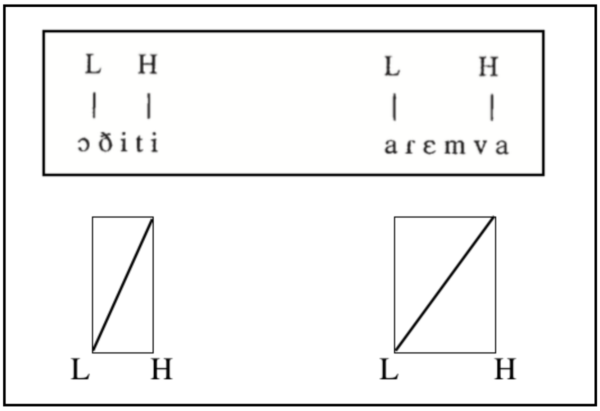


Figure 4.1: *Idealised diagrams of the segmental anchoring of Greek prenuclear bitonal rising pitch accent with different segmental materials. The left panel shows a short duration rise movement while in the right panel the rising movement spans over a longer sequence [ˈrɛmv], but the slope (interpolation between the F0 level at the beginning and at the end respectively) remains unaffected in both conditions.*

'Peak delay' is another important notion in the AM framework. The first studies on 'peak delay' were carried out on American English and Mexican Spanish by Silverman and Pierrehumbert (1990) and Prieto, Van Santen and Hieschberg (1995) respectively. They found that the peak, *i.e.*, the endpoint, in a rising pitch movement is aligned phonetically after the end of the stressed syllable with which the rising pitch movement is phonologically associated. In American English, for example, Silverman and Pierrehumbert (1990) showed that the phonetic alignment of the peak in rising pitch movement is strongly conditioned by the segmental distance to the end of the accented word or by the distance to the following stressed syllable. The greater the distance to the rightward word boundary and/or to the following stressed syllable, the later the peak is aligned in the accented syllable involved. Although segmental anchoring and peak delay are two phenomena of a different nature, they behave in similar ways from at least the following two perspectives: i) the beginning of the F0 rise happens always at the beginning of the accented syllable regardless of the alignment of the peak; ii) the beginning and the end of the pitch movement are independently aligned to the segmental string.

On the basis of these empirical findings and with the attempts to achieve a valid autosegmental interpretation, the 'secondary association' of tones was called for to describe specific patterns of alignment in AM model, which was first proposed by Pierrehumbert and Beckman (1988) to describe the initial L boundary tones in Tokyo Japanese. Figure 4.2 shows the hypothetical autosegmental representation of the notion of 'secondary association', in which the primary association concerns the phonological association between accented syllable in the segmental string and the pitch accent thought to be autosegments, whereas the secondary association is responsible for the 'alignment' of individual L and H tones to specific segments. Thus, the segmental anchoring of the Greek prenuclear accent L+H[23] could be interpreted in the following way: the primary association lies in the association between the accented syllable and the rising pitch accent L+H as a whole; additionally, as to the independent alignment of L and H, the L tone is secondarily associated with the left edge of the accented syllable and the H one with the following vowel.

---

[23] In the AM model, such pitch accent is transcribed as L+H*. The star notation is not adopted in the present dissertation because it is not closely related to the issues discussed here. L+H is used here simply for the purpose that it involves a rising pitch movement.

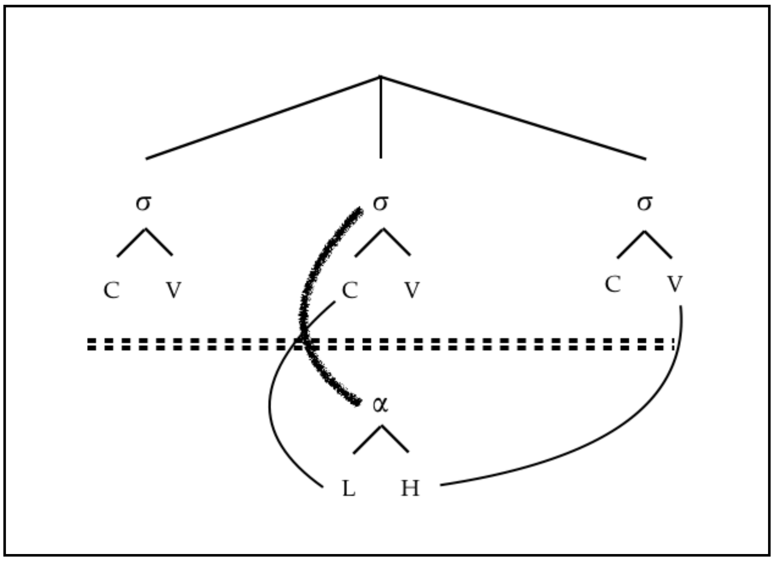


__Figure 4.2:__ Representation of secondary association, the accent L+H as a whole (α) is associated with the accented syllable (σ) as the primary association. The L and the H are associated secondarily with the left edge of the accented syllable and with the posttonic vowel respectively. Adapted from Ladd (2006).

Cross-linguistics evidence shows that the patterns of secondary association between individual tonal events and segmental landmarks can vary according to variations of the phonetic details of the alignment from language to language, or from variety to variety of the same language. For instance, English, Northern German and Southern German can be differentiated in the association of the initial L of the rising pitch accent with the different points found at the segmental level (Atterer and Ladd, 2004). L associates with the left edge of the accented syllable in English, with the onset consonant of the accented syllable in Northern German and with the left edge of accented vowel in Southern German. Similar differences are presented in the association of the H tone as well. The schematised alignment properties are illustrated in Figure 4.3. This behaviour seems to make less plausible the interpretation in terms of secondary association because it is almost impossible to categorise a common phonologically well-defined point in the segmental string for the association of the autosegments. Furthermore, the alignment patterns tend to be a continuum (from the left edge of the syllable to the left edge of the nucleus); and the earlier or later alignment of both the L and the H seems to assume that they are not entirely independent of one another. The alignment somehow applies to the pitch movement as a whole.

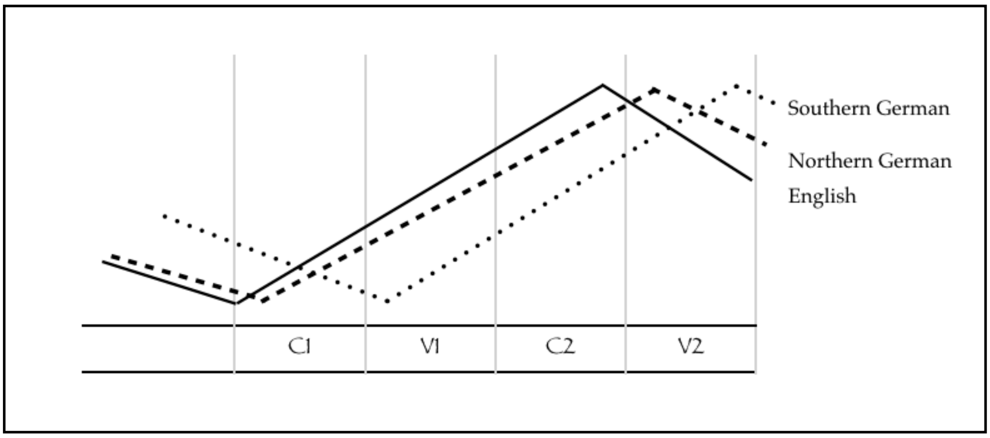


Figure 4.3: *Schematised alignment details found in English, Souther German and Northern German, based on the finding by Atterer and Ladd (2004).*

As confirmed by Ladd (2006), the failure to give an autosegmental interpretation of the segmental landmarks does not cast doubt on the phenomenon of segmental anchoring, but only on their segmental interpretation. Therefore, segmental anchoring might be a mere empirical finding, rather than a phonological hypothesis. Two aspects could further support such view. First, in almost all the languages studied so far, the duration of pitch movement is highly correlated with the temporal interval between the segmental landmarks whereas the amount of F0 excursion remains unchanged by the differences in segmental durations. Second, the precise alignment points can vary from language to language, and from variety to variety within the same language. Specifically, these variations may include: i) simply different phasing relations (*e.g.,* a certain F0 target is aligned earlier or later with the segmental string, as in Northern German vs. Southern German); ii) differences in duration (*e.g.,* in both Dutch and Greek the beginning of the rising movement is aligned in the same way, but the end of rise is later in Greek than in Dutch); iii) differences in the way in which alignment is affected by syllable structure (*e.g.,* in Chinese the end of a rise is always aligned with the end of the syllable, while in Spanish the end of a rise depends crucially on whether the involved syllable is closed).

Since the original autosegmental analysis fails to give a cross-linguistically valid interpretation on how F0 targets in a given pitch movement independently align with the segmental string, it is possible that the temporal coordination of gestures may help in explaining the phenomena. More specifically, if pitch movements are treated as articulatory gestures, then the phenomenon of segmental anchoring becomes a gestural timing problem. Under this assumption, the principles underlying gestural timing should be applied to inter-level phasing, that is, in the mapping of the suprasegmental level onto the segmental one. The control of pitch to achieve a given target in F0 space could thus be modelled as a tonal target, either for lexical tones (*e.g.*, lexical tones in Mandarin Chinese by Gao 2008, 2009) or for phrasal accents (*e.g.*, rising pitch accent 'L+H' in Catalan and in Viennese German by Mücke *et al.* 2012). The alignment differences could then be considered as phonological in nature in the sense that they emerge from differences in relative coupling relationships.

4.2 Pitch movements as articulatory gestures

A tone can be understood as a series of coordinated articulatory actions aimed to achieve a pragmatic goal; in this sense, it can be defined as a dynamic system - a tone gesture - in F0 space (Mücke *et al.,* 2012). Being an articulatory gesture, like all other constriction gestures, a tone gesture has to meet the following three requisites: i) it must be specified for a linguistic goal; ii) it must have intrinsic temporal activation interval per se so that it can evolve in time; iii) it can be coordinated with other articulatory gestures, of both tones and constriction actions. The task of a given tone gesture is to achieve linguistically relevant variations in the fundamental frequency of the vocal folds (see also Fougeron and Jun 1998, Katsika *et al*. 2014), represented by a tonal target in F0 space. Two invariant tone gestures are proposed - with a target at high pitch (H) and low pitch (L) respectively. In order to achieve these targets, the coordinated actions of a list of articulators are employed. The involved articulators are: the lungs, the trachea, the larynx, and a number of muscles such as thyroarytenoid, cricoarytenoid and cricothyroid muscles. At the inter-articulator coordination level, although the details of its functioning have not yet been well understood, it is clear that different sets of structures are responsible for increasing and decreasing the pitch level, *i.e.,* for high and low tone gestures respectively. In particular, in order to produce a lower pitch, parts of the thyroarytenoid muscles have to contract, shortening the vocal folds and making them thicker and heavier; for the production of higher pitches, the vocal folds must be stretched out to decrease their mass per unit length; to achieve this goal, the cricothyroid muscles must contract (for the details of the anatomy and physiology on the pitch control, see also Gick, Wilson and Derrick, 2013). Importantly, in dynamic terms, both types of articulator synergies control the same tract variable - the frequency of vibration of the vocal folds, but different dynamic parameters are assigned to either, namely, a low and a high level of fundamental frequency vibrations, respectively.

Unlike the AM theory, pitch movements such as rising or falling are not seen as the combination of two separate phonological events. Rather, they are considered as a single tone gesture. A rising pitch accent L+H, for instance, involves a 'tonal movement' towards an H target. Such movement is treated as a single tone gesture, for the achievement of a specific goal in the F0 space - H target. As for L, as a component of rising pitch accent it performs a dual role: it is the onset of the H tone gesture and the target of the previous L tone gesture. Since, by hypothesis, the tone gestures have extent in time, this implies an activation interval represented by the temporal latency from the onset to the target of the tone. The onset of a tone gesture is referred to as the temporal point at which the F0 begins to move in the direction of its target/goal. Therefore, in an L+H rise, L is the onset of the H tone gesture. Figure 4.4 gives an illustrative schematisation of the H tone gesture in terms of both AP and AM framework.

In the example illustrated below in Figure 4.4, the onset of H tone gesture coincides with the target/goal of the preceding L tone gesture. However, the onset of the L gesture remains unclear. Considering a tone as a gesture then allows to interpret tone gestural timing as constriction gestures within the inter-gestural timing model.

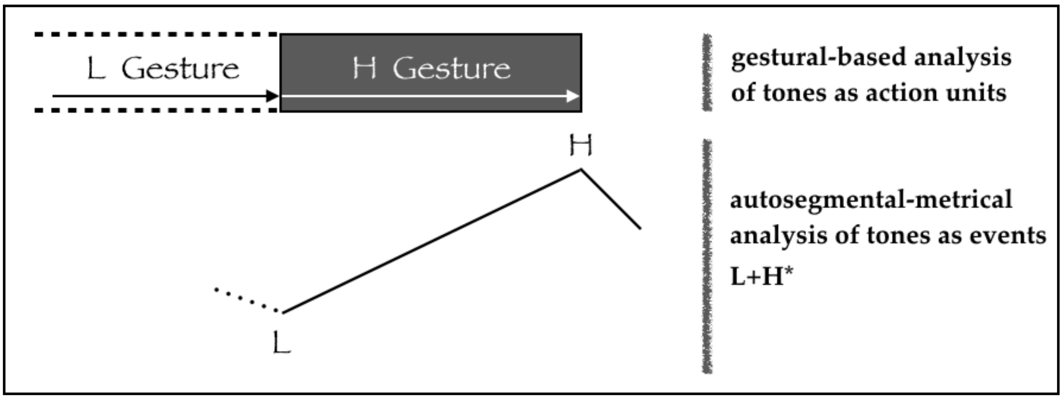


Figure 4.4: *Analysis of L+H rising pitch accent: Tone as gestural action units (above) vs. Tones as tonal events (below). Based on Mücke et al. (2012).*

4.2.1 Coupling of tone gestures in Mandarin Chinese lexical tones

At the inter-gestural coordination level, tone gestures are assumed to be coordinated with both constriction gestures and other tone gestures in either anti-phase or in-phase[24]. Gao (2008) first extended the coupled oscillator model to tone coordination in Mandarin Chinese. Based on empirical findings, Gao (2008) proposes that two invariant tone gestures - the H gesture and the L gesture - can be employed to represent four lexical tones in Mandarin. Tone 1 and Tone 3 are represented by a single tone gesture each, the H gesture and the L gesture respectively, whereas Tone 2 and Tone 4 are seen as combinations of two tone gestures. For Tone 4, the H gesture is followed by an L gesture, thus two tonal gestures are coordinated sequentially. For Tone 2, Gao (2008) supposes that there is an in-phase timing between an L gesture and an H gesture but with different activation interval - the L gesture is much shorter than the H one. A mid level of F0 is observed acoustically before the rise of the H tone. Such mid level F0 target is interpreted as the undershot of an L target due to the simultaneous activation of both the L gesture and the H gesture. Figure 4.5 shows the gestural scores for the

---

[24] It has not yet known clearly if tone gestures could activate one another simultaneously, even though Gao (2008, 2009) claims that the Tone 2 is the result of in-phase gestural timing between the L and the H gestures.

four tones in Mandarin Chinese. A similar interpretation of Tone 4 as two tone gestures simultaneously activated is however difficult to interpret in physical terms, because the gestural control regime for a lower target involves an increase of the mass per unit length of the vocal folds whereas for a higher pitch target the vocal folds have to be stretched and thinner (Gick *et al.*, 2013).

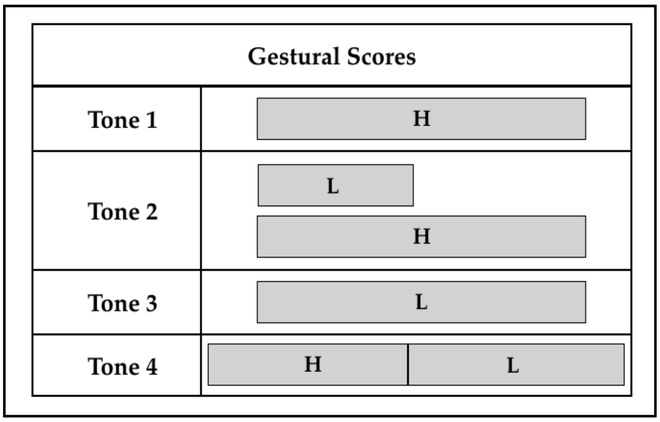


Figure 4.5. *Gestural scores for Mandarin Chinese lexical tones: Tone 1 and Tone 3 are represented by single tone gestures - the H and the L; Tone 4 is represented by two tone gestures sequentially coordinated - an H gesture is followed by an L gesture; while Tone 2 is represented as the simultaneous activation of two tones in which the activation interval of the L is intrinsically shorter than that of the H. Adapted from Gao (2009).*

Besides intra-level coordination, Gao (2008) also studied the inter-level coordination, namely, the coordination between tone gestures and constriction gestures. She showed that inter-level gestural timing is governed by the same coupling modes - simultaneous coordination (in-phase) and sequential coordination (anti-phase). Specifically, two patterns are found for Mandarin Chinese syllables. In Tone 1, Tone 2 and Tone 3, it is shown that the midpoint between consonant gesture (C) and tone gesture (T) co-occurs with the vowel gesture (V). Such pattern suggests a C-centre-like effect (see section 4.1). The T gesture behaves like an additional rightward C gesture, that is, the C and the T gestures are coordinated sequentially to each other and are both coupled synchronously with the V gesture at the same time (see Figure 4.6). As for Tone 4, the first tonal gesture (T1), corresponding to the H gesture, almost coincides with the onset of the V gesture. Being involved in an anti-phase coordination, the two tone gestures can be seen as two additional rightward C gestures sequentially coordinated. As a result, the T1 has to be shifted leftwards to make room for the T2, just as in the case of three-consonantal onset clusters (*e.g.*, split) in which all the three C gestures occur synchronously with the V gesture

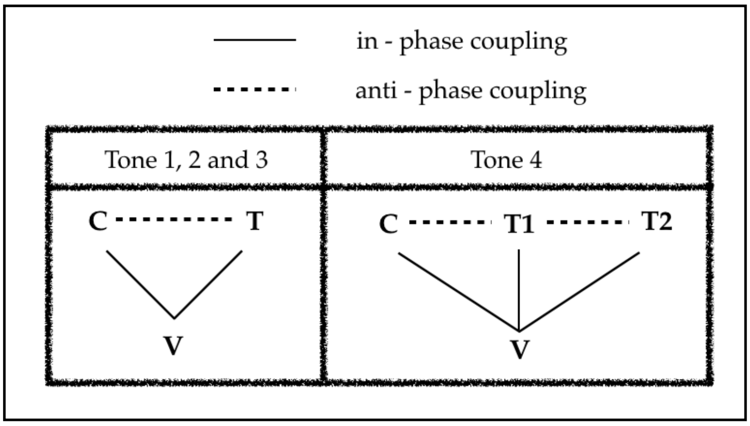


Figure 4.6: *Coupling graphs for Mandarin lexical tones with consonant and vowel. The left panel shows the coordination pattern for Tone 1, 2 and 3, the right panel for Tone 4. Although Tone 2 is interpreted by Gao (2008) as the result of the simultaneous activation of the L and the H gesture, it is represented as a whole by a T gesture in the left panel.*

Hsieh (2011) gives a totally different representation of Tone 3 (see Figure 3.14). Crucially, she proposes that the phonological representation of Tone 3 consists of a low (L) and a high (H) tone gestures, which coordinate independently from each other with the onset C gesture and the nuclear V gesture. The first tone gesture - the L gesture - occurs simultaneously with the V gesture and it is activated following the onset C gesture, thus assuming an in-phase gestural timing with the V and an anti-phase relationship with the C. The second tone gesture in Tone 3 - the H gesture - coordinates only with the V in the sequential fashion, like a coda C gesture. Surprisingly, these two tone gestures do not coordinate with each other. According to Hsieh (2011), this phonological representation can simultaneously capture the three context-dependent variations of Tone 3, namely, a) in isolation in which Tone 3 is fully realised; b) when preceding another tone (not Tone 3), it is realised as a low Tone 3 only (as the situation in Gao 2008, 2009); and c) when a Tone 3 is followed by another Tone 3, the former undergoes a tone Sandhi process giving a Tone 2.

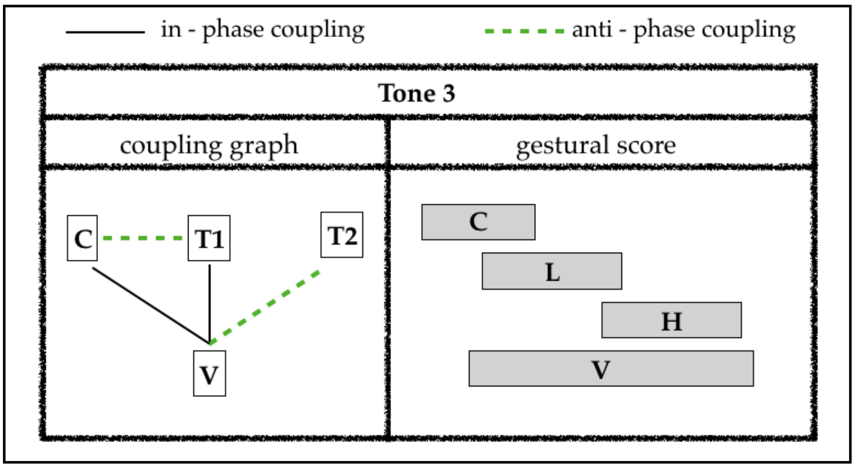


Figure 4.7: *Revised coupling graph and gestural score for the inter-gestural timing among Tone 3, the C and the V gestures in Mandarin Chinese, based on Hsieh (2011).*

In sum, both Gao (2008, 2009) and Hsieh (2011) provide first attempts to an inter-level gestural timing model for Mandarin Chinese syllables. Their results lead to an important implication, that is, the lexical tone gestures are able to affect the lexically specified onset-nuclear coordination. Given that lexical tone is part of mental representation of the word, the gestural score of the syllable involves an extra inter-level coordination, with lexical tone gestures interacting with either C gestures or V gesture at the segmental level. In particular, the coupling mode for the syllable with a Tone 4 provided by Gao (2008, 2009) shows that the onset C gesture is shifted leftward in order to leave the space for the coordination to the C-like sequentially coordinated T gestures.

Besides the attempts reviewed so far, the intra-level gestural coordination of Chinese complex lexical tones still remains as an open issue, the fundamental issue being that of understanding how the tract variable of tones - *i.e.*, the fundamental frequency of vocal fold vibration - can move towards two opposite dynamic goals simultaneously.

4.2.2 Coupling of tone gestures[25] in phrasal accent

Concerning non-tonal languages, a systematic account of inter-level coordination of pitch accents and constriction gestures is provided by Mücke and colleagues (Mücke *et al.* 2012). Specifically, the study concerns the different alignment behaviour of the rising pitch accent in Catalan and in Viennese German.

At the prosodic level, the rising pitch accent in both languages is represented by a tone sequence composed of an initial L tone gesture and a following H tone gesture. These two tone gestures are coordinated sequentially with each other, *i.e.,* in an anti-phase relation. At the level of the syllable, the onset consonant gesture[26] coordinates in-phase with the following gesture of the accented nuclear vowel. For intra-level inter-gestural timing, Catalan and Viennese German share the same patterns. As for the inter-level gestural coordination, however, two different coordination patterns are found, as shown in Figure 4.8 (a-b). The inter-level gestural scores obtained for Catalan assume an in-phase coordination between the H gesture and the nuclear V gesture because it can be observed that the onset of the H gesture occurs synchronous with that of the V gesture. By contrast, in Viennese German, both tone gestures coordinate in-phase with the V, thus suggesting a competitive coupling mode[27] between the prosodic gestures and the V gesture.

Figure 4.8 (a): *The coupling graph and gestural scores illustrate the gestural timing pattern found in Catalan in which the second tone gesture - the H gesture - of a tone sequence coordinates with the accented V in the way of in-phase, represented by solid line. Adapted from Mücke et al. (2012).*

---

[25] The terminology 'tone gesture', consistent with the term widely used in the literature, refers to the articulatory gesture used to achieve an articulatory goal in the F0 space, even for non-tonal languages.

[26] The stimuli in the study contain only the target words with simple onset consonant. Thus, there is no competitive coupling mode presupposed at the level of syllable.

[27] The competitive modes for inter-level gestural coordination is more in analogical sense than in nature. The timepoint at which both prosodic gestures and the V gesture are activated is not a c-centre in a strict sense because, by hypothesis, the onset of the first tone gesture - the L - in a rising pitch movement remains unspecified.

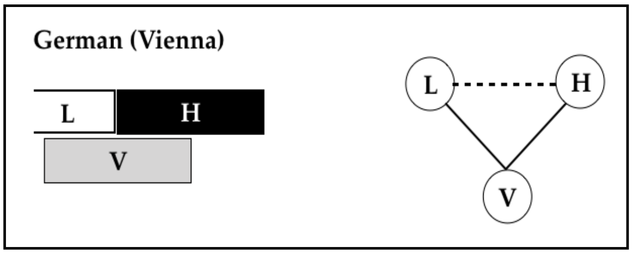


Figure 4.8 (b): The coupling graph and gestural score represent the scenario in German (Vienna), in which both tone gestures in a tone sequence coordinate with the V simultaneously giving a competitive coupling relation. In both languages, tone gestures in a tone sequence coordinate with each other in an anti-phase fashion, represented by dash line. Adapted from Mücke et al. (2012).

The coupling relations between prosodic gestures and constriction gestures in Catalan and Viennese German highlight a fundamental difference between the lexical tone gestures and the prosodic (post-lexical) tone gestures - whether they are able to alter the gestural timing between the onset C and the nuclear V. In Catalan, when a rising tone sequence is imposed onto a syllable, the H tone gesture coordinates with the accented nuclear V gesture in an in-phase fashion, and the gesture of onset consonant is synchronous with the nuclear V gesture. On the contrary, Mandarin lexical tone gestures can influence the C - V gestural timing (*e.g.,* in Tone 4) and appear to be fully integrated into the network of coupling relations within the syllable.

Viennese German is still different. With respect to the onset of the V gesture, the onset of the H tone gesture is supposed to be shifted rightwards due to the competitive relation between the two sequentially coordinated tone gestures. However, importantly, such a c-centre-like effect happens at the purely prosodic level, not involving any C gesture at the segmental level. The onset-nucleus gestural timing remains unaffected.

4.3 Boundary effects: π-gesture

Pitch accent and edge tone are terms traditionally used in the literatures corresponding to the pitch movements that are conventionally associated to stressed syllables of the prosodically prominent words and the words adjacent (pre-boundary and/or boundary-initial) to the phrasal boundaries (both intermediate phrase and intonation phrase) respectively. As introduced in the previous sections, the AP theories model the pitch accent as an articulatory gesture with an articulatory goal in the F0 space.

In the acoustic dimension, boundary-adjacent segments and syllables are shown to be longer both pre- and post-boundary (Shattuck-Hufnagel and Turk, 1998). As far as the articulatory dimension is concerned, the gestures that appear at boundary-initial and/or pre-boundary position are temporally longer and spatially larger, and show less inter-gestural overlap (Cho, 2004; Byrd and Saltzman, 2003). Additionally, in the temporal dimension the articulatory lengthening is found to increase with the boundary strength; in the spatial dimension, the lingual-palatal contact increases alongside the boundary strength increases (Fougeron and Keating, 1997; Keating *et al.*, 1999; Fougeron, 2001). Thus for the edge tones, Byrd and Saltzman (2003) propose that the prosodic boundaries, *i.e.*, the boundaries of intermediate phrase and intonation phrase, could be viewed as gesture as well, termed as the prosodic gesture - π-gesture. Similar to the constriction gesture, the π-gesture is endowed with the intrinsic temporal interval and it could thus extend over time. Unlikely the constriction gestures, the π-gesture does not directly instantiate a constriction target, but instead, it aims to slow the time course of the constriction gestures that are active at the same time (Byrd and Saltzman, 2003; Krivokapić, 2014; Katsika *et al.*, 2014).

These properties of π-gesture make the boundary-adjacent constriction gestures become temporally longer and spatially larger, and less overlapped between/among each other. A stronger boundary strength implies a stronger activation of the π-gesture which in turn results in stronger boundary effects (both in temporal and spatial dimensions). In addition, the π-gesture model allows the hierarchical prosodic structure to be simultaneously gradient and categorical, which is consistent with the basic assumption of AP. As discussed in Byrd (2006), it is structurally gradient due to the fact that the π-gesture has a continuum of activation strength values; while the categorical structure could be achieved by allowing a small number of attractors (*e.g.*, the different degrees of boundary strength) for the corresponding activation strength of the π-gesture (see also Krivokapić 2003 for a further discussion). The following figure gives an illustration that the boundary π-gesture co-occurs with the constriction gestures.

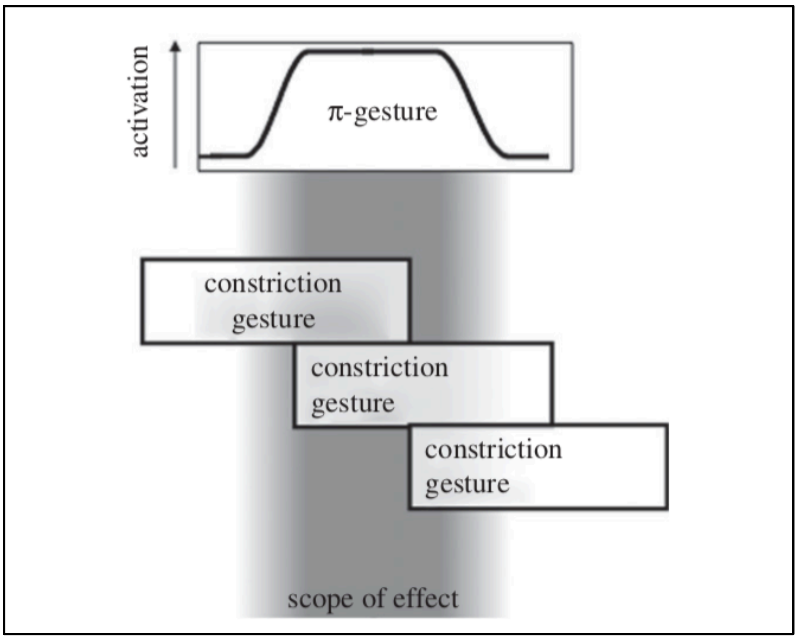


Figure 4.9: *adapted from Krivokapić (2014). A schematic illustration of the π-gesture co-occuring with the pre-boundary constriction gestures. The vertical arrow represents the activation strength of the π-gesture. The boxes indicate the activation interval of both constriction gestures and the π-gesture. The grey area shows the scope of the boundary lengthening effect, the darkness reflects the strength of its effect.*

The approach of π-gesture assumes a novel way of viewing the function of prosodic boundaries as inherently temporal (also spatial, perhaps). In consequence, it could make specific predictions on the behaviour of the boundary-adjacent constriction gestures. One of them might be that the boundary lengthening effects on the constriction gesture decrease with its distance from the prosodic boundary (intermediate phrase and/or intonational phrase), corresponding to the activation shape of the π-gesture, shown in Figure 4.9 (Krivokapić, 2014; Katsika *et al.*, 2014). And precisely for this reason, the stimuli designed for our experiment never occur in the boundary-adjacent position in the carrying sentence (see Chapter 6 for further details).

Chapter V Internal Temporal Organisation of Syllables: A Microscopic Perspective

5.1 Gestural overlap: from macroscopic perspectives to microscopic perspectives

At beginning, it has to be clarified that the term ‘gesture’ is used in the current dissertation with two different extensions. A first meaning refers to the gestures serving as the primitive units in the framework of Articulatory Phonology and produced for reaching a certain phonological/linguistic goal. They can further be specified with gestural attributes like constriction degree and constriction location, and they are equivalent to the tract variables. We have used 'gesture' with this specific meaning in the chapter 2. In addition to this, a second meaning refers to the constriction gestures that are implemented in the oral cavity, usually comprised within one jaw aperture-closure cycle. In this sense, the gesture coincides with a segmental unit like a C or a V, the gesture for [s] or the gesture for [a], for example. From now on, unless specified, the term ‘gesture’ is used with the second meaning, that of an oral constriction gesture.

It was suggested in Chapter 2 that the gestural score of a given utterance can provide a visual understanding on how linearly-ordered articulatory gestures are organised in both temporal and spatial dimensions; for instance, two adjacent gestures can be activated simultaneously (*e.g.*, an onset C gesture and a following nuclear V gesture), sequentially (*e.g.*, a nuclear V gesture and a coda C gesture) or somehow in an overlapping fashion (*e.g.*, C-centre effect for onset C clusters). In the current chapter, we will show that the differences in degree of inter-gestural overlap, illustrated through the gestural scores, might not only map to distinct phonological categorisations, but also correspond to different phonetic realisations, which could be further explained by the interactions between the macroscopic properties and microscopic properties of a given speech system.

5.1.1 Inter-gestural overlap between gestures belonging to distinct articulatory subsystems

Browman and Goldstein (2009) proposed a potential continuum, illustrated below in Figure 5.1, regarding the overlap of articulatory gestures belonging to the distinct subsystems involved in the articulation, ranging from complete overlap, partial overlap to minimal overlap. They argue that depending on the particular articulatory subsystems (oral, velic and glottal) involved, together with the different degrees of inter-gestural overlap, phonological categorisations could be made on the basis of these gestural combinations[28].

Figure 5.1(a) illustrates a case of complete gestural overlap of two articulatory gestures belonging to two distinct cavities. Both of them have approximately an identical temporal extent. In this case, the combination of two completely overlapped articulatory gestures is usually analysed as a single unit at the segmental level, for example, the gestural score for a nasal stop /n/ presented in the left column of the Fig.5.1(a) shows that the velic closure gesture occurs simultaneously and overlaps completely with the oral constriction gesture; similarly, a voiceless unaspirated stop /t/ (middle column) shows a total overlap between oral constriction gesture and glottal gesture. It should be noted that the examples in right colums of Figure 5.1 (a-c) will be discussed altogether below in the section 5.1.2. Noted also that the examples in all three right columns do not involve the gestural activation in glottal and velic subsystems because the voiced C gestures and/or the non-nasal C gestures have the default setting, they thus remain unspecified unless the voiceless C gestures and/or the nasals are concerned (see also gestural activation the /t/ in the middle columns and /n/ in the left column of Figure 5.1 (a) respectively).

---

[28] It is also called gestural constellation in the literature, for the details on gestural constellation, see Chapter 2.

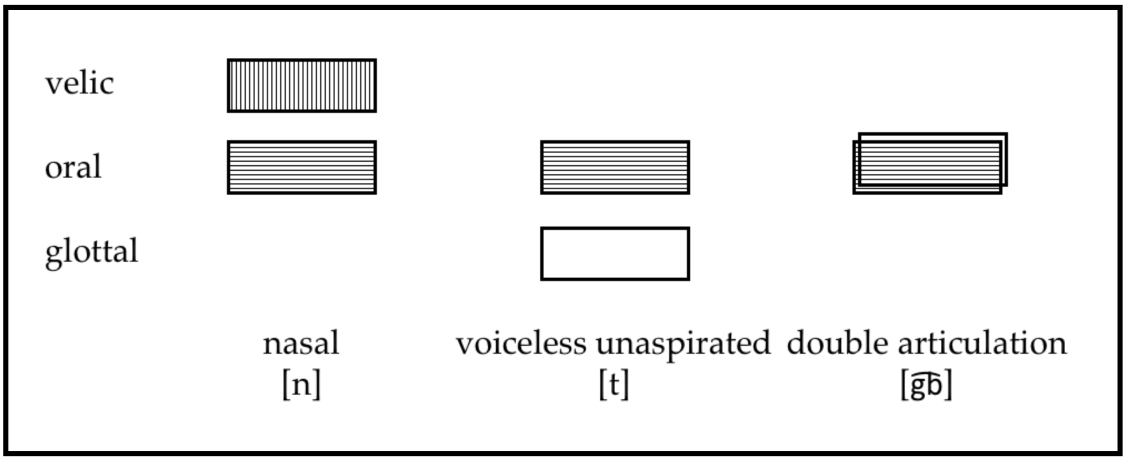


Figure 5.1(a): *based on Browman and Goldstein (2009). The illustration of the case in which two articulatory gestures from two distinct subsystems overlap completely (the left and middle ones). Each box represents the temporal activation of a given articulatory gesture, thus its temporal extent.*

Figure 5.1(b) shows instead the case in which one gesture is slightly moved away from the other in order to produce the effect of partial overlap between the articulatory gestures.

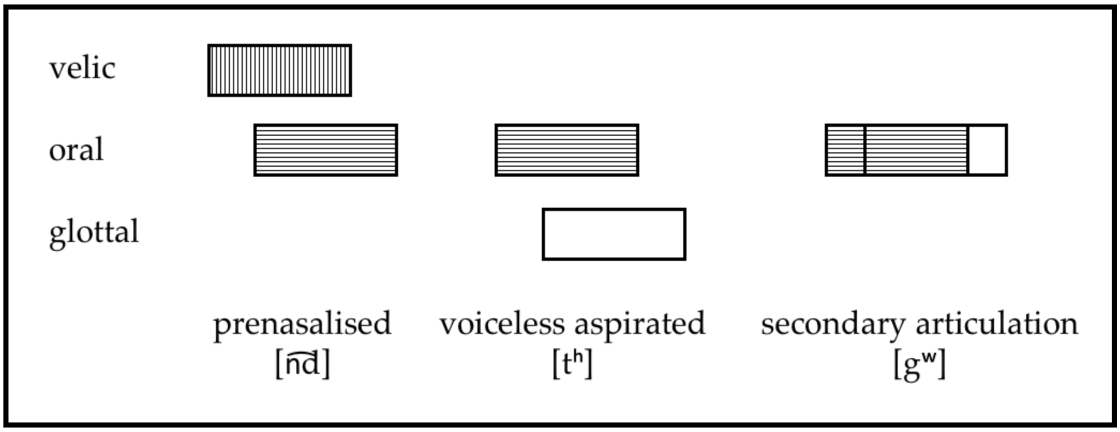


Figure 5.1(b): *based on Browman and Goldstein (2009). The illustration of the case in which two articulatory gestures from two distinct subsystems form a partial overlap (the left and middle ones). Each box represents the temporal activation of a given articulatory gesture, thus the temporal extent.*

Here, the inter-gestural temporal relations are no longer synchronous. The gestural score in the left column gives a direct contrast as opposed to the gestural score in the left column of Figure 5.1(a) — the difference between the prenasalised and nasal stops, from the gestural point of view, simply lies in the different overlap patterns between the articulatory gestures involved (Browman and Goldstein 1986, 2009). A similar gestural interpretation may hold for the difference between aspirated [t^h] and unaspirated [t], compare the middle columns of Figure 5.1(b) and Figure 5.1(a), respectively. In this analysis, phonological differences among segmental units are the natural consequences of different phasing patterns of the articulatory gestures involved.

If one of the gestures continues moving away from the other, two gestures would have a minimal overlap or even no overlap at all so that they might be activated in a sequential fashion. Figure 5.1(c) gives an illustration of this.

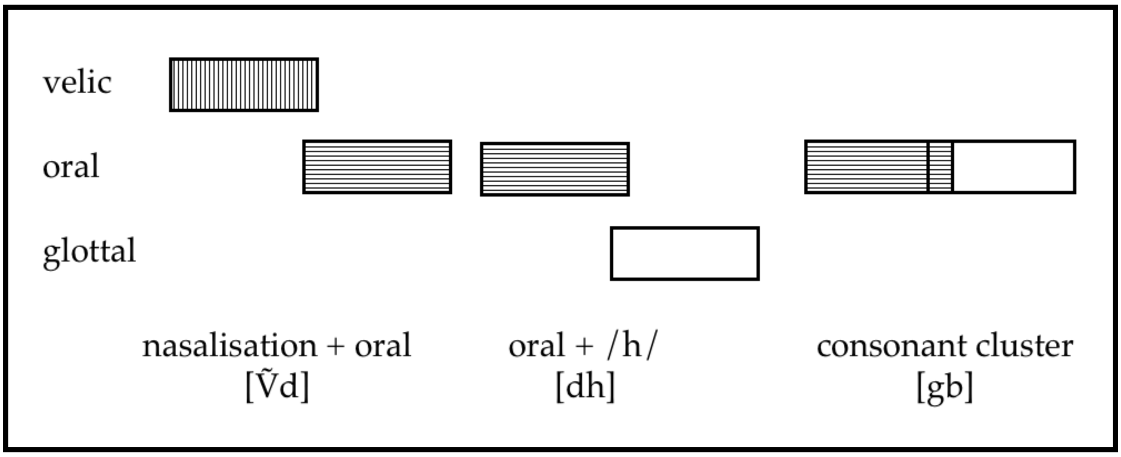


Figure 5.1(c)*: based on Browman and Goldstein (2009). The illustration of the case in which two articulatory gestures from two distinct subsystems form a minimal overlap (the left and middle ones). Each box represents the temporal activation of a given articulatory gesture, thus its temporal extent.*

Minimal overlap is often realised as consecutive activation of the articulatory gestures involved. In the left and middle columns of Figure 5.1(c), for example, the constriction gestures from two different articulatory subsystems — oral-velic and oral-glottal — simply occur one after the other according to a sequential activation order.

5.1.2 Inter-gestural overlap between gestures belonging to the same articulatory subsystem

The inter-gestural overlap between two gestures within the same articulatory subsystems (usually the oral subsystem) differs in various respects from the gestural overlap between two gestures from different articulatory subsystems. The overlap of two consecutive oral constriction gestures (especially for consonants) is relatively rare in the languages of the world and in fact there are several limitations to the overlap possibilities between two oral gestures (Maddieson, 1984). These limitations are caused by the microscopic properties of the individual speech system involved. The right columns of Figure 5.1 (a-c) showed three different cases of overlap between oral constriction gestures; for simplicity, they are represented below in Figure 5.2[29].

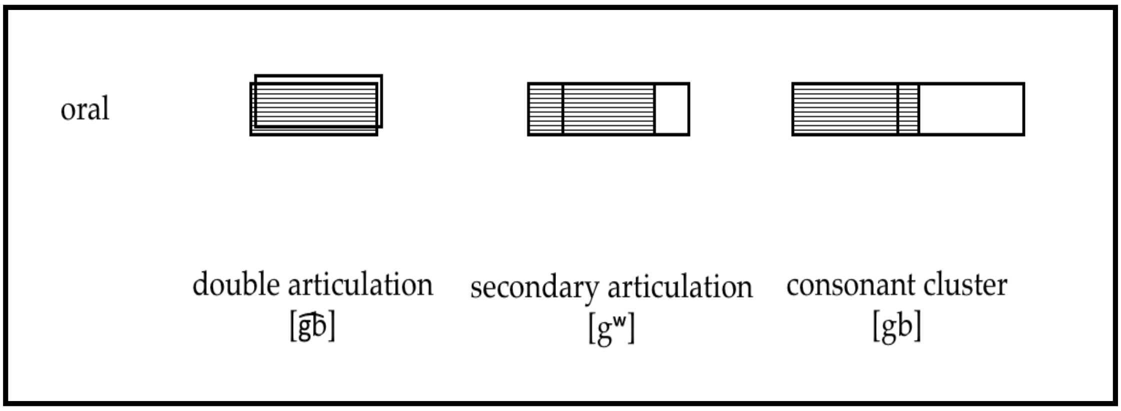


Figure 5.2*: based on Browman and Goldstein (2009). The illustration of the different overlapping cases — from left to right respectively, quasi synchronous, partial overlap and minimal overlap — in which two articulatory gestures belong to the same articulatory subsystem — oral subsystem. The shadow box in the figure represents a gesture with the tongue body constriction location (TBCL) at [velar] and the constriction degree (TBCD) specified for the value [closed], thus corresponding to a [g] in terms of segmental analysis, whereas the blank box indicates a bilabial gesture, thus specified with the TTCL at lips and TTCD specified as [closed], usually interpreted as a bilabial consonant [b] at the segmental level.*

---

[29] The absence of the articulatory gestures in glottal and velic subsystem is due to fact that for the voiced oral consonants the glottal and the velum are left unspecified by default.

The quasi-synchronous overlap between two oral constriction gestures gives a case of double articulation, shown in the left column. Double articulations occur in only 6% of the world languages according to the survey by Maddieson (1984), mostly restricted to the labial and velar constriction gestures. Alveolar and velar closure gestures[30] can also be synchronously activated, but this is associated with a different airstream mechanism which is often used to produce clicks (Traill, 1985, discussed in Browman and Goldstein, 2009).

The partial overlap between a velar stop and a labial closure gesture, giving a labialised stop, is shown in the middle column of Figure 5.2. Stops may partially overlap with a gesture whose constriction degree is wider than complete closure, *e.g.*, similar to frication.

These two types of oral constriction gestural overlap seem to correspond to the macro-microscopic interaction within a given speech system (see above, chapter 3). The macroscopic properties of the system allow two consecutive oral gestures being activated according to various patterns: quasi synchronous, partially overlapping and sequential; the microscopic constraints in turn constrain these overlapping patterns so that only a certain constriction gestures can occur quasi synchronously or with partial overlap.

Since the contrastive inter-gestural overlapping patterns concern the issues of activation order of consecutive articulatory gestures in different subsystems, they could also be understood as different coupling modes between two goal-oriented rhythmic and discrete movements: for example, complete overlap could be analysed as a sort of synchronous coordination, in the sense of an in-phase mode (*i.e.*, tight coordination), while the case of minimal overlap corresponds to the sequentially coordinated anti-phase coupling mode (*i.e.*, loose coordination). Nevertheless, the overall picture of inter-gestural coordination of partially overlapping gestures still remains an open issue and requires further investigation (see *e.g.*, Pouplier and Beňuš 2011 for a discussion of potentially unattested gestural timing patterns).

---

[30] In a research on the velar and alveolar (consonant) gestural overlap in French, Marchal (1987) reported that such airstream mechanism associated with the clicks production due to the velaric suction could be developped automatically as the overlap between consecutive velar and alveolar closure increases.

It could be further observed from Figure 5.2 that although the inter-gestural overlapping degrees form an articulatory continuum — from complete to minimal overlap, the same system of gestural overlapping options is able to explain qualitative phonological categorisations at the segmental level (Browman and Goldstein, 1986, 2009).

The right column of Figure 5.2 represents the case of minimal overlap between two oral constriction gestures, which suggests sequential activation and thus results in a consonant cluster at the segmental level. One might argue that consonant clusters in various syllabic positions differ for the degree of overlap between the oral gestures composing them, a hypothesis which is being thoroughly investigated in recent times. Moreover, gestural overlap patterns may differ across languages, which makes cross-linguistic investigation in this domain particularly urgent.

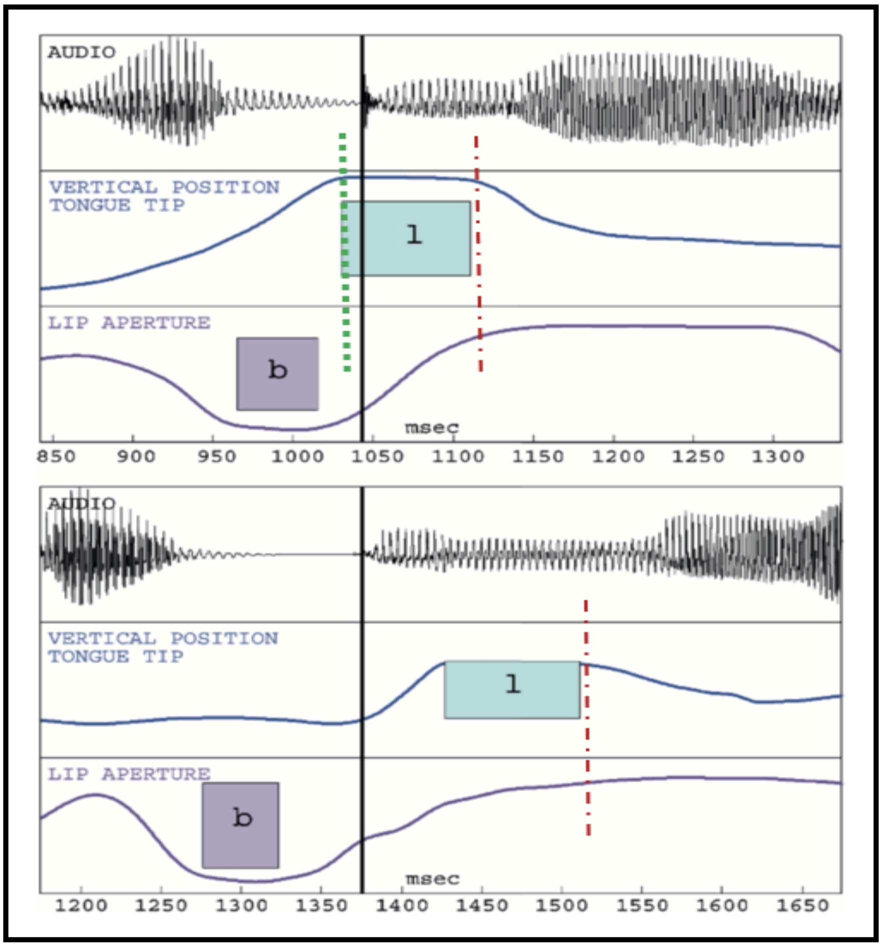


Figure 5.3: *extracted from Pouplier (2015). Illustration of the gestural timing difference between German and Georgian, the upper panel giving the acoustic and kinematic representations of /bl/ in German and the bottom panel*

giving the representation of /bl/ in Georgian. The boxes indicate the temporal interval in which the constrictions of both gestures reach their targets and remain stable.

Figure 5.3 gives an illustrative example for timing differences between Georgian and German for the syllable-initial consonant cluster /bl/ followed by the vowel /e/. In the case of German (upper panel), the tongue tip closure for /l/ occurs just after the accomplishment of the labial aperture gesture for /b/ when the distance between the lips begins to diminish, and the offset of the linguo-palatal closure almost corresponds to the bilabial closure (*i.e.*, the end of a jaw cycle) indicated by the dashed vertical line. The acoustic evidence shows that tongue tip closure assumes a quasi-synchronous relation — less than 10ms latency — with the burst release of /b/. On the contrary, the articulatory data for Georgian give a different inter-gestural temporal organisation, the linguo-alveolar closure happens after the release of the labial constriction. All these evidences suggest that the gestures within the onset consonant clusters are organised differently in German and in Georgian.

Apart from cross-linguistic variation, articulatory timing can also differ within a same language according to place and manner differences. The study of consonant clusters in syllable onset position in various languages over the past decades put the current gestural timing model on syllable structures in a dangerous place. In particular, the hypothesis of C-centre stability is claimed to be not valid for all the constriction gestures of a given individual speech system (Goldstein *et al.,* 2007 on Georgian and Berber; Marin, 2011; Marin and Pouplier 2010 on Romanian consonant clusters in syllable onset position; Hermes *et al.,* 2012 on Italian onset C cluster with *'s impura'*; Pouplier 2014; Brunner *et al.,* 2014 on word-initial and word-boundary consonant clusters in German; Pouplier 2015 on onset-nucleus timing in Polish). In fact, the definition of the C-centre effect (see Chapter 3) implies a sort of articulatory-acoustic interaction (in the sense that the V anchor is defined in terms of acoustics on the one hand, the C-centre is chosen based on articulatory evidences) which is supposed to remain stable with an increase of onset complexity. Pouplier (2015: 46) argues that the C-centre effect simply denotes the surface correlate of articulator timing that the competing underlying coupling graphs give rise to, in the sense that the C-centre provides the evidence for the onset being coordinated as a single entity to the vocalic nucleus – while the timing of each individual consonant in the onset to the vowel changes with the increase of onset complexity. Specifically, the C-centre hypothesis fails to explain the change of the inter-gestural structures inside the onset consonant cluster with the increase of its complexity, that is, how the degree of gestural overlap changes when a gesture is added to or removed from the onset.

Therefore, although the basic idea that syllable constituency can be expressed in terms of relative timing between consonant and vowel gestures is generally supported by a wide range of experimental evidence of the last two decades, it seems clear that the basic gestural timing model of the syllable constituency is not able to fully capture the differences and variations of onset-nucleus timing empirically reported in the literature on various languages and cluster types. One clear example is word-initial /sC/ clusters in Italian, in which the consonants of the branching onset have been found to be coordinated sequentially rather than simultaneously (Hermes *et al.,* 2008), as opposed to English where /s/+C clusters[31] behave more like simple onsets. Similarly, Marin (2013) reported that the order of the consonants in clusters interacts with the C-centre effect in Romanian: if the onset consonant clusters begin with the /s/ (*e.g.*, /sp/, /sm/ and /sk/), the C-centre pattern for the onset-vowel timing is found as predicted by the gestural timing model, but for the stop-initial clusters (*e.g.*, /ps/, /ks/, /kt/ and /kn/), the latency between the /s/ and the following vowel anchor /a/ does not decrease with respect to that in /salm/ - which means that no C-centre effect is observed. Marin (2013) attributed the lack of the C-centre effect to the initial consonant being a stop. These effects cited above are likely to suggest that there may be a complex relationship between hypothesised macroscopic physiological preferences and grammatical/linguistic patterns constrained by the micro-phontatics. As a matter of fact, the implausality of C-centre hypothesis on a cross-linguistically valid competitive organisation within the complex onsets further supports the view that in the gestural

---

[31] Waltl and Marin (2010) failed to find the C-centre effect for English word initial /spl/ and /skl/ consonant clusters. In their data, no rightward shift was found for the /l/ in case of *splayed* as compared to *played*. The authors attributed such unexpected consonant overlap pattern to the low lexical frequency of these clusters (/spl/ and /skl/) in English. However, to which extent and in which ways the lexical frequency may influence the onset-vowel timing is still unclear yet.

model of syllable structure the syllable is considered as an epiphenomenal self-organising structure resulting from inter-gestural temporal organisation conditioned by the language-specific phonotatics, which is also consistent with the early ideas developped by Vennemann (1988), Dressler and Dziubalska-Kolaczyk (1994) and Ohala and Kawasaki-Fukumori (1997).

Furthermore, as anticipated in Chapter 3, the Slovak syllabic consonants do not allow for the default onset-nucleus coordination patterns, since neither in-phase coordination nor C-centre effect are found; by contrast, the segmental identity of the nucleus - whether it is a vowel or a consonant - plays a crucial role. According to Pouplier and Beňuš (2010) there may be a casual link between the absence of in-phase onset-nucleus coorindation for syllabic consonants and the fact that the syllabic consonants with complex onsets are typologically rare (Pouplier and Beňuš 2011; Pouplier 2012). They additionally argue that it is unlikely that the system has a specific timing pattern exclusively for these cases since their frequency is very low. In fact, if this were the case, the cognitive cost for the speaker would increase to an out-of-control level. It is logically and empirically more convenient to hypothesize that apart from the language-specific phonotatic rules, inter-gestural timing patterns could be conditioned by the properties of the specific gestures involved.

Remember that in Articulatory Phonology, coarticulation is usually viewed as coproduction of adjacent or neighbouring articulatory gestures (Fowler, 1980; Fowler and Saltzman, 1993; Browman and Goldstein, 1993; Farnetani and Recasens, 1999; Recasens 2002); in other words, coarticulation is the result of inter-gestural overlap. Articulatory gestures in the speech system are doomed to reach their functional goal through the coordinative structures (Browman and Goldstein, 1988, 1992, 1993; Fowler and Goldstein, 2003). Gestures joining in the coordinative structures could hardly escape from the possible overlap between each other, and inter-gestural overlap reflects indeed their temporal coordination which is further expressed in terms of the model of inter-gestural relative phasing patterns. Under ideal circumstances, an increase in gestural overlap would lead to decrease of the measured gestural duration (since each articulatory gesture has its intrinsic temporal extent of activation: stiffness), and the amount of the coarticulatory effect will be increased. Therefore, coarticulation and coordination seem to two aspects of one same phenomenon — gestural overlap. Coordination is responsible for qualifying the gestures' temporal activation order, while coarticulation reflects the consequence of such differences on the activation order from the articulatory, acoustic and possibly perceptual aspects.

5.2 The DAC model as coarticulation measure

5.2.1 The degree of articulatory constraint (DAC) model

The DAC model was first put forward by Recasens and colleagues in the 1990s for the purpose of giving an empirical explanation on coarticulation. It was originally proposed in order to account for lingual coarticulation in speech. Lingual coarticulation is a complex phenomenon requiring not only collaboration of various tongue regions, but also the participation of different tongue muscles simply because in fact a certain tongue articulator cannot move freely without the activation of other tongue muscles which are responsible for other tongue regions. The DAC model intends to qualify the differences between consonants and vowels in how they are affected by the coarticulatory influences caused by adjacent segments and how they impose the coarticulatory effects on neighbouring phonetic segments. The degree of articulatory constraint is actually inspired by previous studies on coarticulatory resistance (CR) by Öhman (1966, 1967a, 1967b) and Bladon and Al-Bamerni (1976). The idea of CR lies in the fact that the phonetic segments may manifest more or less resistance to the coarticulatory effects induced by the adjacent or neighbouring segments depending on their intrinsic articulatory properties. Öhman (1966) had already shown that the coarticulatory freedom of Russian palatalised stops is very small and similarly Swedish and American English fricatives do not enjoy the same coarticulatory properties as the stops in the two languages. Bladon and Al-Bamerni (1976) reported that in British English, dark [ɫ] is less affected by the change of contextual vowels and consonants, with respect to clear [l]. Bladon and Nolan (1977) also show that apical [n] and [l] tend to become laminal when they are found adjacent to the laminal fricatives - [s] or [z], while the laminal fricatives seldom become apical. Alongside these empirical findings, the basic tenet of the DAC model is that the degree of coarticulatory resistance of a given phonetic segment can be defined by its own production requirements (Recasens 2014).

The DAC model enjoys several advantages, as claimed by Recasens (2014), over other models or theories on coarticulation in several aspects. First of all, the model assigns the values of articulatory constraints to all consonants and all vowels, instead of a subset of the segments, in the phonetic inventory of a given language. Such language-specific hierarchy for coarticulatory resistance is further able to make powerful prediction on the outcomes of sound change (*e.g.*, insertion, elision, assimilation, *etc.*) on both synchronic and diachronic dimensions.

Secondly, the model acknowledges that the coarticulatory outcomes are conditioned not only by the locations in the vocal tract in which the closure or the constriction is found, but also by the manner of articulatory requirements. For example, the manner requirement for the production of the alveolar fricative [s] could control the formation of a narrow medial lingual groove for the passage of the airflow and the generation of the turbulent noise when the airflow impinges onto the upper teeth. It is the manner of articulation controlled by the specific aerodynamic conditions that restricts the coarticulatory effects exerted by the adjacent segments on fricatives.

Thirdly, the studies on coarticulation have shown that some vowels and consonants usually trigger coarticulatory effects on the previous segments (anticipatory effects/regressive coarticulation), whereas others on the following segments (carryover effects/progressive coarticulation). In fact, these two sorts of coarticulatory effects do not share the same mechanism. Anticipatory effects reflect the planning of the upcoming phonemic units ahead of time, whereas carryover effects result from the mechanical-inertial properties of the articulatory structures. The DAC model, in this respect, is able to specify the pattern of coarticulatory directionality given that according to the DAC model, the spatial and/or temporal prominence of the anticipatory and carryover effects induced by a given vowel or consonant gesture is defined by its production demand. For instance, apical consonants are usually anticipated during the production of the preceding segments whereas dorsal consonants manifest coarticulatory effects on both the preceding and the following segments. This is presumably consistent with the different physiological and biomechanical characteristics shown by the tongue tip and the tongue dorsum.

It should be borne in mind that in the DAC model each consonant and vowel in a given speech system is explicitly specified for the degree of articulatory constraint based on their production requirements. Therefore, the alveolar fricative [s] and the alveolar trill [r] might be attributed a rather high value of coarticulatory resistance due to the precise manner of production and the fine control of aerodynamic conditions during the production (see also Ohala and Solé 2010 on /s/; Proctor 2011, Solé 2012 and Celata *et al*. 2013 on /r/). On the contrary, the central vowel *schwa* is considered as the most unconstraint phonetic segment given that the tongue body remains inactive without a specific target that further define the direction of the lingual movement while producing an [ə]; however, the degree of articulatory constraint of the segments at the phonetic level is a language specific issue, rather than a universal one.

5.2.2 Articulatory constraints for vowels and consonants

5.2.2.1 Vowels

Vowels in the physical space of the oral cavity can be classified along two dimensions - vertical and horizontal - the former is responsible for the height-low contrast and the latter for the front-back contrast. The horizontal dimension can also be specified according to constrictions in different locations in the oral cavity (Wood 1979, Recasens 2014), as in table 5.1:

	palatal	upper pharyngeal	lower pharyngeal	velar
front	[i, e, ɛ]			
back		[o, ɔ]		[u]
low			[a, ɑ]	

Table 5.1: *classification of vowels in terms of places of articulation and tongue height, based on Recasens (2012, 2014).*

For the front vowels, the constriction location happens at the palatal region which requires a high degree of constraint of tongue dorsum given that the raising movement of the front dorsum participates actively in the articulation of front vowels. The three front vowels have three rounded cognates [y, ø, œ], characterised by an additional high degree of constraint at the lips. Moreover, vowels sharing the same constriction location can be distinguish from one other by the articulatory constraint depending on the aperture of the constriction - the narrower the constriction, the more constrained is the vowel. Thus, among the front vowels that share the same constriction location in the palatal region using the tongue dorsum (or predorsum), [i] shows the highest degree of articulatory constraint, [ɛ] the lowest, and [e] an intermediate one. In the DAC model, then, 'constrained' means that there is more constriction of the tongue dorsum.

5.2.2.2 Consonants

Consonants in the DAC model[32] (Recasens 1999, 2014, Farnetani and Recasens 1999) are represented in terms of constriction location and manner of articulation (see Table 5.2).

---

[32] Since the DAC model concerns the lingual coarticulation, the consonants whose primary articulator is not found in the tongue are not listed in the table, for example, the uvulars [ʁ, ɹ].

	stops	fricatives	affricates	nasals	laterals	tap	trill	approximants
labials	[p,b]	[f, v]		[m]				[β]
dentals	[t, d]	[θ]						
alveolars		[s, z]		[n]	[l]	[ɾ]	[r]	
(alveolo-) palatals	[c, ɟ]	[ʃ, ʒ]	[tʃ, dʒ]	[ɲ]	[ʎ]			[j]
velars	[k, g]	[x]		[ŋ]	[ɣ]			
labio-velar								[w]

Table 5.2: *classification of consonants in terms of places of articulation and manners, based on Recasens (2012, 2014).*

The degree of articulatory constraints of the consonants depends on the active participation of the lips and/or a specific lingual region in the formation of a closure or a constriction, as well as on manner requirements. Therefore, labials are maximally constrained at the lips, and exhibit a very low degree of constraint for the tongue body since the tongue body configuration plays no significant role in the formation of the closure at the lips. Among the lingual consonants, the (alveolo-) palatals and the velars require active participation of the tongue dorsum, therefore these consonants exhibit a higher degree of articulatory constraint than dentals and alveolars. However, alveolar fricatives [s, ʃ] and the alveolar trill [r] must be specified as highly constrained consonants due to the strict aerodynamic and articulatory requirements. As for laterals, clear [l] requires the lowering of the predorsum for the airflow to pass through the lateral passages in the oral cavity, whereas the dark [ɫ] is articulated with the postdorsum backing towards the pharyngeal wall and lowering of the predorsum. As a result, the predorsum is more constrained in the production of the dark variant. The labiovelar approximant [w] is articulated with the constriction at two different locations - the lips and the velar zone respectively. Therefore, it is also regarded as a highly constrained consonant.

5.2.2.3 The DAC scale and coarticulatory sensitivity

According to Recasens (2014), the more constrained a segment is, the more resistant to the coarticulatory influence of other segments will be. In the original version of the DAC model, vowels and consonants are assigned different DAC values depending on the degree of tongue dorsum constraints: 1 to bilabials and the *schwa*; 3 to the (alveolo-) palatal consonants, vowel (except for the low vowel [a]), the velars, the dark variety of the lateral [ɫ], the alveolar fricatives [s, ʃ] and the trill [r]; 2 to the dentals, alveolars (except for the dark [ɫ] and the trill [r]) and the low vowel [a].

Recent studies (Recasens and Espinosa, 2010; Recasens, 2016) on coarticulation in V_1CV_2 sequences suggest that the current scale still needs to be refined in several respects, e.g. by introducing some intermediate values. For instance, among segments which are assigned the DAC value 3, the (alveolo-) palatals appear to be more constrained than the vowel at the tongue dorsum, and less constrained than the alveolar fricative [s] and the alveolar trill [r]. In addition, velars exhibit some hybrid characteristics, although the tongue dorsum is highly constrained in the articulation, the configuration of the tongue can adapt to the vocalic context along the horizontal dimension, while along the vertical dimension which regards the tongue backwards movement, the velars seem rather resistant to adjacent vowel effects (Zharkova 2007, Chen *et al.* 2017). Thus, velars remain problematic for the current DAC scale.

In a recent version of DAC scale, consonants and vowels are attributed five different values (Recasens 2014). The 0 value is assigned to the *schwa* and the labial consonants as no active lingual involvement during the articulation. The value of 1 is assigned to the back vowels [a, u, o] due to the tongue dorsum and postdorsum coupling. The value 2 characterises a moderate tongue dorsum raising during the production, for example, the [e] and the dental consonants and alveolars (except [s] and the trill [r]). The DAC value 3 is assigned to the phonetic segments with the maximal tongue dorsum raising, like [i] and the (alveolo-) palatals. The maximal DAC value is assigned to the fricatives [s] and [ʃ], the dark variety [ɫ] and the trill [r], apparently due to their strict manner requirements and precise aerodynamic conditions during the articulation. A simple illustration of the updated DAC scale is provided in the following Table 5.3:

DAC Value	Vowels	Consonants
4		[s], [ʃ], [ɫ] and [r]
3	[i]	the (alveolo-) palatals
2	[e]	the dentals, the alveolars
1	[a], [o] and [u]	
0	*schwa*	the labials

Table 5.3: *five-degree articulatory constraints, based on Recasens (2014).*

Once the DAC scale is established, the model is able to predict the degree of coarticulatory sensitivity of phonetic segments. The degree of coarticulatory sensitivity for a given vowel or consonant varies inversely with its DAC value. The coarticulatory sensitivity is manifested by two aspects: coarticulatory resistance (CR) and the coarticulatory aggressiveness. The former is responsible for the extent to which a phonetic segment allows the coarticulatory influences from the adjacent segments, while the latter concerns the extent to which a segment affects the articulatory configuration of the neighbouring segments. The DAC model suggests that CR and coarticulatory aggressiveness are associated with the degree of the articulatory constraint of a given segment. Therefore, the more constrained a phonetic segment, i.e., the higher its DAC value, the more it is immune from the coarticulatory effects of other segments and exerts coarticulatory influences on them. On the contrary, a lower DAC value indicates that the segment is less resistant to the influence of other segments and less aggressive towards them.

5.2.3 Coarticulatory directionality

Coarticulatory directionality means that phonetic segments may favour anticipatory or carryover coarticulatory effects. The prevalence of either one or both coarticulatory directions is closely associated with the articulatory properties of the triggering phonetic segment. For instance, for the syllable-final dark [ɫ] in English, tongue dorsum lowering and retraction often occur before predorsum raising[33] (Sproat and Fujimura 1993); as a consequence, dark [ɫ] favours anticipatory C-to-V coarticulation rather than carryover.

On the other hand, C-to-V effects in VCV sequences with (alveolo-) palatal are often more prominent at the carryover level than at the anticipatory level, since the tongue dorsum gestures for these consonants usually overlap with the following vowel instead of overlapping with the preceding vowel (*e.g.*, the Catalan sequence [aɲa] in Recasens and Espinosa 2010). Thus, a(n) (alveolo-) palatal consonant, like the Catalan [ɲ], is expected to largely prevent the regressive (anticipatory) coarticulatory effects induced by the following vowel and to allow more progressive (carryover) articulatory effects both on the following vowel and from the preceding vowel (Recasens, 2014).

In a V_1CV_2 sequence, C-to-V_1 anticipatory effects might interfere with V_1-dependent carryover effects, while C-to-V_2 carryover might conflict with the V_2-dependent anticipatory coarticulatory effects. In the case where both anticipatory and carryover effects are present, which coarticulatory direction will be the dominant one? The question must be established on empirical grounds. Data have shown that consonants favouring C-to-V anticipatory coarticulation tend to largely block the carryover effects exerted by the preceding vowel; on the contrary, the consonants favouring carryover coarticulatory effects on the following vowel should prevent the anticipatory effects induced by the following vowel (Recasens 2014).

---

[33] Here we use the term 'motion' instead of 'gesture'. The /l/ is considered by many as a multi-gesture segment — composed of a pre-dorsal raising gesture and a dorsal retraction gesture (Giles and Moll, 1975). Crucially, the difference between a clear [l] and a dark [ɫ] properly lies in the different gesture timing observed for two variants (Sproat and Fujimura, 1993; Zharkova, 2007; Scobbi and Pouplier, 2010). However, Recansens (2007) argues that the /l/ is a mono-gestural segment, the dorsal retraction is actually the consequence by the motion of the raise of the pre-dorsum, therefore it is a sort of secondary gesture without the autonomous identity.

Once again, the DAC value for consonants increases with the active participation of the lips and/or a specific lingual region involved in the formation of closure or constriction, and also with the particular manner of articulatory requirements (*e.g.*, /s/ or /ʃ/), therefore, all else being equal, the coarticulatory effects of consonants are usually stronger than those of vowels. In conclusion, the vowel-dependent coarticulatory direction appears to be more or less conditioned by the coarticulatory directionality of the adjacent consonants.

5.3 Macrosopic preferred syllable timing and microscopic relative inter-gestural timing

It has been already argued in the section 5.1 that the two most pervasive inter-gestural coordination patterns – the synchronous and sequential modes - roughly correspond to the way in which two goal-oriented articulatory gestures overlap: complete overlap and minimal overlap respectively, with the *syllable* as the primary domain. Meanwhile, cross-linguistic studies in the last two decades failed to find a universal gestural timing hypothesis for the word-initial consonant clusters (Hermes, 2012; Hermes *et al.,* 2008 for Italian word-initial sC clusters; Pouplier, 2012; Marin, 2010, 2013 for Romanian onset clusters; Brunner *et al*. 2012 for German onset and cross word boundary clusters; Pouplier, 2015 for Polish data). Although the Romanian data reveal that the inter-gestural coordination pattern could even show within-language variations (*i.e.,* beyond C-centre effect in complex onsets) based on the type of consonants (Marin 2010, 2011, 2013), such consonant-induced variations seem to suggest a different view on the biologically/physiologically-constrained (*e.g.,* the correspondence between the synchronous in-phase mode and a jaw aperture-closure cycle) gestural timing modes. That is, the articulatory properties of the segments that participate in a coupling relation are able to condition the inter-gestural timing. In particular, studies on the degree of articulatory constraints (Bladon and Al-Bamerni, 1976; Recasens, 1999, 2006, 2014; Farnetani, 1989; Farnetani and Recasens, 1999, 2010; Recasens *et al.,* 1995, 1997) –

resistance to coarticulation – showed that coarticulatory effects could contribute to the inter-gestural overlap in the spatial dimension.

Following these empirical studies, we thus suggest that the variations in the inter-gestural phasing relations, due to the different articulatory properties are not controversial to the current gestural-timing based syllable structure model. As already introduced in Chapter 2, speech systems are similar to self-organised physical systems, they share the same theoretical descriptions of being simultaneously discrete and continuous, having both macroscopic and microscopic properties which are further mutually constrained. At the syllable level, a given speech system could have its macroscopic inter-gestural timing patterns between syllable constituent on the one hand, and, on the other hand, the microscopic properties of the system, like the articulatory constraints of the constriction gestures allowed in the sound inventory of the speech system, expressed in terms of physiological, biological and aerodynamic characteristics, are able to feed back the entire system and finally contribute to the variations in superficial inter-gestural temporal coordination. On the practical ground, once the DAC model on coarticulation is integrated with the inter-gestural timing model on coordination, the evidences of inter-gestural timing variations could be explained in terms of the contextual re-adaption of speech motor control system as conditioned by the individual languages. In other words, there might be automatical[34] trade-off (which means no changes at the level of plan) in speech motor control, which could lead to a contextual re-definition of the so-called preferred inter-gestural coordination patterns without the necessary categorisation of a new inter-gestural phasing relation at the macroscopic level of the speech system, especially when obstruent clusters are concerned (Fowler and Saltzman, 1993; Farnetani and Recasens, 1999; Marin 2011, 2013; Pouplier, 2012, 2014, 2015; Pastätter and Pouplier, 2014, 2015). This would in turn definitely enlarge our understanding on what actually takes place behind the measured inter-gestural coordination relation.

---

[34] Fowler and Saltzman (1993) claims, when discussing the gestural blending, that no changes may be needed at the plan level and the articulatory consequences of gestural overlap would be shaped by the manner in which the coarticulating gesture blends its influence on the vocal tract with those of an ongoing one.

Chapter VI Experiment: from Data Acquisition to Data Analysis

In the current dissertation, a production experiment was carried out with the purpose of investigating the coordination pattern of Tuscan Italian for various types of syllable constructions in distinct prosodic conditions and with consonants differing for place of articulation and coarticulatory resistance (*i.e.*, /s, l, k/). Tongue mid-sagittal configuration and linguo-palatal contacts are captured by the SychroLing system, an innovative, integrated system for multi-level articulatory and acoustic data acquisition and analysis. This chapter gives a description of the experimental setting, the corpus design, the procedures used for data acquisition, acoustic annotation, prosodic transcription and articulatory measurements.

6.1 Speech materials

An important criterion in the selection of the stimuli for the present study is the possibility of capturing both segmental and prosodic variations in speech production. In such a way, various hypotheses could be evaluated using a single set of data with consistent experimental techniques. The target words contained both real Italian words and pseudo words. The speech materials are given in Table 6.1 (a-b) below. The words listed in the table are the target words that were inserted in two different carrier sentences. The carrier sentences were made as in the following examples:

a) Pronuncia SABA molte volte.
(He/she pronounces 'saba' many times.)
b) A: Pronuncia seba?
(A: Does he/she pronounce seba?)
B: No, pronuncia SABA molte volte!

(B: No, he/she pronounces ‘saba’ many times!)

target words - singletons						
	/s/		/l/		/k/	
	CV	VC	CV	VC	CV	VC
/a/	saba /ˈsaba/	bass /ˈbas/	laba /ˈlaba/	bal /ˈbal/	capa /ˈkapa/	pac /ˈpak/
/i/	siba /ˈsiba/	bis /ˈbis/	liba /ˈliba/	bill /ˈbil/	kipa /ˈkipa/	pic /ˈpik/

Table 6.1(a): *Experimental stimuli used for the study on singleton onsets and codas. For each stimulus, the orthographic and the phonetic transcriptions are given.*

target words - clusters						
/lk/	/lt/	/ks/	/sk/		/st/	
CVCC	CVCC	CVCC	CVCC	CVCCV	CVCC	CVCCV
balk /ˈbalk/	balt /ˈbalt/	pax /ˈpaks/	bask /ˈbask/	basca /ˈbas.ka/	bast /ˈbast/	basta /ˈbas.ta/

Table 6.1(b): *Experimental stimuli used for the study on clusters. For each stimulus, the orthographic and the phonetic transcriptions are given.*

The corpus contains five dimensions of variation. The first one is the contrast between onset versus coda positions.

Second, we contrasted consonants with different places of articulation. According to the DAC model introduced in the previous chapter, /k/ is expected to show the least degree of coarticulatory resistance, /s/ the largest, and /l/ an intermediate one. The reasons why we expect /s/ to be the most, /k/ the least and /l/ the intermediate resistant consonant are the following. Coarticulatory resistance (*e.g.*, Recasens, Pallarès and Fontdevila, 1997; Recasens and Espinosa, 2009) is known to depend on several factors, such as whether a given lingual region is involved in the constriction, jaw height, and the severity of the manner of articulation requirements. With respect to jaw height, /s/ having a higher jaw is known to be more resistant to coarticulation than both /l/ and /k/ (Recasens, 2012). Frication with turbulence is also the most demanding manner of articulation, which reinforces the view that /s/ should show a comparatively higher degree of coarticulatory resistance. However, /l/ as a lateral continuant also has more demanding requirements than stops. In addition, the places of articulation of dorsal /k/ and /g/ are known to vary to a very large extent with vowel fronting (since Öhman, 1966; see also Fowler and Brancazio, 2000).

Third, singleton consonants could occur adjacent to /a/ or adjacent to /i/. This was done to empirically test the degree of coarticulatory resistance of the three consonants.

Fourth, the prosodic structure of the sentence in which the target words were inserted also varied, with sentences with a broad focus contrasting with sentences with a corrective focus[35]. The prosodic contrast on the target words had a twofold function: on the one hand, it served to measure potential prosodic effects on coarticulation (Cho, 1992 on English; Farnetani, 1989; Farnetani and Kori, 1986; Avesani, Vayra and Zmarich, 2009 on Italian); on the other hand, it was used to derive preliminary indications on how Tuscany Italian organises its inter-level (prosodic-articulatory) gestural coordination (Mücke *et al.* 2012; Gao 2008, 2009; Krivokapic 2014, 2016).

Fifth, some coda consonant clusters were also taken into consideration. They were used to determine potential differences between coda singletons and coda consonants followed by another consonant, and also to investigate the directionality of the coarticulatory influences between adjacent consonants.

6.2 Participants

Four native speakers of Tuscany Italian participated in the experiment, one male and three females. Each speaker was recorded on a different session.

---

[35] It is also called contrastive focus in some works because the information is often new and contrast with the previous one and usually realised via the intonation with the relative prominence. However, in a strict sense, it should be the corrective focus although a contrastive prosodic prominence is expected as well. See also Vallduvì (1990) for a description on the typology of focus and the distinction between the corrective and contrastive focus.

6.3 Experimental setting

6.3.1 The *SynchroLing* system

The collection of acoustic and articulatory data used an innovative, integrated and self-acquisition system called *SynchroLing*, elaborated by Laboratorio di Linguistica "Giovanni Nencioni" of Scuola Normale Superiore di Pisa (SNS) with the collaboration of Alpine Laboratory of Phonetic Sciences (ALPS) of Libera Università di Bolzano. The innovative aspect of the system lies in the real time, automatic synchronisation of the data from three parallel channels, which further allows the acquisition of the multi-level phonetic information simultaneously. Figure 6.1 illustrates the entire setting of the *SynchroLing* system.

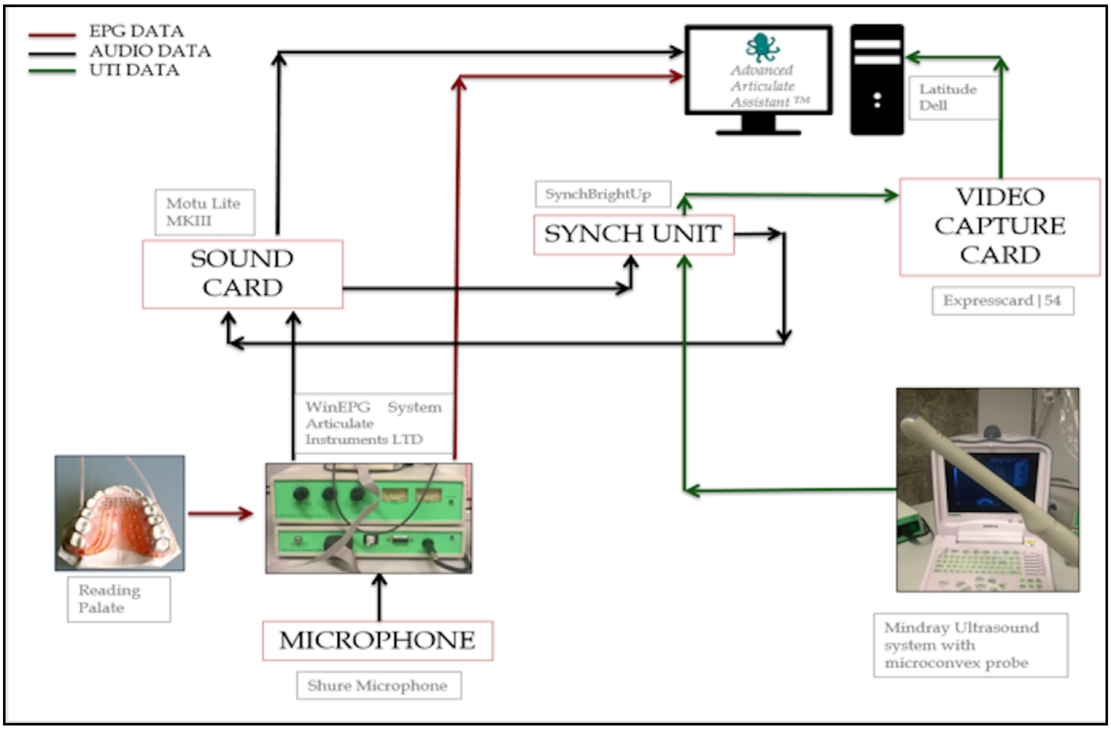


Figure 6.1: *The entire instrumental setting of* SynchroLing *system.*

The system combines three parallel channels: one for the acquisition of audio/acoustic signals; one for the ultrasonographic imagines (UTI) and one for the electropalatalgraphic frames (EPG). The audio/acoutic signals and the articulatory data (UTI and EPG) are synchronised through a synchronisation unit called *SynchroBright* (found in the centre of the Figure 6.1) controlled by the Articulated Assistant Advanced software (AAA; version 2.16.14). The acquisition platform is constructed on the basis of hardware and software components provided by Articulate Instruments Ltd[36]. The UTI data were collected at 60 Hz via a Mindary ultrasound machine with a microconvex probe (Mindary 65EC10EA 6.5 MHz). The EPG data were collected at 100 Hz via the WinEPG™ (SPI 1.0) system. The technical characteristics of the UTI and EPG will be shown in detail in the following sections.

6.3.2 Ultrasound

Ultrasound tongue imaging (hereinafter UTI) provides excellent information on tongue profiles. It involves the use of piezoelectric crystals inside the transducer which are able to emit ultra-high frequency sound waves and then measure the amplitude and delay of the sound waves which are reflected back. The reflection works on the fact that different media have different sound transmission properties (Tabain, 2013). Since tissue and air have different density, their different sound transmission properties result in a bright white line representing the tongue upper profile.

The movement of the whole tongue contour can be identified by UTI inspection. This is one of the most important advantages that UTI possesses over other techniques like Electromagnetic Articulometer (EMA) or Electromagnetic Mid-sagittal Articulometer (EMMA), which are point-tracking techniques and can only track the displacement of single discrete articulatory points of the tongue.

Contrary to EMA/EMMA, UTI has low invasiveness and is relatively inexpensive and potentially portable.

UTI image resolution is rather good: in the machine used for the current study, the frame rate is 30 Hz. Nowadays, with the digital ultrasound machine, the sample rate can reach 100 Hz or above.

---

[36] We acknowledge the technological contribution and technical support of Alan Wrench (Edinburgh).

One important challenges with ultrasound is the difficulty in controlling the position of the probe relative to the head. In this study, a support system especially designed for fixing the probe beneath the chin (a metal helmet) was used. Given its weight, the recording sessions could not be longer than 30 minutes.

6.3.3 Electropalatography

Electropalatography (EPG) measures the contact between the tongue and the artificial palate. The contact is registered by the activated electrodes embedded in a thin, transparent artificial palate which is customised for each individual speaker. The Reading EPG system used in our experiments is composed of 62 electrodes.

EPG provides excellent information on the exact position of linguo-palatal constriction location in lingual consonants; it is also relatively easy to quantify and data acquisition occurs at a relatively high sample rate (around 100Hz in the system used in the current study).

Unfortunately, EPG cannot give any information on the consonants whose articulatory places are lips, tongue root or tongue back; moreover, the customisation required for the construction of the artificial palates limits the number of participants and the quantity of data that can be collected in an experiment.

6.4 Data acquisition and data analysis procedures

6.4.1 Data acquisition procedures

All recordings were made in the sound-proof studio of Laboratorio di Linguistica ‘Giovanni Nencioni’ at Scuola Normale Superiore di Pisa. First, the participants were asked to wear the her/his personal artificial palate nearly half an hour before the recording session, in order to accommodate to the presence of the artificial palate. Meanwhile, the subjects were given a handout with some extra reading material by reading them aloud a sufficient number of times for the purpose of being familiar with words and pseudo-words of the experiment. In addition, the familiarisation phase also served as an adaption phase to reduce salivation and to improve the overall comfort of the subjects in wearing the artificial palate. About 15 minutes later, the ultrasonic transducer was fixed by the helmet beneath the jaw of the participants, orthogonal to the tongue surface so that the ultrasonic wave emitted from the probe was able to catch the entire lingual configuration during the movement of lingual gestures. Approximately half an hour after the wearing of the palate, the recording session began.

In all experiments, the target words were inserted in carrier sentences. During the recording session, the participants read the list of sentences as they appeared on the monitor in front of them. The order of the prompts (sentences) was randomised across speakers but once the order was determined in the first repetition, the following repetitions would keep the same order afterwards. Since the recording session of each participant was limited to 30 minutes, no more than five repetitions of the experimental list were possible.

6.4.2 Segmentation and prosodic annotation of acoustic signals

As presented in the section 6.1, the target segments were those of the first syllable in CVCV sequences and those of the rhyme in CVC sequence. All the speech materials were segmented at a segmental level using Praat (Boersma and Weenink, 2012; version 6.0.29). The sentences in which the target words were in the position of corrective focus had an additional annotation tier, namely, a point tier on which the corrective pitch accent was represented in terms of ToBI labelling system.

It should be mentioned that the star-notation was abandoned in the present annotation of intonation. In the autosegmental-metrical frameworks of intonation, the assignment of an asterisk to a certain tonal event defines a sort of phonological association between such tonal target and the most prominent syllable of the phrase or of the whole utterance. Since the present study follows the principles of Articulatory

Phonology in which the pitch movement is treated as an articulatory gesture, the relationship between segmental level and prosodic level is no more an issue regarding how phonetic alignment maps onto phonological association, and is rather an issue of gestural coordination (see Chapter 4). Figure 6.2 gives an example of segmentation and annotation in *Praat*.

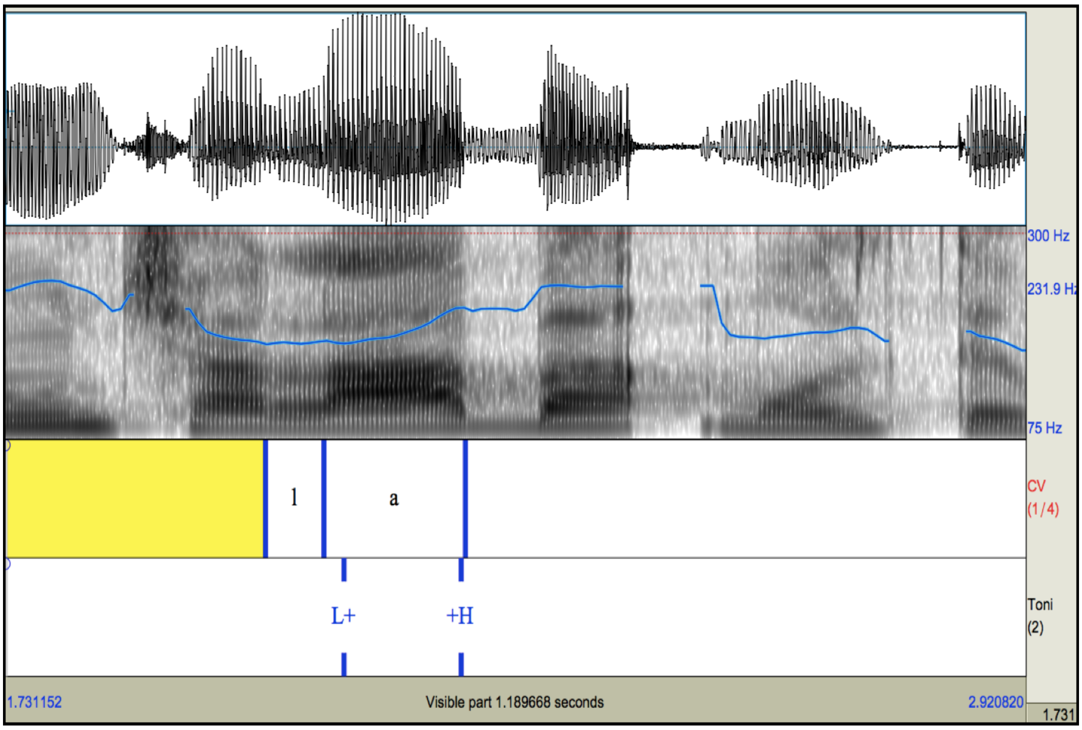


Figure 6.2: *the segmentation of the target syllable /la/ and the transcription of the corrective pitch accent.*

6.5 Data Analysis: from the raw data to the articulatory landmarks

6.5.1 Defining articulatory landmarks

For the purpose of a temporal coordination study, appropriate articulatory landmarks on both segmental and prosodic levels had to be determined. This section explains how the articulatory landmarks were defined. For the analysis of gestural coordination, four articulatory landmarks were defined, and each target syllable was assigned three segmental landmarks according to its structure. Singletons are taken as examples in the following sections, but the same procedure applied to consonants included in clusters.

For CV syllables, Ctarget, Crelease and Vtarget were identified, whereas for VC sequences, Vtarget, Conset and Ctarget were identified. We define:

i) Ctarget as the time-point at which the consonantal gesture reaches its maximum constriction in the relevant lingual and palatal areas and thus bears minimal influences of coarticulation from the adjacent vowels;

ii) Vtarget as the time-point at which the vocalic gesture reaches its target configuration and undergoes minimal influences of coarticulation from the adjacent consonant;

iii) Crelease[37] as a time-point within the acoustic interval of the preceding consonant at which the consonant constriction begins to release and it corresponds to the onset of the gesture for the following nuclear vowel (in the case that the following gesture is a consonantal one, then it also indicates the Conset of the following C gesture);

iv) Conset as a time-point often within the acoustic interval of the preceding vowel at which the constriction gesture for the coda consonant begins.

In order to identify the relative position of the segmental landmarks defined above, gestural stability areas were first detected on the two articulatory channels (EPG and UTI) separately. Then the two sources of information were combined together to identify one reference gestural stability area and the landmarks for coordination measurements. The procedures are described below.

6.5.2 UTI evidence

The analysis of ultrasound tongue images was based on the idea that the target lingual configuration of a consonantal gesture differs maximally from the target lingual

---

[37] In Chen *et al.*, (2017) we referred to Crelease as "Vonset", due to the fact that we only analyzed singletons, and in the case of singletons the following articulatory gesture could only be a vowel gesture. However, in the case of consonant clusters, the articulatory offset of the first consonant is not a Vonset, but rather, it corresponds to the articulatory onset of the second consonant in the cluster. For this reason, and for consistency in the use of terminology, Vonset was here re-nominated as Crelease.

configuration of the adjacent vocalic gesture. The analysis of ultrasound tongue images was divided into three steps:

i) selection of a reference spline for the characterisation of the vocalic gesture. The spline that happened to be the closest to the acoustic midpoint of the vowel was selected as reference.

ii) identification of relevant fan radii for tongue vertical displacement measurement. Depending on the different articulatory properties shown by the lingual sagittal profile, two radii – the 7th (called "first front") and the 14th (called "first middle") from the right – were identified to maximally capture the articulatory differences between vocalic and consonantal gestures in terms of tongue vertical displacement in a relevant area of the oral cavity. The first front radius was seen to be responsible for the difference between the /a/ gesture and the consonantal gestures of /s/ and /l/, as well as for the difference between the /i/ gesture and the consonantal gesture of /k/. The first middle radius was seen to be responsible for the difference between the /i/ gesture and the consonantal gestures of /s/ and /l/, as well as for the distinction between /a/ and /k/;

(iii) Identification of a gestural stability area for the consonant involved and of the relevant Ctarget landmark. Once the reference spline for the vocalic gesture and the fan radius for capturing the C-V difference were established, the distance between the tongue position in the reference spline and the tongue position in all the splines included in the acoustic interval of the consonant along the selected fan radius was calculated (in millimetres). The temporal interval during which this distance reached its maximum was defined as the UTI gestural stability area of the consonant. So the UTI gestural stability area corresponded to the interval during which the tongue was maximally different (in terms of vertical displacement along the selected fan radius) from the vowel.

6.5.3 EPG evidence

A traditional procedure based on the contact index method (Fontdevila, Recasens and Pallarès, 1994) was used for the identification of vocalic and consonantal gestures from EPG data. Since EPG allows the detection of the linguo-palatal contact during speech production, it is possible to get different kinds of information for various regions on the artificial palate by referring to different EPG indices, implemented in the AAA software (version 2.16).

Two of these indices were used to detect the EPG gestural stability area. Specifically, the CAa (Contact Anteriority in the anterior palate) index was calculated on the three most anterior rows of each EPG frame included in the acoustic intervals of each /s/ and /l/ token. CAa served as an indicator of the anteriority of the linguo-palatal contact of /s/ and /l/: the higher the CAa value, the more apical contact on the most anterior region of the palate. Consequently, the temporal intervals characterized by the maximum CAa values within the acoustic intervals of /s/ and /l/ were identified as the gestural stability areas of /s/ and /l/. For /k/, the Qp (percentage of activated electrodes in the four back rows of the palate) index was used: the higher the Qp value, the greater constriction in the palatal area for the realisation of a given consonant gesture. Thus the temporal intervals characterised by the maximum Qp values within the acoustic intervals of /k/ were identified as the gestural stability areas of /k/.

6.5.4 Identification of gestural stability area

We finally put the two gestural stability areas together, the intersection of them being identified as the stability area for the consonantal gesture, because both lingual profiles and palatal contact maximally distinguished such gesture from the adjacent vocalic gestures.

Six different conditions of temporal overlap of UTI and EPG stability areas were found in the data. They are schematically represented in Figure 6.3. The intersected regions represent the final gestural stability areas (a-d). In only two items, the UTI and

EPG stability areas were found not to overlap (e-f). In those cases, the EPG gestural stability area was taken as the gestural stability area[38].

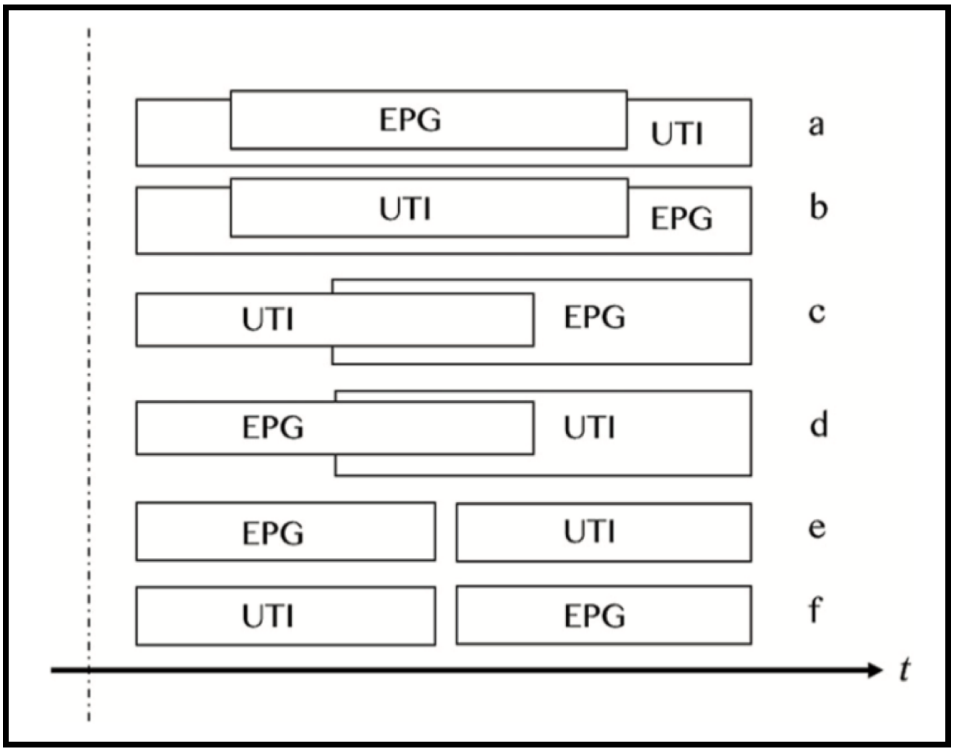


Figure 5.3 (a-f): *six idealised patterns of EPG-UTI gestural overlap and resulting gestural stability areas. The horizontal axis illustrates the temporal extension in which the gestural stability areas of both channels were identified.*

In CV sequences, the time point corresponding to the beginning of the gestural stability area was identified as Ctarget, whereas its offset was identified as Crelease (see footnote 3 for an explanation of how Crelease was originally named as Vonset). As previously introduced (in the section 6.5.2), Vtarget corresponded to the acoustic midpoint of the vowel.

As to the VC sequence, the time point corresponding to the beginning of the gestural stability area was identified as Ctarget. A further step was therefore needed to identify Vtarget and Conset in VC sequences. In these cases, the UTI spline of Ctarget previously obtained was treated as the reference for the identification of the stability area for the vocalic gesture. Then, the C-V distance along the same fan radius

---

[38] Fortunately, throughout the whole experiment, there were only two cases in which we had to rely on the EPG information due to the lack of intersection area. This indirectly proved that the experimental setting is highly reliable.

was calculated within the acoustic interval of the vowel. Similar to the third step described in the section 6.5.2, the distance values plotted as a function of the temporal development gave a plateau representing the gestural stability area of the vowel involved. In the end, the same criteria were applied: the time point corresponding to the first spline of the gestural stability area was identified as Vtarget, the last one as Conset. Figures 6.4 (a-b) illustrate how the articulatory landmarks map onto the acoustic spectrogram, thus showing the discrepancies between the acoustic and the articulatory evidence.

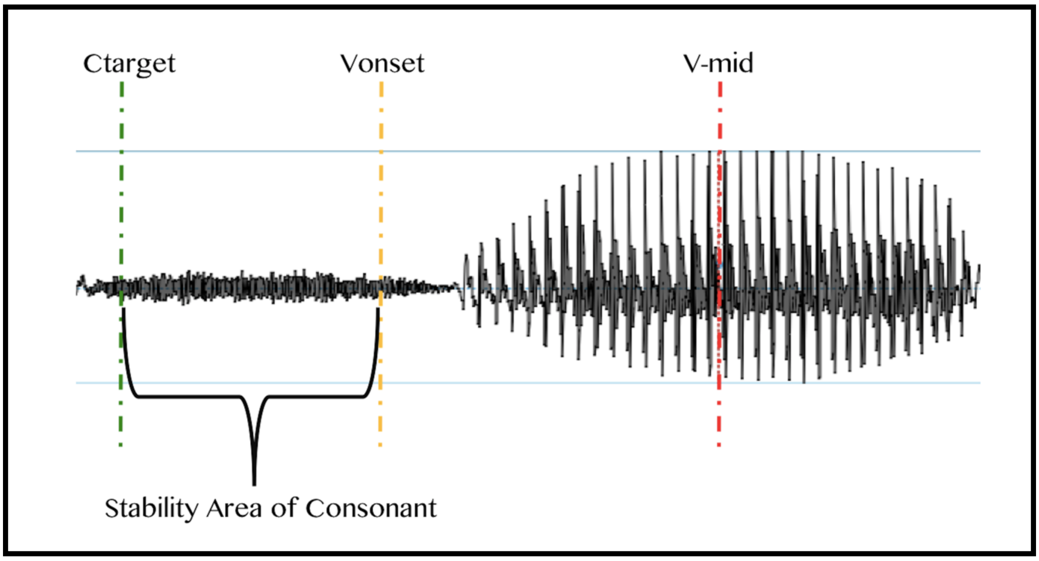


Figure 5.4 (a): *example of the stability area of a consonantal gesture and relative articulatory landmarks involved in the analysis of a CV syllable as mapped onto the acoustic signal.*

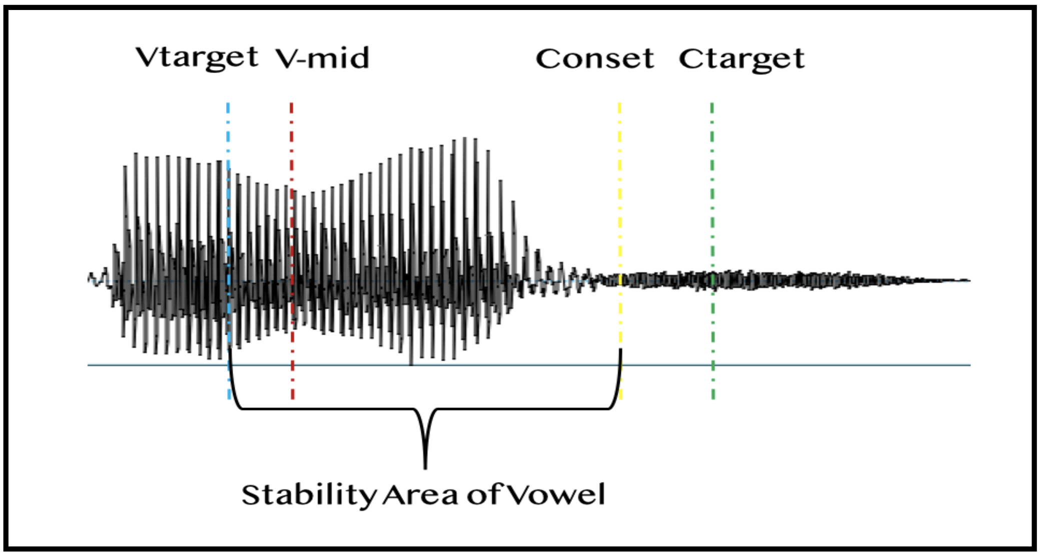


Figure 5.4 (b): example of the stability area of a vowel gesture and relative articulatory landmarks involved in the analysis of a VC syllable as mapped onto the acoustic signal.

Once all the relevant gestural landmarks were identified for both CV and VC sequences, the gestural coordination latencies were calculated. The latency between Ctarget and Crelease was taken as a cue of CV coordination in CV syllables, whereas the latency between Vtarget and Conset was taken as a cue of VC coordination in CVC syllables.

For the postvocalic or intervocalic consonant clusters in our corpora, a similar procedure were applied. In VC1C2 sequences, the stability areas for both consonants could be identified through the integration of UTI and EPG information; the stability area of the vowel gesture was defined as in the VC sequences. Thus, six articulatory landmarks were obtained: from left to right, Vtarget, C1onset, C1target, C1release, C2target and C2release. The latency between the Vtarget and C1onsst was considered as a cue of nuclear-coda coordination, and the distance between the C1target and C1release was taken as a cue of the internal temporal cohesion of the consonants in the cluster. C2target and C2release served as the indication of the C2V2 coordination in V1C1C2V2 sequences[39]. Figure 6.5 shows an example for a VC1C2(V) sequence.

---

[39] The same vowel - /a/ - is chosen for both V1 and V2. The mid-point of V1 is selected as the reference for the identification of the stability areas of consonants because the first vowel bears the lexical accent, therefore it should be maximumly realised and the UTI might give more evident differences when being compared to the profiles of the consonant gestures.

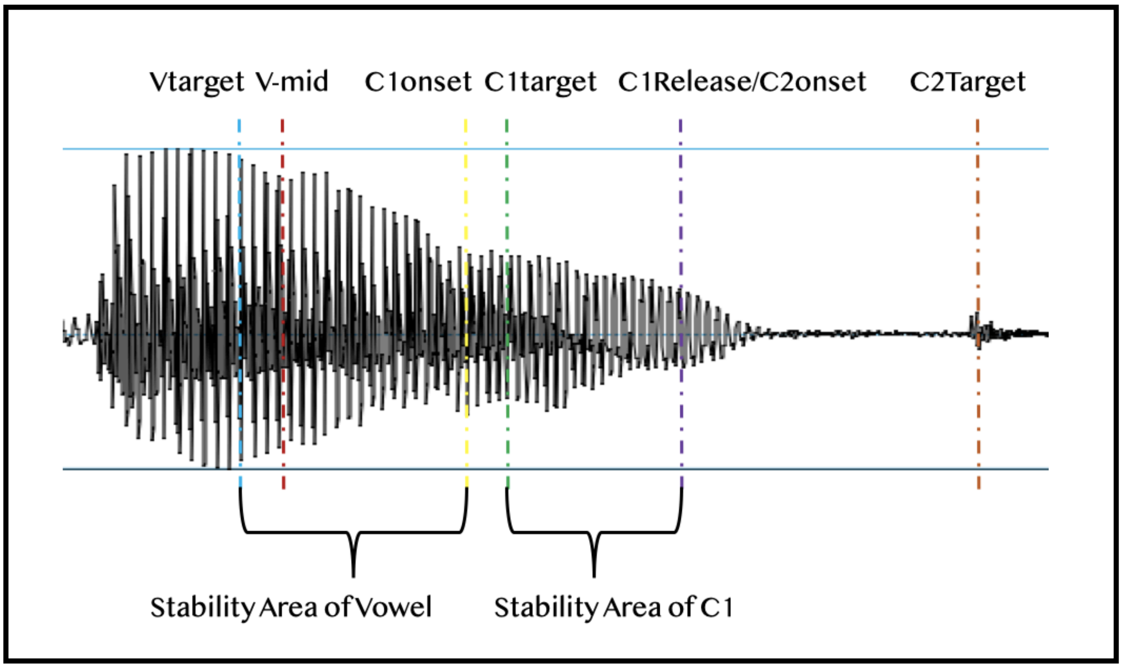


Figure 5.5: *example of the gestural stability areas of vowel and the vowel-adjacent consonant, and relative articulatory landmarks involved in the analysis of VC1C2 syllables as mapped onto the acoustic signal.*

One crucial issue has to be pointed out, that is, unlike in EMA-based studies, the inter-gestural timing is no longer viewed in terms of in-phase and/or anti-phase but represented by means of tight and/or loose coordination in the *SynchroLing*-based study. As introduced above, the inter-gestural timing in the *SynchroLing* system is defined by the temporal latencies between the articulatory landmarks defined above, *i.e.*, the gestural stability area, which in turn, reflects the inter-gestural coupling strength. Thus, whether the inter-gestural timing is tight or loose is always relatively defined.

6.6 Analysis of coarticulation

To evaluate if the three consonants /l/, /k/ and /s/ differed in their overall coarticulatory properties, qualitative inspection of UTI profiles and quantitative evaluation of EPG indices over the acoustically defined consonantal interval were also run. As for EPG indices, in the analysis of singleton consonants, the CAa (Contact

Anteriority in the anterior palate) index for /s/ and /l/ and the Qp (percentage of activated electrodes in the palatal area) index for /k/ were the dependent variables; for the consonant clusters, an additional index – CCa (Contact Centrality in the anterior palate) – was also chosen for laminal /s/ when followed by a coronal stop /t/ (*e.g.*, in target words /bast/ and /bas.ta/).

Following Fontdevila et al., (1994), the Qp index was obtained by averaging all the activated electrodes at the palatal zone by the total number of 24 electrodes (the last 3 rows from the back)[40] and rescaling the resulting values so that a range between 0 and 1 was obtained (Recasens, 2004). CAa, as introduced in the section 6.5.3, reflected the degree of alveolar contact fronting (*i.e.*, the 4 most front rows) (Recasens and Espinosa, 2006). The index CCa reflected the centrality of the constriction, with higher CCa value indicating narrower and more central constriction width. The CCa index is usually applied to /s/ and /ʃ/ in order to provide an estimate of changes in alveolar constriction width (Recasens, 2004: 441).

In both coordination and coarticulation analyses the factors under evaluation were syllable structure (CV *vs*. VC), consonant (/k/ *vs*. /l/ vs. /s/) and vowel (/a/ *vs*. /i/). The statistical analyses were univariate analyses of variance and the non-parametric Mann-Whitney test for distribution differences in independent samples. Non-parametric statistics was run whenever the requirements for ANOVA were not met (*i.e.*, asymmetry and kurtosis higher than |1| and/or unequal variances, as attested by the Levene's test). The statistics were computed using SPSS (version 25.0.0).

6.7 Hypotheses

Two main hypotheses are made following the coupling oscillator model of inter-gestural temporal coordination and the DAC model of coarticulation between/among vocalic and consonantal constriction gestures. The first hypothesis regards the

---

[40] The calculation of the Qp value was originally based on the linguo-palatal contacts in the last 4 rows, thus by a total number of 32 electrodes (see Fontdevile *et al.*, 1994; Recasens and Espinosa, 2006).

consonant-to-vowel cohesion in the case of singleton consonants and the second concerns the articulatory constraints of individual consonants.

6.7.1 Hypotheses on macroscopic consonant-to-vowel inter-gestural cohesion

For singleton consonants, we expected that syllable onset constriction gestures were more tightly coordinated to the upcoming vocalic nucleus than the constriction gestures in the syllable coda were with respect to the preceding vocalic nucleus. Testing this hypothesis was not only meant to replicate a well-known articulatory timing effect on a different speech dataset, but - most importantly - served to establish if the procedure for the identification of articulatory landmarks via *SynchroLing* (see above the section 6.5.3) was reliable. We anticipate that it was; by looking at *SynchroLing*-based gestural stability areas, it was possible to correctly identify the onset and offset of consonantal and vocalic gestures for an analysis of inter-gestural coordination that mirrors EMA-based coil tracking procedures.

For post-vocalic consonant clusters, we expected that the gestural cohesion between the vocalic nucleus and the leftmost consonant in the cluster (V-to-C1 cohesion) showed the same loose coordination pattern that singletons also showed. Moreover, the V-to-C1 cohesion was expected to remain unchanged regardless of the syllabicity status of the rightmost consonant, since in the gestural model of consonant-vowel temporal coordination, the postvocalic word boundary clusters are empirically claimed to be timed in a sequential fashion and no leftword shift of vowel-adjacent consonant is hypothesised (Browman and Goldstein, 1988; Goldstein and Pouplier, 2011).

6.7.2 Hypotheses on microscopic consonant articulatory properties

The second hypothesis concerns the possibility that the coordination patterns vary as a function of the specific consonants involved. This hypothesis was suggested by the findings reviewed above (see Chapter 5), related to the c-centre effect in onset and coda clusters and the effects of coarticulatory resistance on shaping cluster-specific coordination patterns. Since the c-centre effect in consonant clusters is one of the consequences of tight *vs*. loose coordination of syllable onsets *vs*. codas, respectively (since Krakow, 1999), we hypothesised that the degree of coarticulatory resistance could in principle affect not only the way in which singletons are coordinated to the following or preceding vocalic nucleus (C-to-V and V-to-C timing), but also the within-cluster temporal organisation (C1-to-C2 timing).

In the investigation of singletons, three consonants - /s/, /l/ and /k/ - were assumed to vary for their degree of resistance to vowel-induced coarticulatory modifications. Their degree of coarticulatory resistance was established by measuring the articulatory modifications of each consonant adjacent to /i/ as compared to adjacent to /a/. If coarticulatory resistance only affects the magnitude of the C-centre effect (possibly because clusters are more susceptible than singletons to the influence of varying articulatory conditions), then the three consonants should show a similar difference between tight onset-nucleus coordination (*i.e.*, a short latency between the consonantal and the vocalic gesture in CV sequences) and loose onset-nucleus coordination (*i.e.*, a comparatively longer latency between the vocalic gesture and the consonantal gesture in VC sequences). If, on the contrary, consonantal properties such as coarticulatory resistance influence the onset-coda coordination pattern in singletons, the three consonants are expected to vary in the way they are timed to the vocalic nucleus as onsets or codas. In particular, spatial (coarticulatory) and temporal effects of vowel variation over adjacent consonants were expected to be positively correlated. Less resistant consonants (*i.e.*, those that vary a lot as a function of vocalic context) were expected to be influenced by the vocalic gesture both in the sense of modifying their constriction location (spatial coordination) and in the sense of an increased temporal overlap between the vocalic gesture and the consonantal gesture. The effect was expected to be particularly strong for consonants in onset position (whereas, following the gestural timing model proposed by Browman and Goldstein (1988), we expected that the three consonants did not differ in the coda position). Thus, less resistant consonants in onset position were expected to show not only more variability of the constriction location in the /a/ *vs.* /i/ contexts, but also a stronger anticipation of the vocalic gesture. Consequently, they were also expected to show larger onset-coda coordination differences, compared to more resistant consonants. In the case of the consonants of the present study, /s/ was expected to be the most resistant consonant and to show the smallest onset-coda coordination difference; /k/ was expected to be the least resistant consonant and to show the largest onset-coda coordination difference; /l/ was expected to occupy an intermediate position between the two, as far as resistance to coarticulation is concerned, and therefore to show an onset-coda coordination difference smaller than /k/ and larger than /s/.

As for the singleton coda vs. coda clusters opposition, we expected no difference in the way a coda consonant is timed to its preceding vocalic nucleus, thus irresepective of the fact of being word-final or followed by another consonant.

Finally, as far as within-cluster gestural organisation is concerned, the properties of both consonants in the clusters were expected to be able to affect C1-to-C2 cohesion. Based on the assumed hierarchy of degrees of coarticulatory resistance explained above (/s/ > /l/ > /k/), it is hypothesised that, on one hand, the C1-to-C2 coordination within /sC/ clusters is looser than that within /lC/ clusters; on the other hand, the C1-to-C2 coordination is expected to be tighter in /Ck/ clusters than in /Ct/ clusters.

In the following two chapters, we will present the results of the experiment described above.

Chapter VII Results (I): Temporal Coordination

The current chapter will present the results of the experiments introduced in the previous chapters. They are divided into two parts. The first part will concentrate on temporal coordination results, and positional differences on the gestural overlap between C gestures and V gestures. Onset-nucleus and nucleus-coda will be presented in the first place, then the consonant clusters will be presented, including three different categories of data: word-final branched coda and intervocalic consonant clusters. The second part will focus on the coarticulatory analysis of segment-induced effects from two different channels - UTI configurations and EPG indices. As in part one, the results on singleton consonants will be discussed first, and then followed by the results of consonant clusters.

7.1 Coordinative relations between C and V in onset and coda: singleton consonants

7.1.1 Variations on sylalble structures

In this section, we verified if there is variation in the consonant-vowel temporal coordination patterns according to the different syllabic status of the consonant (onset vs. coda). In addition, we verified whether the coordination patterns are affected by the articulatory properties of the consonants involved, with specific reference to varying degrees of coarticulatory resistance. Table 7.1 shows the gestural coordination latencies across stimuli and subjects.

Syllable Type	*N*	*mean (ms)*	*st. dev.*
CV	103	38	20
VC	102	105	20.4

Table 7.1: *Gestural coordination latencies in CV and VC sequences across subjects, vocalic contexts and prosodic conditions.*

For the onset-nucleus coordination (CV syllables), the mean value of the Ctarget-Vonset interval across all valid tokens was 38 msec. For the nucleus-coda coordination (VC sequences in CVC syllables), the mean value of the Vtarget-Conset interval across all valid tokens was 105 msec. The temporal interval occurring between the consonantal gesture and the vocalic gesture in onset-nucleus sequences was therefore much shorter than the temporal interval between the vocalic gesture and the consonantal gesture in nucleus-coda sequences (negative difference: 67 msec), which suggested an opposition between in-phase coordination (quasi-simultaneous activation) and anti-phase coordination (sequential activation) in CV *vs*. VC sequences. In an ANOVA with C-V latency as dependent variable and Syllable as between subject factor, the difference between CV and VC was statistically significant (F (1, 205) = 523.869, p < .001).

7.1.2 Variations on prosodical prominent: broad focus vs. corrective focus for /a/

We also examined if the variations induced by syllabic constituency are shaped by prosodic prominence. The stimuli analyzed here all have /a/ as nuclear vowel (see the section 6.1).

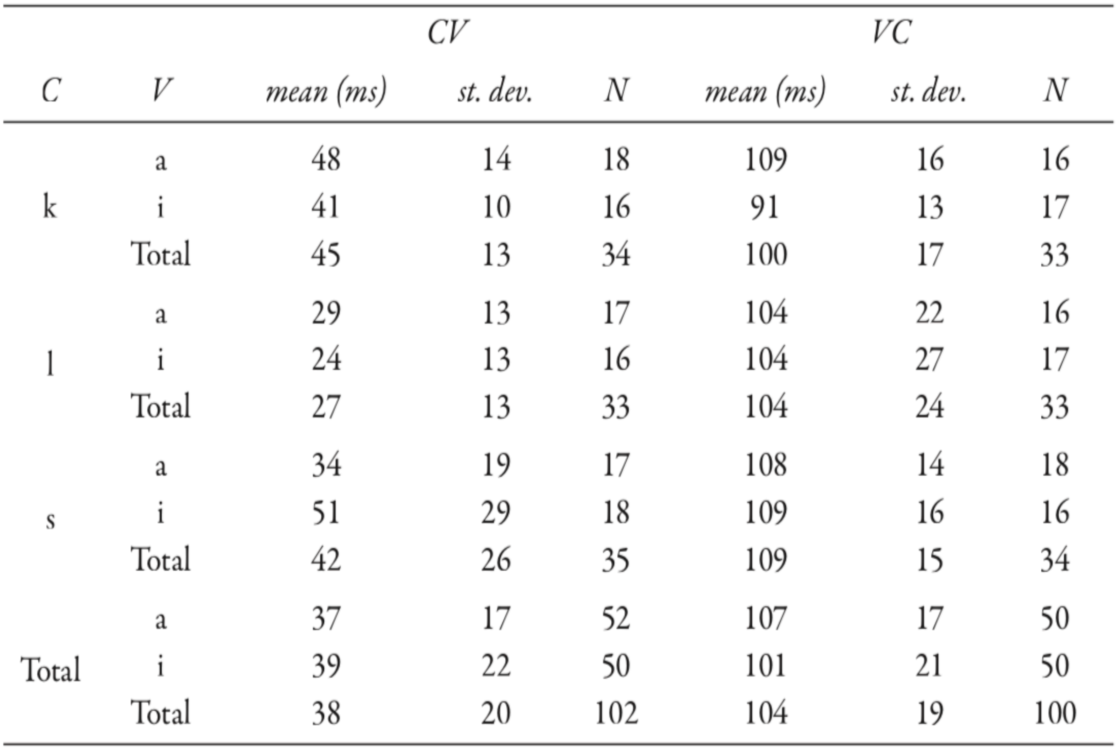
Syll.	*PA*	*mean (ms)*	*st.dev.*	*N.*
CV	0	38	17,8	53
	1	34	19	53
	Total	35,9	18,4	106
VC	0	108,2	19	51
	1	130,9	29,7	52
	Total	119,7	27,3	103

Table 7.2*: Temporal coordination between C and V in two different prosodic contexts: broad focus (PA = 0) vs. corrective focus (PA = 1).*

Table 7.2 provides the mean and standard deviation values. For CV sequences, the presence of a corrective focus has no effect on onset-nucleus timing (F (1, 104) = 1,317, p > .05). However, in nucleus-coda coordination, the nonparametric test on the median VC latency shows that the average latency from Vtarget to Conset does significantly vary depending on whether the corrective focus is present (p < ,005), syllables with a contrastive focus showing a longer coordination interval than syllables with broad focus. This can be easily explained by considering that vowels in prosodically prominent syllables are significantly longer than vowels in prosodically non prominent syllables[41] (220,11 ms for the /a/ in bass with corrective focus *vs*. 173,82 ms for the /a/ in bass with broad focus).

7.1.3 Variations on the properties of consonants

The following Table 7.3 shows the gestural coordination latencies as a function of consonantal (/k/, /l/, /s/) and vocalic variations (/a/, /i/).



		CV			*VC*		
C	*V*	*mean (ms)*	*st. dev.*	*N*	*mean (ms)*	*st. dev.*	*N*
k	a	48	14	18	109	16	16
	i	41	10	16	91	13	17
	Total	45	13	34	100	17	33
l	a	29	13	17	104	22	16
	i	24	13	16	104	27	17
	Total	27	13	33	104	24	33
s	a	34	19	17	108	14	18
	i	51	29	18	109	16	16
	Total	42	26	35	109	15	34
Total	a	37	17	52	107	17	50
	i	39	22	50	101	21	50
	Total	38	20	102	104	19	100

Table 6.3: *Temporal coordinations between C and V, and between V and C as a functiuon of variations on consonant property and vowel quality.*

---

[41] It should be noted that a longer acoustic duration does not necessarily correspond to a longer gestural stability duration at the articulatory level. On the contrary, a long gestural stability duration might suggest a long acoustic interval since the acoustics could be interpreted as the physical representation (both quantitative and qualitative) of the articulatory activities.

The mean values confirme that the sharp distinction between CV and VC coordination patterns was present in all three consonants of the study, with a negative difference between the two conditions of 55 msec for /k/, 67 msec for /s/ and 77 msec for /l/.

The interaction between Syllable and Consonant turned out to be statistically significant (F (5, 201) = 5.906, p < .05), thus confirming that the coordination pattern was different across consonants. In particular, post-hoc Tukey HSD test revealed that /l/ was significantly different from both /s/ (p < .01) and /k/ (p < .05), whereas /s/ and /k/ were not significantly different. As a matter of fact, average latency values for VC sequences were similar across consonants, whereas for CV sequences there was a much lower latency value for /l/ (27 msec on average) than for the two other consonants (45 msec for /k/ and /42 msec for /s/).

The difference between CV and VC coordination was equally present in /a/ (70 msec) and /i/ stimuli (65 msec), as confirmed by the non-significant Syllable by Vowel interaction (F (3, 201) = 0.713, p > .05).

We could therefore conclude that the properties of the consonantal gesture influenced the gestural coordination patterns, whereas vowel quality did not.

Finally, as far as the three-way interaction 'Syllable*Vowel*Consonant' is concerned, the Leven's test turned out to be significant which means that the prerequisites for the analysis of variance are not met in this dataset. For this reason, a graphical representation of the the differences is only provided to show the interaction between the consonant and the vocali context (see Figure 7.1 and Figure 7.2). This shows that, for the nucleus-coda coordination (see Figure 7.1), the three consonants do not differ to a great extent from one another in the /a/ context; on the contrary, the differences are much larger in the case of the /i/ context, in which /s/ shows the longest interval between C and V, /l/ the shortest and /k/ is in-between. For the onset-nucleus coordination, the situation is more complex, however, it is consistent with the articulatory scenario of the segments involved. In particular, for /s/, the less tight inter-gestural cohesion in the context of /i/ comparing to that of /a/ is uncontroversial to the fine control of laminal constriction since during the formation of the laminal constriction the tongue region involved in the production of /a/ has more independence than the tongue region participating in the palatal constriction of /i/.

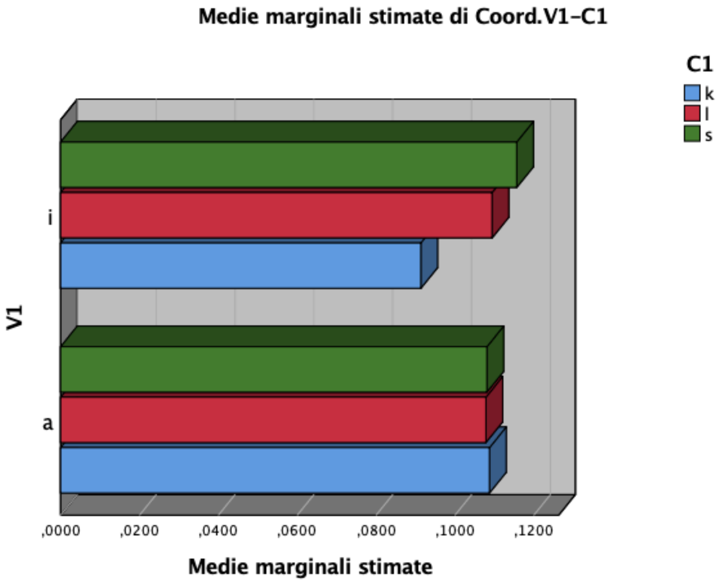


Figure 7.1: *the impact of the interaction between vowel quality and consonant property on the nucleus-coda coordination.*

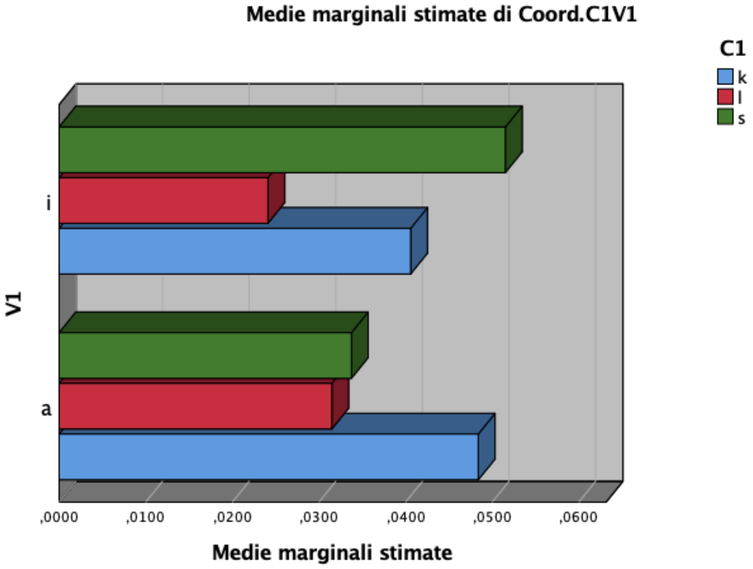


Figure 7.2: *the impact of the interaction between vowel quality and consonant property on the onset-nuleus coordination.*

7.2 Coordinative relations between C and V: consonant clusters

The results on the inter-gestural coordination for complex coda and intervocalic consonant clusters will be presented in the following section. The focus will be put on nucleus-coda coordination (*i.e.,* V1-to-C1), temporal organisation within coda clusters (*i.e.,* C1-to-C2) and inter-gestural timing between the rightmost consonant in the cluster and the following vowel (*i.e.,* C2-to-V2). Moreover, we will examine the coordination from three specific viewpoints: the syllable structure (*i.e.,* VC, VC1C2 and V1C1C2 V2), the articulatory property of the leftmost consonant in the consonant cluster and the articulatory property of the rightmost consonant in a cluster.

7.2.1 V-to-C1 temporal coordination

The nucleus-coda coordination for three different syllable structures, *i.e.,* VC, VC1C2 and VC1C2V is shown in the following Table 7.5. For VC sequence, the tokens with nuclear vowel /i/ and prosodic prominence of the corrective focus in the /a/ context are excluded from the current analysis for the reason that the further comparion between singltons and clusters should be based on the identical segmental and prosodic condition. The mean value of V-to-C latency is inverse to the complexity (or the length) of the phonetic string to the right of the stressed vowel: the more complex (or long) the phonetic string is, the more tight the temporal coordination between the nuclear vowel and the adjacent coda consonant turns out to be. A nonparametric test on the median of the nucleus-coda latencies as dependent variable and the segmental complexity as within-subject factor gives a significant result ($p < .01$), suggesting that the degree of the segmental complexity indeed can influence the inter-gestural timing between the nuclear vowel and the vowel-adjacent coda consonant.

Syll.	*mean (ms)*	*st.dev.*	*N.*
VC	108,2	19	51
VCC	106,4	20,4	86
VCCV	99,7	12,7	35
Total	105,6	18,8	172

Table 7.5: *nucleus-coda gestural coordination as a function of segmental complexity to the right of the stressed vowel across speakers and consonant properties.*

The biggest difference is found between the disyllabic VCCV and the monosyllabic stimuli (VC and VCC), the former showing a much shorter temporal latency (or tigther coordination) between the nuclear vowel and the adjacent coda consonant than the latter ones. A much smaller difference is found in the comparison between VC and VCC, which indicates that the number of the coda consonants in itself (*i.e.,* by the fact of having a single or a branching coda) is not very relevant in predicting the degree of temporal cohesion between the vocalic and the consonantal gestures. On the contrary, the fact of being or not a word-final coda has a bigger impact on the temporal cohesion with the vocalic nucleus. This finding suggests that looking at mono- vs. polysyllables might affect the results of the investigation. When non-word-final codas are considered, the difference in temporal cohesion with respect to consonantal onsets is less extreme.

In the following sections, we will examine the contribution of the individual consonants in complex codas and the intervocalic consonant clusters respectively.

In Table 7.6 the mean inter-gestural timing values as a function of the vowel-adjacent consonants are presented.

V-to-C1 coordination				
C1	*Syll.*	*mean (ms)*	*st. dev.*	*N*
s	VC_1C_2	108,3	17,9	35
l	VC_1C_2	113,9	21,6	33
k	VC_1C_2	89	11,4	18
Total	VC_1C_2	106,4	20,4	86

Table 7.6: *nucleus-coda gestural coordination as a function of C1 identity in VC1C2 sequences.*

For V-to-C1 coordination in VCC sequence, when the vowel-adjacent consonant is an apical fricative - /s/ or a lateral stop - /l/, the mean values of temporal latency are similar (108 and 114 ms, respectively); the nucleus-coda coordination interval is much shorter when the vowel-adjacent consonant is the velar stop /k/. The difference is statistically significant: the post-hoc Tukey HSD test reveals that the /k/ context is significantly different from both the /s/ (p < .005) and the /l/ context (p < .001).

Table 7.7 exhibits the mean values of V-to-C1 latency as a function of C1 identity and segmental complexity. Once again, things change with changing consonants.

V-to-C1 coordination				
C1	*Syll.*	*mean (ms)*	*st. dev.*	*N*
s	VC	108,1	14,2	18
	VCC	108,3	17,9	35
	VCCV	99,7	12,7	35
	Total	104,9	15,6	88
l	VC	107,9	26,1	17
	VCC	113,9	21,6	33
	Total	111,9	23,1	50
k	VC	108,8	15,6	16
	VCC	89	11,4	18
	Total	98,3	16,7	34
Total	VC	108,2	19	51
	VCC	106,4	20,4	86
	VCCV	99,7	12,7	35
	Total	105,6	18,8	172

Table 7.7: *nucleus-coda gestural coordination as a function of consonant properties of the vowel adjacent consonant across the subjects in different syllabic affiliation.*

In particular, for /s/, the mean value of the V-to-C1 coordination is the same in singleton and branched codas, whereas if the coda /s/ is followed by another CV sequence (dysillabic stimuli), the temporal latency is shorter.

When the vowel-adjacent consonant is /l/, the vowel-coda coordination remains stable in both singleton and branching codas (F (1, 48) = ,772, p > .05).

Finally, when the vowel-adjacent consonant is the velar stop /k/, the V-to-C1 latency in the branching coda is significantly shorter than in the singleton coda (F (1, 32) = 18,042, p < .001).

It can therefore be concluded that only /k/ as a C1 is more tightly coordinated with the preceding vocalic nucleus if the complexity of the coda increases, whereas /s/ and /l/ do not show any change. Additionally, the clusters with /s/ indicate that if the coda is non-word-final, the coordination is also tighter than when the coda is word-final. taken together, these results suggest that the temporal coordination difference between onsets and codas is maximal when word-final, singleton codas are considered; this difference can eventually be reduced, at least with some particular consonants, when the coda is non-word-final and part of a branching coda.

A further hypothesis that can be put forth is that, in complex codas, the nucleus-C1 coordination could vary as a function of the identity of the C2. Table 7.8 provides information about that.

V-to-C1 coordination					
C1	*C2*	*Syll.*	*mean (ms)*	*st. dev.*	*N*
s	k	VC_1C_2	115,4	20,9	18
	t	VC_1C_2	100,8	9,9	17
	Total	VC_1C_2	108,3	17,9	35
l	k	VC_1C_2	120,3	24,6	16
	t	VC_1C_2	108	16,9	17
	Total	VC_1C_2	113,9	21,6	33
Total	Total	VC_1C_2	111	19,8	68

Table 7.8: *nucleus-coda gestural coordination as a function of the interaction between the vowel-adjacent consonant and the vowel-remote consonant, excluding the tokens with /ks/.*

When the vowel-adjacent consonant (C1) is an apical fricative /s/, the nonparametric test on median of the V-to-C1 temporal latency confirms that C2 plays a significant role to the nucleus-coda coordination (p < .01), with the coordination being tighter when the /s/ is followed by a /t/ than when it is followed by a /k/. However, when the C1 is the lateral stop /l/, the result of an ANOVA reveals that the nucleus-coda coordination is insensitive to the articulatory property of the vowel-remote consonant (F (1, 31) = 2,845, p > .05). It could therefore be tentatively concluded that homorganicity of the segments in the branching coda might play a role in determining the degree of articulatory cohesiveness between the nuclear vowel and the following consonant, *i.e.*, by increasing the articulatory cohesiveness, but the effect is evident only in the case of some particular consonants and particular clusters; laterals as C1, for instance, appear to be insensitive to cluster homorganicity.

In the specific case of /s/ as a C1, we can also look at the potential role of C2 identity in the intervocalic condition (VCCV) (Table 7.9). A nonparametric test of the mean reveals that the identity of C2 (either /k/ or /t/) is, unlikely the case of the /sC/ clusters in the absolute position (*e.g.*, postvocalic word-final), irrelevant for the variation of V-to-C1 coordination (p > .05). This could probably be attributed to the presence of an additional /a/ with which the C2 takes part in the inter-gestural overlap. As a result, V-to-C1 temporal relation undergoes less C2-induced impact.

V-to-C1 coordination				
C2	*Syll.*	*mean (ms)*	*st. dev.*	*N*
k	VC_1C_2V	102,3	15,8	17
t	VC_1C_2V	97,4	8,6	18
Total	VC_1C_2V	99,7	12,7	35

Table 7.9: *nucleus-coda gestural coordination as a function of the interaction between the vowel-adjacent consonant and the vowel-remote consonant in VC1C2V sequences.*

7.2.2 C1-to-C2 coordination

In the current section, we will examine inter-gestural coordination within consonant clusters as a function of syllabic affiliation, the identity of C1 and C2.

The consonant clusters involved in the study have two different patterns of syllabication depending on the presence or absence of the vowel /a/ to the right of the clusters. Table 7.10 shows the inter-gestural coordination values within the consonant clusters as a function of the different postvocalic segmental compositions (VC1C2 *vs.* VC1C2V, when C1 = /s/).

C1-to-C2 coordination			
Syll.	*mean (ms)*	*st.dev.*	*N.*
VCC	163,5	22,2	35
VCCV	164,1	29,7	35
Total	163,8	26	70

Table 7.10: *inter-gestural coordination within the consonant clusters as a function of the syllabification patterns.*

When the leftmost consonant in the cluster is the sibilant /s/, the mean C1-to-C2 latency is almost the same when the consonant clusters form the complex coda as a whole and when the clusters are in intervocalic position. In the case of word-final clusters, the identity of C1 and C2 may shape the degree of their cohesiveness. Tables 7.11 and 7.12 illustrate the C1-to-C2 coordination values separately for the various conditions.

When C2 = /k/, the inter-consonantal coordination is tighter when C1 = /l/ than when C1 = /s/ (F (1, 32) = 67,616, p < .001). The same applies when C2 = /t/: inter-consonantal coordination is tighter when C1 = /l/ than when C1 = /s/ (F (1, 32) = 37, 212, p < .001). This is possibly due to the fact that /l/ is much shorter than /s/ in both types of clusters (88,4 ms for /l/ and 130,11 ms for /s/ when C2 = /t/; 85,23 ms for /l/ and 115,06 ms for /s/ when C2 = /k/). The further one-way ANOVA test suggests that when C2 is a coronal /t/, the durational difference between /s/ and /l/ is statistically signficant (F (1, 31) = 20,444, p < .001), and so does the difference between them when followed by a palatal /k/ (non parametric test on the median data, p < .05).

C1-to-C2 coordination					
C1	*C2*	*Syll.*	*mean (ms)*	*st. dev.*	*N*
s	k	VC$_1$C$_2$	159,6	21,5	18
l		VC$_1$C$_2$	107,9	13,9	16
Total		VC$_1$C$_2$	135,3	31,8	34
s	t	VC$_1$C$_2$	167,5	23	17
l		VC$_1$C$_2$	124,3	18	17
Total		VC$_1$C$_2$	145,9	29,9	34
Total	Total	VC$_1$C$_2$	140,6	31,1	68

Table 7.11: *inter-gestural coordination within complex coda as a function of the articulatory property of the leftmost consonant sorted by C2, excluding the tokens with /ks/.*

C1-to-C2 coordination					
C1	*C2*	*Syll.*	*mean (ms)*	*st. dev.*	*N*
s	k	VC$_1$C$_2$	159,6	21,5	18
	t	VC$_1$C$_2$	167,5	23	17
	Total	VC$_1$C$_2$	163,5	22,2	35
l	k	VC$_1$C$_2$	107,9	13,9	16
	t	VC$_1$C$_2$	124,3	18	17
	Total	VC$_1$C$_2$	116,4	17,9	33
Total	Total	VC$_1$C$_2$	140,6	31,1	68

Table 7.12: *inter-gestural coordination within complex coda as a function of the articulatory property of the rightmost consonant sorted by C1, excluding the tokens with /ks/.*

C1-to-C2 coordination				
C2	*Syll.*	*mean (ms)*	*st. dev.*	*N*
k	VC$_1$C$_2$V	151,8	28,2	17
t	VC$_1$C$_2$V	175,8	26,9	18
Total	VC$_1$C$_2$V	164,1	29,7	35

Table 7.13: *inter-gestural coordination within consonant clusters as a function of the articulatory property of the rightmost consonant.*

When C1 =/s/, the fact of having /k/ or /t/ in the C2 position does not affect the C1-C2 coordination degree (F (1, 33) = 1,093, p > .05). However, when C1 = /l/, the coordination with dorsal /k/ is tighter than with the homorganic apical/t/ (F (1, 31) = 8,463, p < .01).

When the consonant clusters are found in the intervocalic position (VCCV) and C1 = /s/, C2 has a significant influence on the inter-gestural coordination of consonants (F (1, 33) = 6,636, p < .05), /k/ showing a shorter latency than /t/ (Table 7.13).

Although the within coda temporal organisation appears to be insentive to the C2-induced effect when the leftmost consonant is an apical /s/, there is nevertheless a tendency, in both word-final and intervocalic clusters, that C1 (/s/ and /l/) has a more cohesive relation with an upcoming dorsal /k/ than with a homorganic apical /t/.

7.2.3 Internal temporal organisation in /sk/ and /ks/ coda clusters

The comparison between the two mirror clusters /ks/ and /sk/ could provide important evidences on how positional differences might interact with the consonant articulatory properties and further regulate the coarticulatory pattern between consonants.

Table 7.14 illustrates the mean /s/-to-/k/ and /k/-to-/s/ inter-gestural timing values. The results show that vowel-adjacent /s/ has a longer gestural stability area than vowel-adjacent dorsal /k/, however such difference is not statistically significant (F (1, 34) = 4,002, p > .05).

	C1-to-C2 coordination		
stimuli	*mean (ms)*	*st.dev.*	*N.*
/bask/	159,6	21,5	18
/paks/	140	35,7	18

Table 7.14: *inter-gestural coordination for /sk/ and /ks/ coda clusters across all the speakers.*

7.3 Summary of the temporal coordination results

This chapter has dealt with the temporal cohesion between vocalic nuclei and following consonantal gestures as measured through a multi-level acoustic-articulatory measure system based on *SynchroLing*. Temporal cohesion has been analyzed for two different phenomena: coordination differences between onset and coda consonants (CV *vs.* VC), and coordination in consonant clusters.

As far as coordination between the nuclear vowel and singleton onset vs. coda consonants is concerned (CV *vs.* VC), we obtained the following findings. The onset-nucleus inter-gestural coordination was much tighter than the nucleus-coda, as expected on the basis of preceding studies on different languages. This difference held valid regardless of variations in vowel quality (/a/ - /i/) and prosodic prominence (except for the fact that, as much as vowels are elongated as an effect of prosodic prominence, the beginning of the gesture the coda consonant was also delayed). Variation depending on the consonant was found to affect the onset-nucleus coordination, to the extent that the lateral turned out to have a tighter temporal relation with the nuclear vowel than the sibiliant and the velar stop; however, this variation did not disconfirm the validity of the general hypothesis about tighter onset-nucleus coordination vs looser nucleus-coda coordination. Although the interaction between vowel quality and consonant property did not meet the prerequisites for an analysis of variance, there is a tendency for the nucleus-coda coordination to vary more according to the variation of the consonant property when the vocalic nucleus involved is /i/ than when it is /a/.

As far as the coordination in consonant clusters is concerned, two measures of gestural cohesion were considered, that is, the cohesion between the vocalic nucleus and the following consonant (V-to-C1) and the cohesion between the two consonants of the cluster (C1-to-C2). We first summarize the findings for the V-to-C1 measure. Segments' identity was found to condition the temporal coordination between the nucleus and the leftmost consonant in the cluster (V-to-C1 cohesion), in particular, the dorsal consonant /k/ showed a significantly tighter temporal coordination with the preceding nucleus than /s/ and /l/. Another aspect which differentiates /k/ from the other consonants is that, although the cohesion with the vocalic nucleus is not strongly affected by the number of consonants in the cluster (VC vs VCC) when C1 is /l/ or /s/, it does change when C1 is /k/, the coordination in VCC being significantly tighter than in VC. When /k/ is the rightmost consonant in the cluster (/sk/ and /lk/), the vowel and the apical /s/ are coordinated in a looser way in /sk/ than in /st/, suggesting a role for cluster homorganicity to increase the temporal cohesion between C1 and the preceding vowel, but at the same time the lateral /l/ did not change its coordination pattern with the preceding vowel in /lk/ as compared to /lt/, thus suggesting that other factors than (or in addition to) homorganicity may play a role. When the cluster is intervocalic (/sta/, /ska/), the V-to-C1 coordination is tighter than when the cluster is in word-final position (/st/, /sk/); however, no difference was found between /sta/ and /ska/, thus suggesting that in this case, cluster homorganicity plays no role.

As to the internal temporal organisation within the clusters (C1-to-C2), the intrinsic duration of the leftmost consonant was naturally found to affect the termporal cohesion with the rightmost consonant, /s/ having a longer gestural stability area and, consequently, a looser temporal coordination with the upcoming consonant (either /t/ or /k/), compared to /l/. However, when each consonant was separately analyzed, the gestural cohesion between the two consonants of the cluster was found to be shorter in /lk/ than in /lt/, in /sk/ than in /st/ (though not significant at the statistical level) and in /ska/ than in /sta/, thus suggesting that the gesture for /k/ can be anticipated during the production of the preceding coronal consonant more than the gesture for another coronal consonant.

The implications of these results will be thoroughly discussed in the final chapter.

Chapter VIII Results (II): Spatial Coordination

8.1 Spatial coordination for singleton consonants

8.1.1 Lingual configurations: UTI evidence

The average lingual configurations for three different consonants included in the study (average of all lingual configurations attested in the acoustic interval of the consonant of each stimulus) - /s/, /l/ and /k/ - in both vowel conditions (when adjacent to /a/ and to /i/, either in onset or coda position) will be displayed below in Figures 8.1 to 8.3, being compared in subject-by-subject fashion[42]. In each of the following figures, the tongue root is located on the left handside and the tongue tip on the right handside of the images. The average tongue profile is represented by the bold solid line in the centre, whereas the upper and lower lighter lines specify the variance, say, the standard deviation of the curves, the larger the space between them, the larger is the variation on the lingual configuration during the realisation of the articulatory gestures. It might also be the case that the overall tongue profile dose not vary a lot, but only a part of it shows a huge variance (the tongue profile in the pharyngeal region, for example).

---

[42] For the inspection on the lingual configurations, we highlight also the between subject differences in cases of both singleton coda and consonant clusters for the reason that each subject is supposed to have the distinct anatomic properties of her (or his) tongue. Moreover, by showing the lingual profiles in a speaker-by-speaker fashion, we are able to capture the different intra-speakers' articulatory strategies as well.

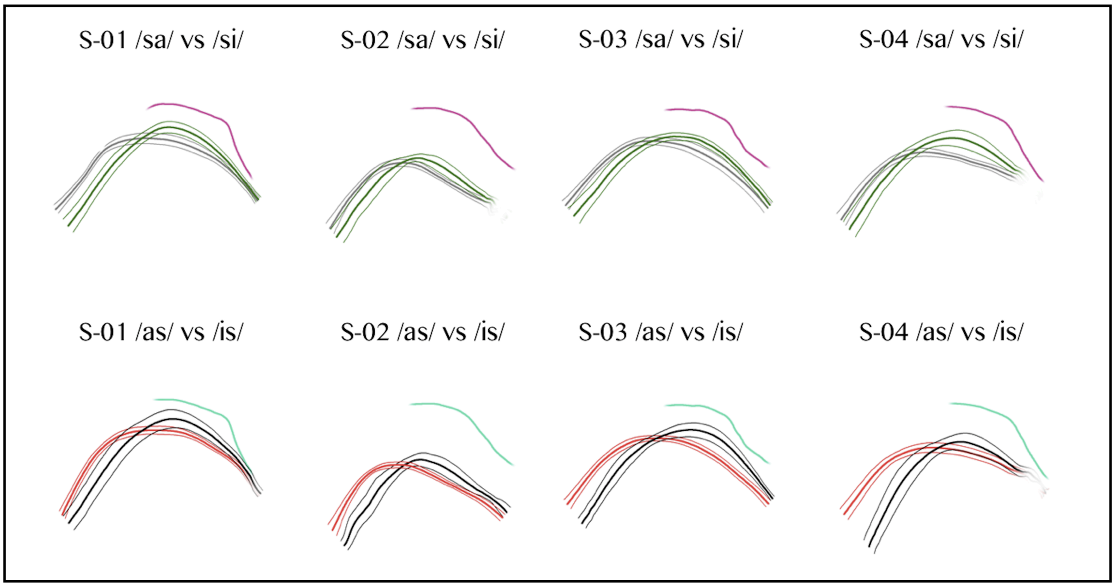


Figure 8.1: *comparison between average tongue configurations (with standard variations) of /s/ in both onset and coda position when adjacent to /a/ and /i/ for the four speakers of the study. In the upper part, the profile of the artificial palate in violet, /s/ in /si/ in dark green, and /s/ in /sa/ in grey; in the lower part, the artificial palate in turquoise, /s/ in /is/ in black, and /s/ in /as/ in red.*

Figure 8.1 shows the average lingual configurations during the production of /sa/ and /si/ (upper part) and /as/ and /is/ (lower part) across subjects. For all of them, the mid-sagittal tongue profile is always higher and more advanced in /si/ as opposed to /sa/; the difference can be observed in the pre-dorsal and dorsal areas. Three out of four speakers do not distinguish their tongue configuration in the anterior region of the tongue: tongue blade and tongue tip. Only Subject 04 (hereinafter S04) appears to neatly differentiate the anterior tongue configuration during the production of the sibilant /s/ in /sa/ *vs*. /si/, with a relatively higher position in /si/ than in /sa/, shown in the rightmost image in the upper panel of Figure 8.1.

As far as coda /s/ is concerned, the overall advancement of the tongue profile, when to the right of /i/, catches the eye for all the speakers. In particular, S03 shows the most evident anterior tongue advancement with respect to the other speakers (S02 comes seconde in this). Note that for some of the speakers the tongue tip is raised to a very limited extent, in both /a/ and /i/ contexts.

These results therefore reveal that, although /s/ is articulated with the lateral sides of the tongue raised towards the alveolar ridge, and the medial part of the tongue and the tongue tip slightly lowered forming a groove to allow the airstream to be able to escape from the front, there are nevertheless significant differences in the mid-sagittal profile as far as the dorsal and pre-dorsal area are concerned. In addition, there might also be the speaker-based differences in the anterior part of the tongue: the (pre-)dorsal height found for S04 when the onset /s/ is concerned; the (pre-)dorsal advancement for S03 and S02 when /s/ is in coda.

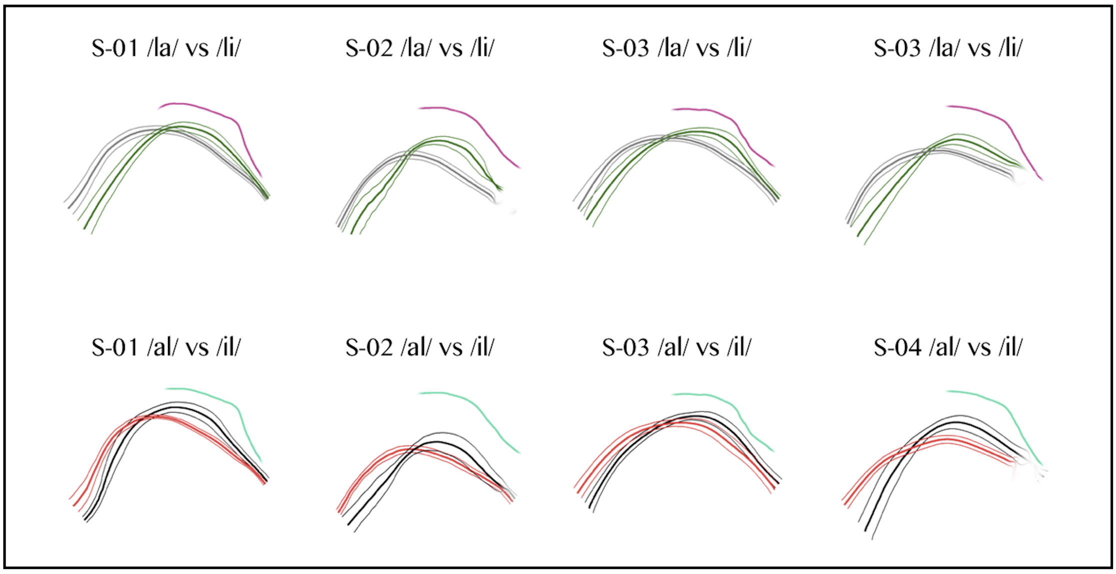


Figure 8.2: *comparison between average tongue configuration (with standard variations) of /l/ in both onset and coda position when adjacent to /a/ and /i/ for the four speakers of the study. In the upper part, the artificial palate in violet, /l/ in /li/ in dark green, and /l/ in /la/ in grey; in the lower part, the artificial palate in turquoise, /l/ in /il/ in black, and /l/ in /al/ in red.*

Figure 8.2 displays average overall tongue configurations of the gesture /l/ during the articulation of /la/ *versus* /li/ in the upper part, and /al/ *versus* /il/ in the lower part. As in the ultrasound tongue images for /s/, the tongue tip gesture is sometimes hardly visible. When /l/ is in onset, for all the speakers, its overall tongue profile is in a higher and more advanced position in the vowel context of /i/ than in the one of /a/. Moreover, the tongue tip, where exactly the constriction for /l/ occurs, is significantly higher in only two of the speakers: S02 and S04. When it is in coda, tongue tip differences might only be recognised for S04. These results tend to suggest that there is slightly less coarticulatory variations on /l/ when in coda than in onset position.

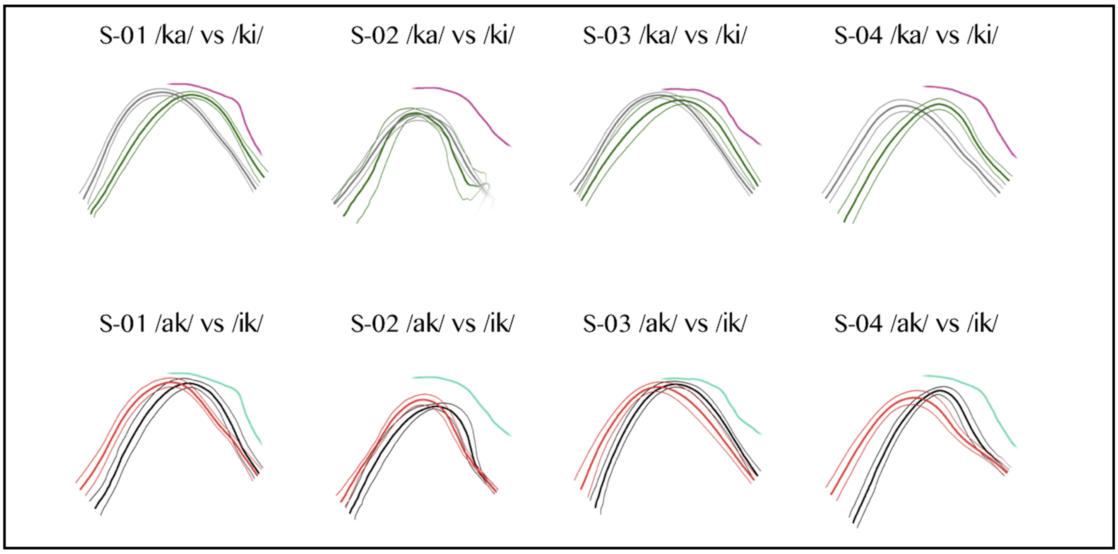


Figure 8.3: *comparison between average tongue configuration (with standard variations) of /k/ in both onset and coda position when adjacent to /a/ and /i/ for the four speakers of the study. In the upper part, the artificial palate in violet, /k/ in /ki/ in dark green, and /k/ in /ka/ in grey; in the lower part, the artificial palate in turquoise, /k/ in /ik/ in black, and /k/ in /ak/ in red.*

Figure 8.3 gives the average tongue configurations of /k/ as it functions as both onset (upper part) and coda (lower part) when adjacent to the vowels /a/ and /i/ for all the speakers. As far as the onset position is concerned, three speakers out of four show fairly different tongue profiles in two vowel contexts. In /i/ context, the overall profile of /k/ is more advanced than in /a/ context. However, it seems that S02 does not differentiate the tongue configurations in those phonetic contexts. If one takes a closer observation on her realisation, the constriction locations of /k/ in two cases appear to overlap (see the second image from the left in the upper panel of Figure 8.3).

The tongue profile patterns for /k/ in coda position is similar to that for /k/ in onset. In this case, all four speakers show the same pattern, that is, the overall tongue profile considerably advances when adjacent to /i/. Thus we can conclude that the dorsal consonant /k/ is realised in a more advanced tongue position when adjacent to /i/ than to /a/ regardless of syllable constituency, despite the fact that S02 fails to distinguish onset /k/ in different vowel conditions.

8.1.2 EPG indices

In the present section, we will examine the EPG indices for all three consonants in the study in order to see the coarticulatory patterns (in other words, spatial coordination) in different vowel contexts. The values of EPG indices for different consonants in different vocalic contexts could further provide complementary evidence on how inter-gestural temporal coordinations could be conditioned by interaction between vowel quality and the consonants' articulatory properties. As anticipated in Chapter 6, each of the singleton consonants is attributed one EPG index: CAa values for /s/ and /l/, and Qp values for /k/. It should be underlined that the stimuli with the nuclear vowel /a/ bearing the pitch accent of the type of corrective focus are excluded from the present study.

The CAa value is expected to be higher when the consonant /s/ is found adjacent to /i/ than when it is adjacent to /a/, because during articulation of a high vowel the tongue position is more fronted than during the production of a low vowel. Table 8.1 shows the average CAa values of /s/ as a function of vocalic contexts and its position in the syllable (*i.e.*, onset /s/ *vs*. coda /s/).

average CAa values for /s/				
Syll.	*V*	*Average CAa*	*st. dev.*	*N*
CV	/a/	,918	,063	17
	/i/	,949	,032	18
	Total	,934	,051	35
VC	/a/	,958	,024	18
	/i/	,953	,039	18
	Total	,955	,032	36

Table 8.1: *average CAa values for /s/ in CV and VC sequences as a function of vowel quality across the speakers, excluding the stimuli with the pitch accent of the type of corrective focus.*

The data shown in the table fail to confirm the expectation of higher CAa values in the context of /i/ than in the context of /a/. The Mann-Whitney with the average CAa values of /s/ in various contexts as dependent variable and vowel quality as with-subject factor reveals that the CAa values do not differentiate the categories dipending on syllabic constituency and vocalic contexts: when /s/ is in syllable onset position, U = 201, z = 1,854, *p* > .05; when /s/ is in coda position, U = 155,500, z = -,206, *p* > .05. It could thus be confirmed that the gesture for /s/ as measured by the CAa index turns out to be insensitive to variation of quality of the adjacent vowel.

However, the results show that there is a significant effect caused by position in the syllable. Coda /s/ has a relatively higher average CAa value than onset /s/ (The Mann-Whitney with the average CAa values as dependent variable and syllabic constituency of /s/ as with-subject factor: U = 817,500, z = 2,156, *p* < .05). Thus, the overall linguo-palatal contact of /s/ is more fronted in coda position than in onset position.

A similar assumption for the CAa values of /l/ is made when it is adjacent to different vowels — a higher average CAa value in the context of /i/ than in the context of /a/. The following table 8.2 illustrates the average CAa values of /l/ as a function of vowel quality and the its syllabic affiliation (*i.e.*, onset /l/ *vs*. coda /l/).

average CAa values for /l/				
Syll.	*V*	*Average CAa*	*st. dev.*	*N*
CV	/a/	,951	,036	18
	/i/	,984	,017	16
	Total	,967	,033	34
VC	/a/	,985	,025	17
	/i/	,982	,023	18
	Total	,983	,024	35

Table 8.2: *average CAa values for /l/ in CV and VC sequences as a function of vowel quality across the speakers, excluding the stimuli with the pitch accent of the type of corrective focus.*

The data for /l/ confirms expectation that the value of CAa is higher when /l/ is adjacent to /i/ than when it is adjacent to /a/ only for onset /l/ (U = 221, z = 2,657, *p* < .05). In VC sequences, the CAa value does not change in a significant way across the vocalic contexts (U = 144, z = -,297, *p* > .05). The results thus suggest that the anteriority of the linguo-palatal contact during articulation of onset /l/ is sensitive to the variation on the context of vowel quality. In particular, there is more /i/-to-/l/ spatial overlap than /a/-to-/l/.

As far as syllabic constituency is concerned (*i.e.*, onset /l/ *vs*. coda /l/), similar to the situation for /s/, the average CAa value is significantly higher when the /l/ is in coda than when it is in onset (U = 770,500, z = 2,107, *p* < .05). The linguo-palatal contact during the production of /l/ is more fronted for coda /l/ than for onset /l/.

For the gesture of /k/, the Qp index is concerned. The Qp value is expected to be higher when /k/ is adjacent to /i/ than when it is adjacent to /a/ because there is more dorsal activity during the production of a high, palatal vowel. Table 8.3 gives the average Qp values of /k/ as a function of vowel quality and syllabic constituency (*i.e.*, onset /k/ *vs*. coda /k/).

average Qp values for /k/				
Syll.	*V*	*Average Qp*	*st. dev.*	*N*
CV	/a/	,105	,036	18
	/i/	,283	,054	16
	Total	,189	,101	34
VC	/a/	,105	,035	16
	/i/	,252	,050	17
	Total	,181	,086	33

Table 8.3: *average Qp values for /k/ in CV and VC sequences as a function of vowel quality across the speakers, excluding the stimuli with the pitch accent of the type of corrective focus.*

The results are consistent with the expectations in both CV sequence and VC sequence. The Qp value is significantly higher in the context of the high vowel /i/ than of the low vowel /a/. As to the effect of syllabic affiliation (*i.e.*, onset /k/ *vs*. coda /k/), the ANOVA analysis with the QP value as the dependent variable and the syllable position as the within-subject factor shows that the difference of the Qp values across the different syllabic conditions is not statistically significant, thus suggesting that the discrepancy of the Qp values caused by the context of vowel quality is equally present in both CV and VC sequences. Based on these results, it might be conclude that the dorsal activity during the articulation of /k/ is conditioned, regardless of its syllabic constituencity, by the vocalic context in which it is found.

8.1.3 Summary of the coarticulatory results for singleton consonants

The cumulative evidences of UTI lingual configurations and EPG indices regarding the location of lingua-palatal contacts suggest that the coarticulatory patterns and the degrees of coarticulatory resistance of the consonants do vary across different contexts of vowel quality. In addition, the UTI profiles presented above show that coarticulatory activities between the adjacent articulatory gestures could vary across the speakers as well. Moreover, the coarticulatory patterns could be conditioned by syllabic structure, that is, the position in the syllable that a given consonant occupies, which provides additional information about the dynamics of the spatial overlap, in particular, of sibilant /s/ and liquid /l/ in different vowel contexts (/a/ and /i/), so that the articulatory gestures of an onset consonant could anticipate the lingual properties of the following nuclear vowel, and the same consonant could accommodate to the preceding nuclear vowel when it is in coda position.

In specific terms, the sibilant is expected to show a comparatively higher degree of coarticulatory resistance (Recasens, 2012, 2014), *i.e.*, to vary the least as a function of vowel contexts. A higher degree of coarticulatory resistance of sibilant /s/ is confirmed by the linguo-palatal contacts in the anterior palate, inasmuch as no variations is found across different vowel conditions. Nevertheless, the analysis on the mid-sagittal lingual configuration shows that the lingual configuration in the context of /a/ is slightly different from that in the context of /i/. Thus, the tongue might adapt somehow to the position required for the adjacent vocalic gesture (by reaching an overall more anterior part in the context of /i/, and for some speakers, a higher blade and tongue tip, see Figure 8.1), at the same time leaving a central groove in the anterior palate in order to guarantee the achievement of the target constriction characterising the gesture for /s/.

There is, however, syllable-induced variation, with the linguo-palatal contact more fronted (both on average and for some specific speakers) in coda than in onset.

The lateral consonant /l/ is expected to be less resistant to coarticulatory effect than sibilant /s/, thus exhibiting more variation as a function of vowel quality. The expectation is met, inasmuch as the linguo-palatal contact varies in a statistically significant way in two different vowel contexts, in particular when the lateral is in onset position. However, the lateral is similar to the sibilant in two respects. First, the vowel-induced variation on the mid-sagittal profiles are successfully captured by the Ultrasound machine. The UTI for both consonants shows an overall more anterior tongue configuration in the context of /i/. For a subset of the speakers, a relatively higher blade and tongue tip could be observed in their realisations. Interestingly, although this vowel-induced variation is proven to influence the anteriority of linguo-palatal contact in a significant way only for the lateral consonant /l/, it is consistent with the observations that /l/ is always articulated with the central apical contact. Second, a syllable-induced variation is found for both consonants, being more fronted in coda than in onset.

The dorsal consonant /k/ exhibits a consistently flexible lingual configuration and a statistically significant change regarding of the EPG values of anteriority as a function of vowel quality. During the production of /k/, the lingual configurations are overall more advanced when adjacent to /i/ than to /a/ regardless of the position in the syllable. The analysis further confirms that there is no impact on the coarticulatory patterns, which are strongly modified by the requirements for the vowel, no matter where it is found, preceding or following /k/.

We could therefore conclude that, on the basis of the linguo-palatal contact analysis, with the additive evidence from the mid-sagittal tongue images, a hierarchy regarding the degree of coarticulatory resistance could be established for the present three consonants with sibilant /s/ exhibiting the highest resistance, dorsal /k/ the lowest and lateral /l/ an intermediate degree of coarticulatory resistant.

8.2 Spatial coordination for consonant clusters: EPG indices

In the following section, we will present the EPG indices for three consonants in various contexts, including their syllabic affiliation and the different articulatory environments where they are found. Afterwards, the UTI tongue profile evidence as a function of the speakers will be provided.

8.2.1 C1 as a function of segmental complexity

We first verify whether the EPG indices of an individual consonant vary as the degree of segmental complexity increases from singleton to branched coda and intervocalic consonant cluster. The relative EPG values are expected to be higher in singleton coda than in complex coda, at least for the two pre-dorsal consonants (/s/ and /l/), they be more or less resistant to the coarticulatory effects by the adjacent gestures, since they are supposed to make some compromise in their articulatory space according to the articulatory environment created by the rightmost consonant in the cluster. The EPG indices for the three consonants in various conditions of syllabic affiliation are presented in Table 8.4.

EPG values as a function of Syll.

C1	*Syll.*	*mean value*		*st. dev.*	*N*
/s/	VC	CAa	,959	,025	18
	VCC		,911	,086	35
	VCCV		,918	,074	34
	Total		,923	,074	87
/l/	VC	CAa	,985	,025	17
	VCC		,948	,089	33
	Total		,961	,075	50
/k/	VC	Qp	,105	,035	16
	VCC		,104	,043	18
	Total		,105	,039	34

Table 8.4: *mean EPG values for /s/, /l/ and /k/ in VC, VCC and VCCV sequences as a function of segmental complexity to the right of the stressed vowel across the speakers, excluding the stimuli with the pitch accent of the type of corrective focus.*

The mean value of CAa is chosen as an EPG index to verify the anteriority regarding of the linguo-palatal contact during the production of the lateral /l/. Although the CAa value reveals that the constriction location appears to be a little bit more fronted in the singleton condition than in branched coda (0,985 vs. 0,948), which suggests the articulatory gesture for /l/ might adapt somehow to the following C gesture with a backward constriction location, the non parametric test on the median values across all syllabic constituency shows that such difference is not statistically significant ($p > .05$).

The CAa value of the apical fricative is selected as an index for the constriction location of the gesture for /s/. The non parametric test on the median values shows that the linguo-palatal contact location does not change as the variations on the coda complexity ($p > .05$). However, the results on the mean CAa value in three different conditions of syllabic affiliation (see Table 8.4 above) seem to draw a sharp distinction between the singleton coda /s/ and /sC/ clusters (including both word-final and intervocalic consonant clusters). Furthermore, the post-doc LSD (least significance difference) test suggests that the singleton coda /s/ contrasts with the word-final /sC/ clusters in a significant way ($p < .05$).

For the dorsal consonant /k/, the mean value of Qp remains rather stable across the conditions of coda complexity. The ANOVA analysis with the Qp value as dependent variable and coda structure as within-subject factor confirms that the constriction location for /k/ does not change as a function of coda complexity (F (1, 32) = ,006, $p > .05$).

In order to get a full better understanding of how the constriction location displacement for the apical fricative gesture is conditioned as a function of segmental complexity (as the leftmost C in a branched coda *vs*. followed by another CV sequence) in different articulatory environments (as followed by the homorganic consonant /t/ *vs*. the non-homorganic consonant /k/), we further verify the EPG indices for the leftmost

/s/ in the consonant cluster as a function of the articulatory properties of the rightmost consonant in the cluster (*i.e.*, /t/ and /k/). The results are presented in Table 8.5.

mean values for C1 (/s/)									
C1	*C2*	*Syll.*	*mean CCa*	*st. dev.*	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/s/	/k/	VCC			,912	,091	,143	,024	18
		VCCV			,909	,093	,145	,018	17
		Total			,910	,090	,144	,021	35
	/t/	VCC	,483	,182	,910	,084			17
		VCCV	,515	,105	,927	,051			17
		Total	,499	,148	,918	,069			34

Table 8.5: *mean EPG values for /s/ in VCC and in VCCV sequences as a function of its syllabic affiliation sorted by the rightmost consonant in the cluster across the speakers.*

i) /bask/ *vs*. /baska/

For /s/ in /bask/ *vs.* /baska/, we test the CAa and Qp values for /s/ to see whether anteriority and posteriority of the articulatory properties may change in these different conditions of segmental complexity. The ANOVA test with the CAa value as dependent variable and syllabification as within-subject factor shows that the anteriority of the constriction location for /s/ is unaffected by the variations on post-laminal segmental complexity, F (1, 33) = ,057, $p > .05$. As far as the Qp index for /s/ is concerned, the result reveals that the dorsal area of the tongue is equally 'coarticulated' with the following /k/ during the production of the apical consonant regardless of the fact whether /sk/ is word-final or intervocalic cluster (F (1, 33) = ,008, $p > .05$).

ii) /bast/ *vs*. /basta/

Now we turn to /s/ followed by an homorganic occlusive /t/, the CAa and CCa indices are used. Since /t/ is articulated with a slightly more advanced constriction location and full closure in the tongue tip, the former value regards the anteriority of the constriction location and the latter one is responsible for the centrality of the anterior closure.

Neither of these two articulatory properties appears to be affected by two conditions of postvocalic segmental complexity (*i.e.*, stressed vowel /a/ followed by an /sC/ sequence *vs*. by an /sCV/ sequence) in a significant way. The ANOVA analysis on CAa values suggests that the constriction location of /s/ is insensitive to the variation on the segmental structures when followed by a /t/ (F (1, 32) = ,514, *p* > .05), the non-parametric test on the median of CCa values reveals the similar result (the centrality of the constriction in the anterior tongue for /s/ does not vary across the two conditions, *p* > .05).

8.2.2 C2 as a function of segmental complexity

In this section we verify whether the articulatory properties of the rightmost consonant in the cluster may be modified by the phonotatic context. The three EPG indices are again selected: CAa and Qp for the dorsal consonant /k/, CAa and CCa for the (dental-) alveolar consonant /t/. All values are illustrated in Table 8.6.

mean values for C2 (/k/ and /t/)									
C1	*C2*	*Syll.*	*mean CCa*	*st. dev.*	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/s/	/k/	VCC			,774	,216	,206	,048	18
		VCCV			,686	,192	,188	,037	17
		Total			,731	,207	,197	,044	35
	/t/	VCC	,597	,116	,958	,034			17
		VCCV	,740	,080	,989	,009			17
		Total	,669	,122	,974	,029			34

Table 8.6: *mean EPG values for /t/ and /k/ in VCC and in VCCV sequences as a function of its syllabic affiliation in the cluster across the speakers.*

i) C2 = /k/: /bask/ *vs*. /baska/

We first check the Qp values for the constriction location of /k/. The results of the ANOVA analysis fail to discriminate the two /k/s in a significant way: F (1, 33) = 1,537,

p > .05. The primary constriction location for /k/ is invariated in both segmental conditions regardless of whether or not it is followed by /a/. As to the anterior tongue region, the non-parametric test on the median of CAa values draws a clear distinction between the two cases (p < .05), suggesting that the pre-dorsal region undergoes different /s/-induced influence during the formation of dorsal constriction in these two segmental conditions.

ii) C2 = /t/: /bast/ *vs*. /basta/

Although (dento-) alveolar /t/ is not one of three consonants directly involved in our study, the EPG indices regarding the articulatory properties in different segmental and prosodical[43] contexts could provide us complementary evidences for the preceding /s/ on both the anteriority and centrality value of the constriction, especially in the case where neither of the index of /s/ is itself able to tell the difference in the two phonotatic patterns considered.

We can find in Table 8.6 above that the CAa value for /t/ in VCCV sequences is higher than that in VCC sequences (0,989 *vs*. 0,958). A non-parametric test on the median values confirms that the differences between the two cases is significant (p < .01), that is, the constriction location for /t/ is significantly more advanced when /t/ is the rightmost consonant of an intervocalic cluster. As far as the centrality of constriction is concerned, the ANOVA analysis on CCa values gives a similar result: the constriction is significantly more central in intervocalic consonant clusters than in branched codas (F = (1, 32) = 17,527, p < .001).

8.2.3 /s/-series complex coda: /bask/ *vs*. /bast/

We now evaluate the coarticulatory patterns between the consonant gestures in the /s/+C clusters. The EPG indices for /s/ in different consonantal contexts (/k/ *vs*. /t/) are presented in Tables 8.7 and 8.8:

---

[43] The prosodical context refers to the syllabic condition of the consonant: being singleton onset of the rightward syllable, or being a part of the branded coda in the stressed syllable.

mean values for C1 (/s/) as a function of C2 in VCC

C2	*mean CCa*	*st. dev.*	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/k/	,536	,162	,912	,091	,143	,024	18
/t/	,483	,182	,910	,084	,144	,021	17
Total	,510	,172	,911	,086	,144	,022	35

Table 8.7: *mean EPG values for /s/ in branched coda as a function of the articulatory properties of C2 across the speakers.*

mean values for C1 (/s/) as a function of C2 in VCCV

C2	*mean CCa*	*st. dev.*	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/k/	,506	,174	,909	,093	,145	,018	17
/t/	,515	,105	,927	,051	,145	,012	17
Total	,511	,142	,918	,074	,145	,015	34

Table 7.8: *mean EPG values for /s/ in intervocalic clusters as a function of the articulatory properties of C2 across the speakers.*

The statistical results show that in both types of stimuli (VCC, VCCV), the three EPG indices of /s/ do not vary according to the variation of the subsequent consonant (/t/, /k/) (Table 8.9).

C_1 = /s/	**VC_1C_2: /bask/ *vs.* /bast/**	**VC_1C_2V: /bas.ka/ *vs.* /bas.ta/**
CAa	F (1, 33) = ,006, p > ,05	F (1, 32) = ,485, p > ,05
Qp	F (1, 33) = ,012, p > ,05	F (1, 32) = ,000, p > ,05
CCa	F (1, 33) = ,819, p > ,05	non-parametric test on median: p > ,05

Table 8.9: *statistical comparison between VCC and VCCV stimuli, three EPG indices.*

8.2.4 /l/-series complex coda: /balk/ *vs.* /balt/

This section investigates whether the constriction location for the lateral is altered by the articulatory properties of the following consonants and the eventual differences between lateral /l/ and laminal /s/ in the identical articulatory environments.

mean values for C1 (/l/) as a function of C2 in VCC					
C2	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/k/	,907	,112	,099	,026	16
/t/	,987	,028	,066	,028	17
Total	,948	,089	,082	,031	33

Table 8.10: *EPG indices for /l/ as a function of C2 in VCC sequence across the speakers.*

Table 8.10 shows that for the CAa index, which is responsible for the primary constriction location for the lateral, the average value for the /l/ is relatively higher when it is adjacent to /t/ than to /k/. A non-parametric test on the median confirms that such difference is statistically significant, suggesting that the constriction location of /l/ is more advanced before a pre-dorsal consonant (thus, before a homogenous articulator) than before a dorsal consonant.

The mean Qp values for /l/ potentially reveals the degree to which the dorsal constriction is anticipated during the production of the lateral consonant. The ANOVA shows that Qp is significantly different before /t/ *vs.* /k/ (F (1, 31) = 12,361, $p < .005$), which means that the tongue dorsum has prepared for movement towards the palate during the production of the lateral consonant.

8.2.5 /bask/ vs. /balk/

We further verify how the articulatory configuration of the dorsal consonant occupying the second position in a branching coda varies as a function of the previous consonant (Table 8.11).

mean values for C2(/k/) as a function of C1 in VCC					
C1	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/s/	,774	,216	,206	,048	18
/l/	,487	,325	,180	,045	16
Total	,639	,305	,194	,048	34

Table 8.11: *EPG indices for /k/ as a function of* C_1 *in VCC sequence across the speakers.*

An ANOVA with the Qp values as dependent variable and the articulatory property of the vowel-adjacent consonant as within subject factor suggests that the dorsal constriction is homogenous across the two contexts (F (1,32) = 2,740, $p > .05$). The anterior tongue, by contrast, seems to be more affected by the anteriority of apical /s/ than by that of the lateral consonant /l/ during the production of dorsal /k/; a non-parametric test on the distribution by Mann-Whitney U test confirms that CAa is significantly higher before /s/ than before /t/ ($p < .005$).

8.2.6 /bask/ *vs.* /paks/

In this section, we verify the coarticulatory patterns between the tongue blade gesture (/s/) and the dorsal gesture (/k/) within two complex codas whose segmental sequences are mirror images of each other: /sk/ and /ks/ (Table 8.12).

/bask/ vs. /paks/						
C	*stimuli*	*mean CAa*	*st. dev.*	*mean Qp*	*st. dev.*	*N*
/s/	/bask/	,912	,091	,143	,024	18
	/paks/	,944	,048	,159	,023	18
	Total	,928	,074	,151	,025	36
/k/	/bask/	,774	,216	,206	,048	18
	/paks/	,277	,289	,104	,043	18
	Total	,525	,356	,155	,069	36

Table 8.12: *EPG indices for /s/ and /k/ in the stimuli /bask/ and /paks/ across the speakers.*

We evaluate the primary constriction location in the first place. For laminal /s/, the average CAa value remains rather stable, irrespective of coda composition (F (1,34) = 1,776, $p > .05$). For /k/, the average Qp value is confirmed to be different in the two case, that is, the dorsal constriction is more posterior in /ks/ than in /sk/ (F (1,34) = 44,654, $p < .001$).

For the tongue regions that are not directly involved in the formation, we find that the tongue pre-dorsal area for the dorsal consonant /k/ is more advanced (that is, it adapts to the pre-dorsal configuration for /s/) to a larger extent in /sk/, compared to /ks/ (F (1,34) = 34,084, $p < .001$). This is in line with the observation above (concerning Qp) that /k/ is less posterior in the /sk/ context. Concerning the dorsal area during the production of laminal /s/, a non-parametric test on the median of Qp reveals that no difference is statistically significant between the dorsal activity in /sk/ as compared to /ks/.

8.3 Spatial coordination for consonant clusters: lingual configurations

The average lingual configurations for the three different consonantal gestures included in the current study in various coda conditions are presented in this section, according to a subject-by-subject analysis. In all figures, the tongue tip is located on the right handside and the tongue root on the left.

8.3.1 /s/ in postvocalic word-final consonant clusters /bask/ *vs.* /bast/

Figure 8.4 shows the comparison between the average lingual configurations during the production of the laminal consonant /s/ in the branched codas /sk/ and /st/. None of the speakers distinguishes the lingual configuration of /s/ in the pre-dorsal area in the two articulatory environments. As for the dorsal area, three speakers out of four (S01, S02 and S03) realise /s/ with a relatively higher tongue position in /sk/ than in /st/.

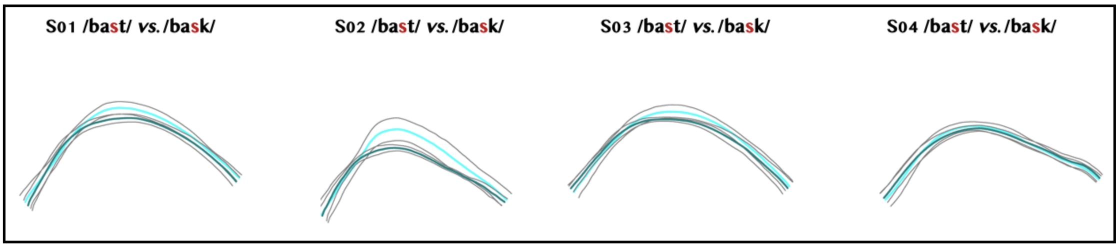


Figure 8.4: *average tongue configurations (with standard variations) of /s/ followed by /k/ (light blue) and by /t/ (green) in branching codas.*

8.3.2 /s/ in intervocalic consonant clusters /baska/ *vs.* /basta/

Figure 8.5 shows the average tongue configurations of /s/ in two intervocalic clusters. S01 and S03 do not distinguish the height nor the anteriority of the apical constriction; S04 shows a slightly higher tongue tip position in the pre-dorsal area for /s/ in /basta/ and S02 shows an overall higher tongue position for /s/ in /baska/. In the dorsal area, two patterns are observed: on the one hand, S01, S02 and S03 have a higher tongue position when /s/ is followed by /k/ (also found in branching codas, see the section 8.3.1 above); on the other hand, S04 has an overall more fronted and higher tongue profile in /basta/ than in /baska/, which might further explain the different tongue positions in the pre-dorsal areas between S02 and S04.

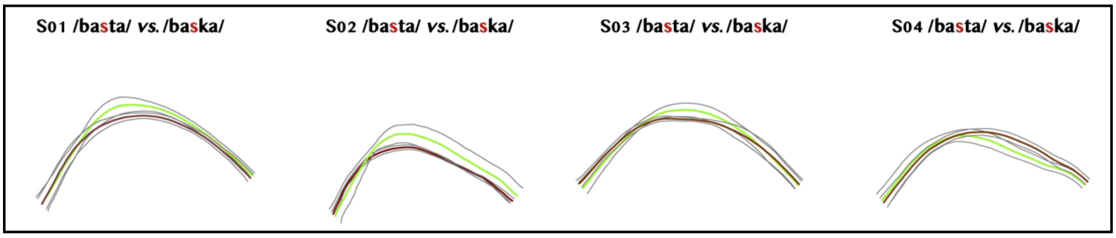


Figure 8.5: *average tongue configurations (with standard variations) of /s/ when followed by /k/ (green) and by /t/ (brown) in intervocalic consonant cluster.*

8.3.3 /l/ in postvocalic word-final clusters /balk/ *vs.* /balt/

A clear distinction between the tongue profiles of /l/ after /k/ and after /t/ is shown in Figure 8.6. The effects are fairly clear in the realisations of S01, S02 and S04 with a much higher tongue position in the dorsal area. Some differences are observed in the realisation of S03 as well, but to a smaller degree. These data should be considered in conjunction with the results of the EPG index CAa, which also reported large differences as far as the anteriority of the constriction was concerned (see above Table 8.10). It should be recalled that /l/ being articulated with the lateral regions of the tongue curled up, the midsagittal profile as revealed by the UTI analysis is not directly informative on the exact position of the constriction for this consonant.

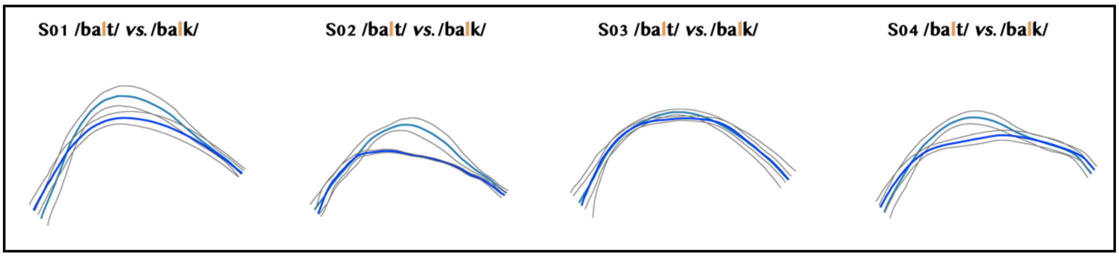


Figure 7.6: *average tongue configurations (with standard variations) of /l/ when followed by /t/ (blue) and by /k/ (azure) in coda cluster.*

8.3.4 /s/ *vs.* /l/: /bask/ and /balk/

In Figure 8.7, we compare the average lingual configurations of lateral /l/ and laminal /s/ when they are followed by dorsal /k/ in branching codas. The focus is on the dorsal area, to see whether the lingual area that is not directly involved in the constriction undergoes anticipatory effects from the following consonant. The dorsal configuration of /l/ is generally higher (S01, S02 and S04) and more fronted (S02 and S04); the relatively more posterior dorsal area for /s/ is consistent with the higher Qp values which were reported in the previous section (the section 8.2.5) for this consonant.

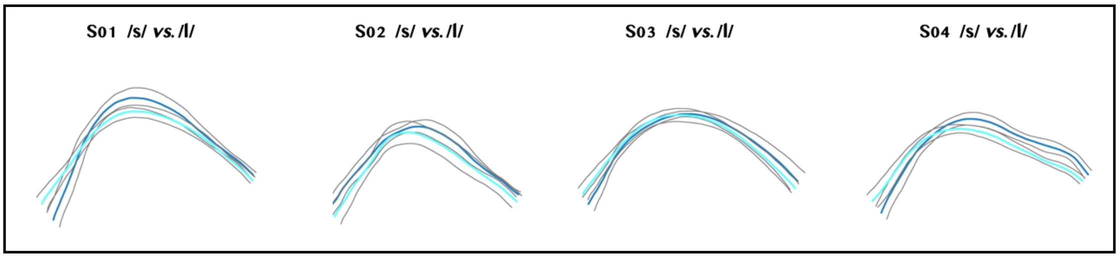


Figure 8.7: *average tongue configurations (with standard variations) of /s/ (light blue) and /l/ (dark blue) when followed by /k/ in coda cluster.*

8.3.5 /s/ *vs.* /l/: /bast/ and /balt/

A comparison is also shown for the lingual configuration of /s/ and /l/ when they are followed by the apical consonant /t/ (Figure 8.8). In all speakers a higher pre-dorsum can be found for the lateral (especially in S03 and S04); moreover, the shape of the /s/ profiles is much more consistent across repetitions (as revealed by the smallest standard deviation area). These observations suggest that /s/ is more constraint in the pre-dorsal area than /l/. This might be explained by the fact that a central groove is formed during the laminal constriction for /s/, which allows the airflow from the lungs to travel continuously through the oral cavity, and this behaviour requires the fine control of all muscles in the pre-dorsal area.

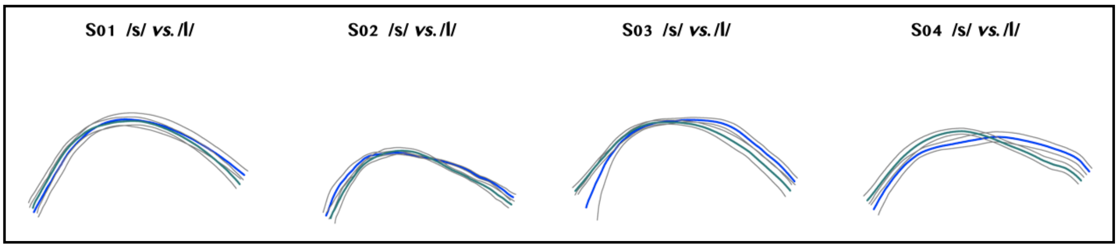


Figure 8.8: *average tongue configurations (with standard variations) of /s/ (green) and /l/ (blue) when followed by /t/ in coda cluster.*

8.3.6 /k/ in postvocalic word-final clusters /bask/ *vs.* /balk/

The average lingual configuration of /k/ in different articulatory environments is presented in Figure 8.9. Three speakers out of four (S01, S02 and S04) neatly distinguish the tongue configuration in the dorsal constriction as a function of the preceding consonants; S01 produces the /k/ with a higher tongue position in the dorsal area when preceded by the lateral; S04 produces the same gesture with the dorsal constriction more fronted; and S02 chooses a hybrid strategy to vary the tongue configurations in both vertical and horizontal dimensions - higher and more fronted.

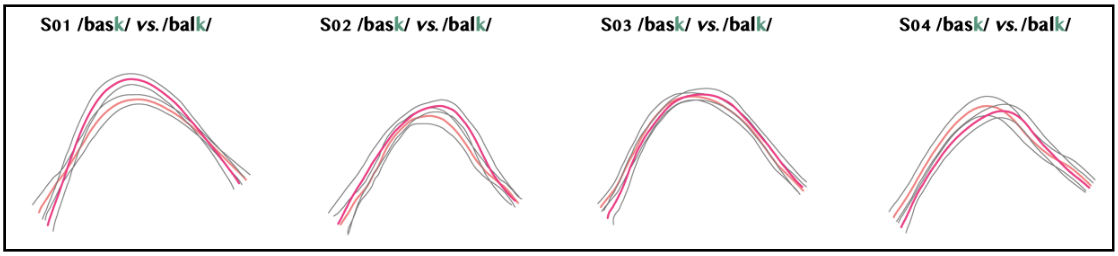


Figure 8.9: *average tongue configurations (with standard variations) of /k/ when preceded by /s/ (red) and by /l/ (pink) in coda cluster.*

8.3.6 /s/ in /bask/ and /paks/

The average tongue profiles in Figure 8.10 reveal that the apical constriction of laminal /s/ appears to be quite stable when /s/ is adjacent to /k/, regardless of whether it is preceded or followed by dorsal /k/. All of the four speakers show almost the same tongue position in the anterior tongue. The realisations of /s/ by S01, S03 and S04 are characterised by a higher tongue position in the dorsal area; moreover, the configurations are more fronted for S03 and S04. S02, on the contrary, has a lower tongue position in the dorsal area for /s/ in /paks/ than in /bask/.

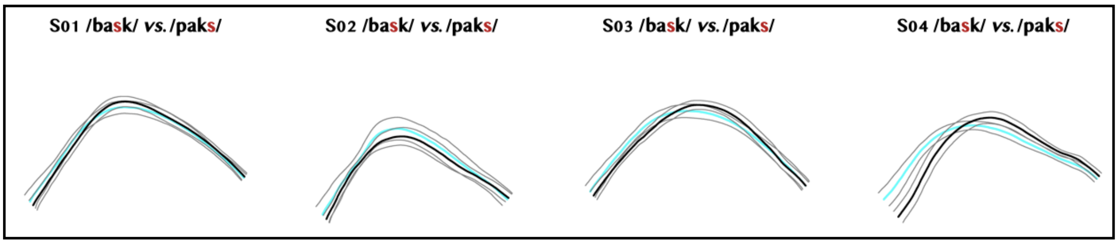


Figure 8.10: *average tongue configurations (with standard variations) of /s/ when followed by /k/ (black) and precede by /k/ (light blue) in coda cluster.*

8.3.7 /k/ in /bask/ and /paks/

Consistently with the relative stability of the apical constriction of /s/ in /bask/ and /paks/ (Figure 8.10 above), the dorsal constriction of /k/ also appears to exhibit very few variability in the same clusters (Figure 8.11). Some speakers show a slightly more fronted realisation (S01, S02 and S03). A relatively higher tongue pre-dorsum in the tongue profiles in /bask/ is also visible, but all these effects are of a very little magnitude.

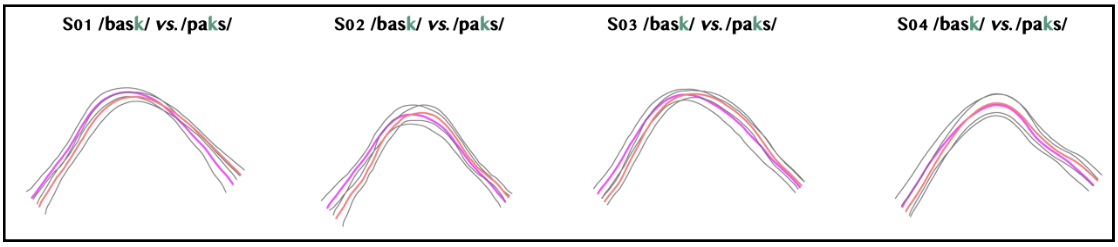


Figure 8.11: *average tongue configurations (with standard variations) of /k/ when followed by /s/ (pink) and precede by /s/ (red) in coda cluster.*

7.4 Summary of the spatial coordination for consonant clusters

The coarticulatory patterns for consonant clusters were expected to exhibit a more complicated scenario with respect to singleton codas, in the sense that consonants might undergo the coarticulatory influence of an adjacent consonant at the level of constriction location but also depending on phontatics. As a matter of fact, like in singleton codas, EPG indices and UTI configurations suggest that the coarticulatory patterns between the consonants with various degrees of coarticulatory resistance vary according to the articulatory environment in which they are found. The UTI data additionally suggest that specific articulatory strategies are adopted by the individual speakers.

Positive evidence in support of the hypothesis has been found in the data.

Coda complexity was indeed found to impact on the constriction location only for laminal /s/, but not for the lateral consonant /l/ and the dorsal consonant /k/: the apical constriction is more fronted for a singleton coda /s/ than for an /s/ as C1 in a branching coda. Again, the linguo-palatal constriction of /t/ when it is C2 in a cluster varies as a function of segmental composition in both anteriority and centrality when preceded by a homorganic /s/, that is, /t/ has a more fronted and centralised constriction in /VsCV/ sequence than in /VsC/ sequence.

Concerning the differences among the three consonants of the study, the laminal fricative /s/ was supposed to have the highest degree of coarticulatory resistance in tongue pre-dorsum (Recasens 2004, 2012, 2014, 2017; Recasens and Pallarès, 2001; Ohala and Solé, 2010; Chen *et al.,* 2017). As a consequence, we expected that the linguo-palatal constriction of /s/ in the anterior tongue varied only to a limited extent as a function of articulatory environment. The EPG indices regarding constriction anteriority/posteriority confirm that /s/ is the most resistant consonant among the three: not only the tongue predorsum remains at a stable constriction location, but also the dorsal area is seldom altered by the adjacent consonant. The UTI profiles, however, suggest that the dorsal area of /s/ in the production of some speakers can accommodate to the following /k/: a relatively higher tongue dorsum is found for /s/, with respect to the case in which /s/ is followed by /t/.

The discrepency between the EPG index and the UTI profiles for /s/ in /sC/ cluster lies in the fact that the these two techniques are able to provide articulatory information from two different aspects. For the laminal consonant /s/ in /sk/ and /st/ clusters, the absence of the linguo-palatal contact in the palatal region does not necessarily imply the absence of the vertical displacement of the tongue. In the case where EPG evidences fail to tell the difference between the two /s/, the mid-sagittal lingual configurations could provide the complementary articulatroy information regarding of the vertical displacement.

The lateral consonant /l/ was supposed to have an intermediate degree of coarticulatory resistance, thus it was expected to find a more variable tongue predorsum than with /s/. The results are consistent with the expectation in two aspects: first, the anterior constriction is more anterior when followed by homorganic /t/ than by /k/; second, similar to laminal /s/, the tongue region not directly involved in the constriction is affected by dorsal /k/, as revealed by both the Qp index and the UTI configurations (see Figure 7.6). In addition, /l/ has a higher dorsum and an overall more variable lingual configuration than /s/ (in both tongue dorsum and pre-dorsum) when followed by dorsal /k/. By contrast, the UTI profiles fail to distinguish /s/ and /l/ in the tongue dorsum when they are followed by /t/. However, as far as the anterior part of the tongue is concerned, the lateral appears to have a more variable pre-dorsum than the laminal fricative.

The dorsal consonant /k/ was supposed to be the least constrained, especially in the tongue pre-dorsum area. The coarticulatory patterns as revealed by the CAa index confirm that dorsal /k/ adapts well to the articulatory configuration of the adjacent consonant with the tongue blade and tongue tip constriction (*e.g.*, /s/ and /l/) in coda consonant clusters. The activity of the tongue dorsum is not sensitive to the adjacent articulatory configuration (*e.g.*, in the clusters /lk/, /sk/ and /ks/) probably due to the fact that the articulatory gestures of /s/ and /l/ do not have a rigid requirement on the dorsal configuration. However, the UTI tongue configurations suggest that /k/ is produced with a more variable dorsal constriction when preceded by a lateral than by a laminal fricative.

Taken together, these observations lend further support to the hypothesis of a hierarchy in the degrees of articulatory constraint, that is, laminal /s/ is the most resistant since it allows less dorsal variability compared to /l/ when adjacent to /k/; dorsal /k/ is the least articulatory constrained, and lateral /l/ is in the middle.

In the following chapter, we discuss the relationship between the spatial (coarticulatory) and temporal (coordinatory) aspects of consonant-vowel and consonant-consonant interactions that have been uncovered in this and the preceding chapter.

Chapter IX General Discussion

9.1 Introduction

The current study had two fundamental purposes as discussed in Chapter 6. The first purpose was that of implementing of a methodologically challenging experimental setting, designed in particular to advance our knowledge of inter-gestural temporal coordination and coarticulation by allowing simultaneous inspection of linguo-palatal contact and tongue midsagittal contours. The second purpose was that of verifying whether the macroscopic patterns of inter-gestural behaviours could be altered by the microscopic (co-)articulatory properties of the linearly-ordered gestures in a selection of stimuli produced by Tuscan speakers of Italian, at the same time looking at how the macroscopic and the microscopic properties interact with each other in the speech system.

As far as the first purpose is concerned, the experimental recordings were successfully run within the *SynchroLing* system. Furthermore, the speech materials were easily obtained and accurately analysed in the AAA software environment, which enabled us to manage complex data sources in a unified experimental environment. Therefore, the *SynchroLing* system proved itself a fruitful instrument for the real time elaboration of multi-level synchronised speech data, including acoustic, ultrasonic tongue imaging and electropalatalgraphic data.

The speech data captured by *SynchroLing* were used to characterise the dynamics of articulatory gestures - inter-gestural temporal coordination and coarticulation in specific terms. This enriched window over speech movements was exploited in the present study in two different ways. On the one side, it was used to determine the temporal dynamics of the consecutive constriction and opening gestures of the linearly-ordered consonants and vowels. In such a way, we could identify the onset and offset of the involved C and V gestures from multiple sources of articulatory information which cover a large area of the articulatory organs, as opposed to, for instance, EMA-based point tracing of the movement of individual coil. On the other side, the multi-level analysis enabled us to collect information on the spatial adjustment of C and V lingual gestures during the production of speech according to various perspectives, *e.g.*, by showing the effects of variable constriction 'points' for those lingual regions that are not directly involved or even not involved at all in the constructions themselves.

The pilot study in the current dissertation was somewhat limited in scope and additional studies are undoubtably needed to improve the methodological aspects and to develop a more sophisticated technique of data analysis. In particular, concerning the issue of measurement, it would be important in future analyses to directly map UTI-based lingual profile information onto EPG-based information of linguo-palatal contacts on the surface of artificial palate and vice versa. These improvements will further provide advantages for both articulatory phonetics (*e.g.*, to quantify cross-instrument reliability of instrument-specific measures) and for fundamental theories of laboratory phonology (*e.g.*, to determine the accommodation patterns of tongue back as a response to the movements in the tongue front, *etc.*) (see also Recasens and Rodriguez, 2016; Spreafico *et al.*, 2016 for similar considerations).

In the following section, we will discuss the implications of the coordination and coarticulation patterns which were found in the experiment. The findings for singleton consonants will be discussed first, followed by those concerning complex codas and intervocalic consonant clusters.

9.2 Singleton consonants as onsets and codas

A production experiment was conducted to investigate the differences between syllable onset and coda consonant in their temporal coordination with the syllable nucleus. The hypothesis that we intended to verify with Italian CVCV and CVC word stimuli was based on well-known effects of syllable structures on inter-gestural coordination (see above Chapters 3 and 6). In particular, we hypothesised that onset consonant gestures are more tightly coordinated to the following nuclear vowel gestures than the same consonant gestures in coda position. Our hypothesis was confirmed in the current study by the data concerning the temporal interval between Ctarget and Crelease in CV sequences, which proved to be much smaller than the temporal interval between Vtarget and Conset in VC sequences. The difference between the two categories was statistically significant, suggesting that the procedures for the identification of articulatory landmarks in *SynchroLing* (see Chapter 7) were reliable and that it was possible to correctly identify the onset and the offset of C and V gestures in the mono- and disyllabic experimental stimuli used in the current study, by looking at the gestural stability areas defined concurrently by UTI and EPG data.

An additional hypothesis concerned the possibility that the patterns of inter-gestural coordination varied as a function of the consonant. This hypothesis was based on the finding that, in some languages (*e.g.*, Marin and Pouplier 2010 for American English, Pouplier 2012 for German, Pouplier 2015 for Polish, Marin 2009 and Marin and Pouplier 2010 for Romanian, *etc.*), different consonants have distinct articulatory behaviours within the domain of syllable and the onset-coda opposition could show up differently according to the articulatory properties of the segments/gestures involved. As reviewed in the Chapter 3, the most frequently investigated coordination pattern in this respect is the so-called C-centre effect for consonant clusters in syllable onset. In that case, the magnitude and the regularity of such "macroscopic" coordination pattern, have been shown to be influenced by the "microscopic" properties of the segments, characterised by hierarchically ordered degrees of resistance to coarticulation.

In the current study, we hypothesized that different consonants are timed to the vocalic nucleus as onsets or codas in different ways. In particular, less resistant consonants were expected to show a larger onset-coda coordination difference than more resistant consonants.

We tested three consonants, that were assumed to vary for their degree of resistance to vowel-induced coarticulatory modifications: /s/, /l/ and /k/. Their degree of coarticulatory resistance was established by measuring the articulatory modifications of each consonant adjacent to /i/ as compared to adjacent to /a/. If, as reviewed in Chapter 3, coarticulatory resistance only affects the magnitude of the c-center effect (possibly because clusters are more susceptible than singletons to the influence of varying articulatory conditions), then the three consonants should show a similar difference between tight onset-nucleus coordination (*i.e.*, a short latency between the consonantal and the vocalic gesture in CV sequences) and loose onset-nucleus coordination (*i.e.* a comparatively longer latency between the vocalic gesture and the consonantal gesture in VC sequences). If, on the contrary, consonantal properties such as coarticulatory resistance influence the onset-coda coordination pattern in singletons, the three consonants are expected to vary in the way they are timed to the vocalic nucleus as onsets or codas. In particular, spatial (coarticulatory) and temporal effects of vowel variation over adjacent consonants were expected to be positively correlated. Less resistant consonants (*i.e.*, those that vary a lot as a function of vocalic context) were expected to be influenced by the vocalic gesture both in the sense of modifying their constriction location (spatial coarticulation) and in the sense of an increased temporal overlap between the vocalic gesture and the consonantal gesture. The effect was expected to be particularly strong for consonants in onset position. Thus, less resistant consonants in onset position were expected to show not only more variability of the constriction location in the /a/ vs. /i/ contexts, but also a stronger anticipation of the vocalic gesture. Consequently, they were also expected to show larger onset-nucleus coordination differences, compared to more resistant consonants. In the case of the consonants of the present study, /s/ was expected to be the most resistant consonant and to show the smalles onset-coda coordination difference; /k/ was expected to be the least resistant consonant and to show the largest onset-coda coordination difference; /l/ was expected to occupy an intermediate position between the two, as far as resistance to coarticulation is concerned, and therefore to show an onset-coda coordination difference smaller than /k/ but larger than /s/.

We discuss the results of the coarticulation analysis first. The multi-level approach adopted in this study provided us with information on variation in both constriction location and uncontacted portions of the tongue (midsagittal profile). Constriction location was predominantly evaluated on the basis of EPG index values. In this respect, /s/ was found to maintain the same linguopalatal constriction location before /a/ and /i/, in both onset and coda positions; /l/ showed fronting before /i/ in onset position only; /k/ changed in both onset and coda positions. EPG evidence appeared therefore to support our hypothesis concerning coarticulation degrees across consonants. The target constriction for /s/ notoriously requires fine control over the articulators in order to narrow the air channel and sustain turbulence during frication, which can explain the absence of vowel-induced variation. However, the same EPG indices also revealed that /s/ and /l/ were overall more fronted in coda than in onset. This effect of syllable position on constriction location did not generalize to /k/. According to Recasens (2004: 451-452), in Catalan /VC1#C2V/ stimuli /s/ and /l/ are more fronted when they are in coda position (C1) than when they are in onset (C2), provided that the adjacent consonant is articulatorily less constrained at tongue front. By contrast, if the adjacent consonant is more constrained at tongue front (*e.g.*, /ʃ/), then the pattern is reversed (*i.e.*, /s/ and /l/ have higher fronting values in onset than in coda). Since our stimuli include intervocalic /s/ and /l/, and vowels have a less constricted lingual configuration than the two apical consonants, we can conclude that our results are consistent with what has been observed for Catalan and lend further support to the generalizations stemming from an account of coarticulation patterns in terms of segments' coarticulatory resistance.

Besides constriction location, tongue dorsum and postdorsum as revealed by UTI images were found to vary more for /l/ than for /s/. For both consonants, tongue tip position changed as a function of vowel quality only in a minority of the cases, thus indirectly confirming the results of the EPG analysis. However, there were nevertheless significant differences in the midsagittal profile as far as the dorsal and pre-dorsal areas were concerned. This clearly revealed that, even for strongly coarticulation-resistant consonants, such as /s/ (and slightly less resistant consonants, such as /l/), areas of the tongue that are not directly involved in constriction realization may accommodate to the articulatory requirements of the adjacent vowel, without substantially modifying the position and the amount of linguo-palatal contacts. In the UTI analysis, /s/ and /l/ were therefore similar with respect to coarticulation (whereas EPG data differentiated more clearly between the two). However, while /s/ showed (in one of the subjects) more vowel-induced variation in coda than in onset, /l/ showed (in two of the subjects) the opposite pattern, *i.e.*, less vowel-induced variation in coda than in onset. Whether and how this finding is important for the overall picture of /s/ and /l/ coarticulatory behavior sketched so far should be ascertained in future analyses, using quantitative measures of tongue profile variations and possibly comparing the behavior of a larger number of informants. The tip and blade of the tongue were also found to consistently and significantly vary before /a/ as compared to /i/ for the dorsal consonant /k/. This finding suggested that variation in constriction location additionally implies a change in the overall tongue configuration. Thus UTI-based analysis apparently highlights a basic distinction between dorsal /k/, whose place of articulation changes a lot as a function of a modification of the vocalic context, and the two anterior consonants, that still change, but not necessarily for constriction location, and to a smaller extent.

In sum, the multi-level approach adopted here was found to increase the amount of information on vowel-induced modifications in lingual consonant production, thus showing up as a potentially relevant way of improving our understanding of coarticulation patterns. There are major and minor differences among consonants. Major differences (*e.g.*, the difference between dorsal and apical consonants) are highlighted by both EPG and UTI. Minor differences (such as the extent to which the overall configuration of the tongue accommodates to varying constriction locations) may emerge differently, according to the specific source of articulatory data from which information is drawn. More sophisticated ways of establishing correspondences and overlapping functions between the two instrumental outputs (as well as between the articulatory output and the acoustic data) have to be developed, to fully exploit the advantages of the multi-level analysis of speech movements.

We additionally tested the hypothesis that the onset-coda coordination difference varied as a function of the specific consonants involved. We found that the three consonants did vary in the way they differentiate onset-nucleus from nucleus-coda coordination, but not always in the expected direction, which was a larger latency difference for less resistant consonants and a smaller latency difference for more resistant consonants. As a matter of fact, our findings met the expectations to the extent that the latency difference was bigger for /l/ than for /s/, but the finding that the latency difference for /k/ was even smaller than for /s/ did not fit our initial hypothesis. In a comparison between /s/ and /l/, the lateral was found to show a very short CV latency in onset position compared to /s/. This finding can be interpreted that, consistently with our initial hypothesis, not only the articulation of the following vowel influences the constriction location of the lateral more than it influences the constriction location of the sibilant, but also its temporal anticipation is stronger in the case of an onset /l/ than of an onset /s/. However, if this was the case, onset /k/ should have shown an even stronger anticipation of the vowel and a consequently shorter CV latency than /l/, which was not the case.

We think that there are at least two possible explanations for these findings. The first is that the onset-coda coordination effect is actually not affected by the coarticulatory resistance of singleton consonants. Resistance to coarticulation has been found to influence the c-center effect in clusters and maybe singletons are not as prone as clusters to suffer the influence of varying articulatory conditions. The second possibility is that the onset-coda coordination effect is affected by the coarticulatory resistance of the singletons, but only for a subset of them. In particular, one could hypothesise that it is only within the range of apical consonants that the degree of articulatory constraint determine the temporal effects of segments coordination; dorsal consonants could be hypothesized to be influenced by adjacent vowels only in terms of variable tongue configuration and changing constriction location. This difference might be due to the fact that dorsal consonants are produced by the movement of a less flexible and therefore slower part of the tongue, which might obscure the effects of tight C-V coordination in syllable onsets. The finding that /k/ showed the smallest difference between onset C-V and coda V-C coordination latencies (55 msec, as opposed to 67 msec for /s/ and 77 msec for /l/) might bring support to this view, since it shows that the vocalic gesture is not strongly anticipated during the production of onset /k/. However, to be fully ascertained this hypothesis should be tested by comparing two dorsal (or non-apical) consonants with different degrees of coarticulatory resistance, to verify if the same hierarchy found for /l/ and /s/ applies to consonants produced more in the back in the oral cavity. Additional experiments would therefore be needed.

An additional aspect which was analysed with respect to singleton consonants is the role of prosodic prominence in shaping consonant-vowel gestural coordination. The results showed that prosodic variation (that is, the presence or absence of corrective pitch accent on the relevant syllable) affected the nucleus-coda temporal relation in a significant way, but not the onset-nucleus temporal relation. The stimuli bearing corrective focus exhibited a looser VC coordination than the stimuli bearing broad focus. In fact, the stimuli bearing corrective focus had a longer vowel stability area, which corresponded in acoustic terms to vowel duration. In addition to the observation that prosodic variation had no influence on the onset-nucleus coordination pattern, we can conclude that, in our data, there is no direct evidence of prosodic prominence showing an impact on the articulatory dynamics of onset and coda consonants; the observable modications are rather related to the instrinsic duration of vowels, which are longer in prosodically prominent positions.

9.3 Post-vocalic consonant clusters: word-final and intervocalic clusters

9.3.1 Articulatory organisation for /ks/ coda cluster

For the postvocalic consonant clusters, both V-to-C1 relations and the internal organisation within the clusters were investigated in order to verify whether the coordination patterns vary as a function of the articulatory properties of the consonants. Browman and Goldstein (1988) claimed that the timing between the vocalic nucleus and the vowel-adjacent coda consonant remains unchanged irrespective of coda complexity and coda compositionality. Following their hypothesis on the nucleus-coda inter-gestural timing, the leftmost coda consonant always achieves its articulatory target at the same time with respect to the preceding vowel and there is no automatic increase in the amount of consonant-vowel inter-gestural overlap as the number of the consonants in the coda cluster increase. They then argue that, in coda clusters, the first consonant in the sequence will be linked to, and partially overlapping with, the vowel, but the consonant(s) following the leftmost postvocalic one will not be linked to or overlap with the vowel (which motivates the absence of the C-centre in coda clusters). Our data only partially confirmed the prediction put forward by Browman and Goldstein. For laminal /s/ and lateral /l/, the V-to-C1 timing was fairly stable across the two conditions, whereas for /k/ we found a significantly tighter coordination (with a difference of almost 20 ms) in the complex coda cluster /ks/ with respect to the singleton coda /k/. This is on pair with the reduced length of /a/ in /aks/ as compared to /ak/: 145,55 ms *vs*. 154,1 ms, respectively. However, such durational difference at the acoustic level across two different cases of segmental complexity was not statistically significant ($p > .05$).

At first sight, the shortening of /a/ seems to be consistent with the observation that the vowel shortens as the number or length of subsequent consonants increases (Munhall, Fowler, Hawkins and Saltzman, 1992; Byrd, 1994, 1995). In particular, Munhall *et al.* (1992) claim that in English monosyllabic words, vowel duration is shortened as the duration of the following coda cluster increases, and this is due to an earlier jaw raising gesture for the following consonant gesture in a cluster than in a singleton, thus creating a shorter interv al of the jaw lowering movement for the vocalic gesture which precedes the consonant cluster. Nevertheless, such compensatory shortening of the syllabic nucleus is only found for a subset of coda consonants in subsequent studies (Marin and Pouplier, 2010; Katz, 2012). In a recent study, Katz (2012) reported that in English monosyllabic word, the shortening of the nuclear vowel is only found when the syllable in question changes from open to closed with one coda consonant. He failed to find (incremental) shortening when a singleton coda changes to a complex coda, except for obstruent-liquid sequences. In addition, Byrd (1995) finds a significant lengthening of the preceding nuclear vowel when the coda clusters are three-obstruent sequences (/sks/ and /skt/) as opposed to two-obstruent coda clusters. Sugahara and Turk (2009) argue that this is probably due to the fact that these codas are unambiguously polymorphemic in English and the English polymorphemic coda sequences are usually elongated with respect to the monomorphemic ones.

This evidence seems to suggest again that the segmental properties of the postvocalic consonants in the microscopic dimension play a role in regulating the nucleus-coda temporal relation, similarly to what happens in onset clusters reviewed above (*e.g.,* Marin 2011, 2013 for Romanian complex onsets). Our data, as already mentioned, confirm this view inasmuch as the three consonants provide different results, with /k/ showing a significantly tighter articulatory coordination when the coda is complex than when it is simplex, and /s/ and /l/ showing no significant differences. it should also be noticed that, although vowel shortening as a function of syllable structure and coda complexity is traditionally and empirically investigated at the acoustic level (acoustic vowel duration), in our study we found evidence for shorter articulatory stability area. The articulatory perspective can thus provide a fresh look at vowel shortening phenomena.

For the singleton coda /k/, the articulatory movement requires 226,8 ms to travel from the Vtarget to Ctarget, while in case of the complex coda /ks/, such temporal latency is significantly reduced to 187,8 ms ($p < .001$), which could confirm that there is an increased gestural overlap between /a/ and /k/ in case of a complex coda relative to a singleton coda.

We hypothesise that the effect found for /k/ might be interpreted from the following two perspectives. First, the tongue region involved in the formation of the dorsal constriction is relatively less flexible than the apical constriction needed to produce /s/ or /l/, and therefore the tongue dorsum raising gesture for the production of /k/ is activated earlier, with a larger amount of overlap with the preceding dorsum flattening gesture needed to produce /a/, compared to what happens in the cases of the coronal consonants. Second, the tongue dorsal constriction is independent of the tongue tip constriction. In a coda dorso-apical sequence like /ks/, the tongue tip gesture is more constrained and requires more articulatory precision. Thus, the dorsal constriction gesture for /k/ is anticipated in order to guarantee the enough 'articulatory space' for the movement of the tongue tip constriction. This in turn explains why such temporalancticipatory effect is absent in the V-to-/s/ timing for the word-final cluster /sk/ as compared to the singleton coda /s/. The laminal-dorsal sequence /sk/ is composed of both tongue tip and dorsal gestures. However, dorsal /k/ is preceded by the articulatorily highly demanding gesture /s/ in this case. Consequently, the realisation of the laminal constriction of /s/ does not impact the dorsal constriction and no anticipatory effect takes place.

In addition, we calculated the time for /k/ to reach its articulatory target from the gestural onset (*i.e.,* the temporal latency from Conset to Ctarget) in the two cases. This value could further suggest whether there is an acceleration for the consonant gesture to reach its articulatory target due to the presence of a /s/ to its right. A nonparametric test on the median latency confirms that there is an acceleration of the dorsal raising gesture when it is followed by a laminal /s/ (98,8 ms for /aks/ as opposed to 118 ms for /ak/, $p < .05$). Laminal /s/ is considered to be more coarticulatory resistant (and articulatorily more complex) than /k/. Since the dorsal gesture for /k/ easily undergoes the coarticulatory effects induced by the adjacent laminal consonant /s/, there is less room left for /k/ to reach its target, and therefore the dorsal raising gesture has to accelerate.

All this evidence points out that the highly constrained pre-dorsal narrowing and fronting gesture required for /s/ might be said to favour the temporal anticipation of the tongue dorsum raising gesture during the production of the vowel in /aks/ sequences, thus creating an increased inter-gestural overlap between vowel /a/ and the leftmost consonant /k/. By contrast, in the cases of word-final clusters /sk/, /st/, /lk/ or /lt/, where no shortening of the stability area interval was found compared to items with simplex codas /l/ and /s/, the highly constrained gestures - /s/ and /l/ - are followed or by the dorsal consonant /k/ (/sk/ and /lk/) or by the consonant with the homorganic articulator (/st/ and /lt/). In the former case, since the tongue tip gesture and the dorsal gesture can be realised independently, there is no need for the the apical gesture to be anticipated in the preceding articulatory domain of vowel simply because the following /k/ is less articultorily constrained; in the latter case, the sequence made up of consonants with the homogeneous articulator could[44] increase the within-cluster inter-consonantal gestural overlap instead of being anticipated in the vowel articulatory domain. That is why a tighter V-to-C1 timing is only attested for /ks/ and not for other clusters in our study.

What we have found for the V-to-/k/ timing for complex coda /ks/ might lend further support to the hypothsesis proposed before for the singleton onset vs. coda consonant-vowel inter-gestural timing, namely, that only a subset of the consonants (*e.g.*, whose constriction requires control of the tongue pre-dorsum area, like /s/ and /l/) in the current hierarchy of the degrees of coarticulatory resistance are able to condition the consonant-vowel temporal relationship. This is probably also linked to intrinsic motor limitations of dorsal consonants (*e.g.*, /k/), in which a less flexible tongue dorsum area travels more slowly and somewhat independently from the tongue tip or blade. Consequently, the dorsal consonants are temporally more impacted than apical consonants by the presence of a following apical gesture, especially when this following apical gesture is articulatorily highly demanding, as in the case of coda /ks/.

9.3.2 Articulatory organisation for /sC/ clusters

The postvocalic word-final /sC/ clusters (/sk/ and /st/) exhibited, as already mentioned, left-edge stability, by showing the same timing relation with the preceding nucleus compared to the singleton coda /s/. This is consistent with cross-linguistic evidence found for American English (Marin, 2009; Marin and Pouplier, 2010) and German (Pouplier, 2014) /s/-series coda clusters. Additionally, we found that the V-to-/s/ temporal coordination was conditioned by the following consonant in the cluster: the /s/ in the /st/ sequence showed a significantly tighter coupling relation with the

---

[44] But not necessarily so, as shown in the following sections, the degree of the inter-consonantal gestural overlap may also depend on the DAC value of the individual consoanants.

preceding vowel than the /s/ in the /sk/ sequence. Tighter coupling relations with the preceding nucleus when C2 is /t/ (compared to when it is /k/) were also reported for the intervocalic /sC/ clusters (/VstV/ vs. /VskV/) and the word-final /lC/ clusters (/Vlt/ *vs*. /Vlk/, not statistically significant).

This is consistent with the findings for the /ks/ complex coda reported above. When the laminal consonant /s/ is followed by a consonant whose articulation involves an independent region of the tongue, the dorsum constriction of /k/, for example, the two gestures requiring different tongue regions could independently realise their articulatory targets respectively without necessarily overlapping with each other. Instead, if the consonant cluster (/st/ or /lt/) consists of two gestures with the same articulator, *e.g.,* the tongue tip, a competition betweent the two gestures might take place and as a conseguence the first apical might be anticipated in the articulatoty domain of the preceding vowel in order to leave more 'room' for the rightward apical gesture.

The main result, *i.e.,* that coda clusters show the same timing relation as simplex codas, are not at odds with the gestural model of syllable timing (Browman and Goldstein, 1988, 1992, 1993; Goldstein and Pouplier, 2014). On the contrary, they further confirm the lack of the C-centre effect in coda consonant clusters, which produces the same V-to-C1 timing in coda clusters irrespective of the number and type of consonants involved. However, our data also support a partial revision of the traditional approach to gestural timing, to the extent that they strongly suggest that a major role in determining the detail of the coupling relations is played by the articulatory and coarticulatory properties of the individual coda consonants, as discussed above for the difference between homorganic and non-homorganic clusters and for the difference between /sk/ and /ks/ coda clusters.

As to within-cluster gestural organisation, /s/ was found to have a longer gestural stability interval than /l/, no matter if the following consonant is a dorsal /k/ or a coronal /t/. This lends further support to our original hypothesis that /s/ is more coarticulatory resistant than /l/. /s/ and /l/ are both highly resistant to the coarticulatory effect from the adjacent gestures at the tongue predorsum zone. However, the laminal constriction needed for production of frication requires more articulatory control in the predorsal area than the central closure needed for an /l/ and is therefore less prone to a compromise with adjacent gesture(s). This also corresponds to the findings of the CAa index of /s/ and /l/ when they precede the dorsal /k/ and the coronal /t/, respectively. For /s/, the predorso-palatal contact index remains stable across contexts, whereas for the lateral there is less fronting when it is followed by /k/ and more fronting when it is followed by /t/.

Again, with respect to the issue of within-cluster gestural organisation, /s/ was found to coordinate less tightly with an upcoming coronal /t/ than with an upcoming dorsal /k/. This difference is statistically significant for intervocalic clusters /asCa/ (175,8 ms for /asta/ *vs*. 151,8 ms for /aska/) but is also present in word-final clusters /asC/. This difference is probably due to the already mentioned fact that the tongue dorsum raising gesture for /k/ requires the movement of a tongue region that is at the same time less flexible and independent from the tongue tip and blade, that are involved in the production of the other consonants; no articulatory conflict with the preceding /s/ is thus created.

The different coarticulatory patterns of /sk/ and /st/ sequences revealed by EPG correspond to the UTI evidence regarding on the midsagittal tongue configurations of /s/ when followed by /k/ and /t/ in the two different segmental conditions /VsC/ *vs.* /VsCV/. The tongue dorsum of /s/ is higher when followed by a dorsal consonant /k/ than when it is followed by a homorganic apical consonant /t/ (see Figure 8.4 and 8.5). The vertical displacement of the tongue dorsum in the spatial dimension captured by the UTI in the articulatory domain of /s/ further supports the claim that the dorsal movement is independent of the tongue tip constriction and thus results in a tighter inter-gestural coordination. For /st/ sequence, however, the more articulatorily constrained apical gesture of /s/ prevents the following homorganic constriction from being anticipated. Consequently, the tongue tip closure gesture of /t/ has to be activated afterwards, and this in turn guarantees a relatively longer gestural stability for the preceding laminal constriction (as compared to the /s/ in /sk/ sequence). The spatial coordination therefore confirms once again that the articulatory properties of the individual consonants within the cluster play an important role in conditioning the microscopic shaping of inter-gestural overlap, regardless of syllabic constituency.

When considering the intervocalic /sC/ clusters /aska/ and /asta/, we found that, as far as their within-cluster temporal organisation is concerned, they are not different from the word-final ones (/ask/ and /ast/). The inter-consonantal cohesion within the intervocalic clusters showed almost the same temporal latency between the relative articulatory anchors as compared with the word-final /sC/ clusters. The results confirmed the expectation that laminal /s/ has a loose temporal relation with the consonant to its right. Such loose temporal coordination within the intervocalic clusters successfully prevents the leftmost consonant from being analysed as part of the onset of the following syllable.

Although the intervocalic clusters failed to differentiate themselves from the word-final clusters in terms of temporal organisation, the spatial coordination patterns reflected by the EPG results provide a partly different scenario. For the sequence /sk/, mean Qp for /k/ remains stable regardless of the syllabic constituency of the rightmost /k/ (*i.e.*, in /ask/ and /aska/). However, mean CAa for /k/ in word-final /ask/ sequence is higher than for /k/ in intervocalic /aska/. Since CAa is an index of anteriority and /k/ is a consonant primarily articulted in the back region of the palate, this difference suggests that there is more contextual anteriority in the production of a /k/ when /sk/ is found in word-final position (as in /bask/) than when an intervocalic /sk/ sequence is concerned (as in /aska/). Similarly, for /st/, mean CCa of /t/ is higher in intervocalic /asta/ than in word-final /ast/. Since CCa is an index of centrality and /t/ is a fully contacted consonant midsagittally, this difference suggests that there is more contextual decentralization in the production of a /t/ when /st/ sequence is a word-final cluster (as in /ast/) than when it is an intervocalic cluster (as in /asta/)[45]. We believe that the articulatory behaviours of /k/ and /t/ in the spatial dimension as highlighted by the EPG indices should further be interpreted in terms of inter-gestural cohesion, that is,

---

[45] The fact that the constriction location of /t/ is less central when adjacent to /s/ is also reported by Bladon and Nolan (1977) for English. Such behaviour of /t/ lends further support to the claim that the laminal fricative /s/ is more (co-)articulatorily constrained than the coronal /t/ at the level of tongue pre-dorsum due to strict requirements of its manner of articulation.

the coarticulatory effects induced by the laminal /s/ are reduced by the presence of another vowel to the right of the consonant cluster. Less (co-)articulatorily constrained gestures (such as those for /k/ and /t/) better preserve their own articulatory properties in the spatial dimension when the syllabication of intervocalic /sC/ clusters is undecidable (Bertinetto, 1999a, 1999b, 2004), much like the temporal dimension of inter-gestural overlap also reveals. Therefore, these results from the perspective of spatial coordination are in favour of the claim that inter-gestural overlap is sensitive to language-specific micro-phonotatic variations.

9.3.3 Articulatory organisation for /lC/ clusters

As already mentioned, our data revealed that the nucleus-coda temporal relation did not change with the increase of coda complexity in the case of /l/, that is, the raising of the tongue tip for the lateral /l/ occurs at almost the same time with respect to the preceding vowel regardless of whether it is word-final or it is followed by another consonant (with either homogeneous or heterogeneous articulators, /t/ and /k/). This is consistent with the hypothesis that, at the macroscopic level, the vowel-adjacent coda consonant always entertains a loose temporal relation with the preceding nuclear vowel, regardless of coda complexity.

Our results for /l/ however conflict with the findings in Byrd (1994, 1995) and Marin and Pouplier (2010) for American English complex codas with a leftmost /l/. In fact, Marin and Pouplier (2010) show that for two coda clusters - /lp/ and /lk/ - the tongue tip gesture occurs temporally closer to the preceding vowel anchor point in gulp and bulk compared to gull and ball respectively, that is, no stable left-edge temporal coordination is found for the liquid. The authors argue that such anticipation of the tongue tip gesture could be attributed to the fact that, in English, a coda /l/ is known to be composed of two gestures - a tongue tip raising gesture and a dorsal retraction gesture (Giles and Moll, 1975). It is the overlap between the dorsal retraction gesture and the preceding vowel gesture that results in a tighter coordination of complex vis-à-vis simplex coda liquids, which does not however limit the liquid's perceptibility.

By contrast, in a study on German complex codas by the same author (Pouplier, 2014), it is shown that /l/-initial tautosyllabic coda clusters (/lm/ and /lp/) have the same V-to-C1 inter-gestural temporal coordination as compared to the heterosyllabic intervocalic /lm/ and /lp/ clusters, which suggested a homogeneous nucleus-coda coordination pattern regardless of coda complexity (*i.e.*, singleton /l/ vs. /lm/ and /lp/ coda clusters). Additionally, other complex codas like /ks/, /kt/ and /bt/ showed an identical pattern, thus suggesting that such left-edge stability holds true irrespective of the coda compositionality. The author attributed the difference with the findings in Marin and Pouplier (2010) to the fact that German /l/ is not dark and thus it is not supposed to have any retraction gesture. Supposedly, the retraction gesture of coda /l/ in English conditions the result that, in this language, the V-to-/l/ timing is sensitive to coda complexity. The same interpretation could extend to our Italian data. Since Italian /l/ is clear, the absence of the dorsal retraction gesture would guarantee that the pre-dorsum fronting gesture reaches its target within a relatively stable temporal domain.

As for the internal organisation within the /lC/ clusters, the lateral-dorsal sequence /lk/ showed a significantly tighter inter-consonantal cohesion than the lateral-coronal sequence /lt/. This is on a par with the tighter coordination between /l/ and the preceding nuclear vowel in /alk/ as opposed to /alt/, mentioned above (the section 9.3.2). Recall that the same two effects were also found for /s/. As above (the section 9.3.2), we can interpret these findings as probably due to the fact that the tongue dorsum raising gesture for /k/ requires the movement of a tongue region that is at the same time less flexible and independent from the tongue tip and blade and therefore no articulatory conflict is created with the preceding /l/, if the dorsal gesture is anticipated during the production of the lateral. This confirms once again that the articulatory properties of the individual consonants within the cluster play an important role in conditioning the microscopic shaping of the inter-gestural overlap.

The spatial coordination patterns revealed by the EPG data show that the CAa value for /l/ is higher when followed by a coronal /t/ than when followed by a dorsal

/k/; the Qp value was also significantly higher when a dorsal raising gesture occurred afterwards, suggesting that the tongue region not directly involved in the formation of the constriction for /l/ could undergo the (co-)articulatory effects from the adjacent dorsal gesture. Accordingly, at the level of UTI tongue profiles, a higher tongue dorsum position for the lateral in the context of a lateral-dorsal sequence was also found (see Figure 8.6). Such spatial coordination differences are consistent with the inter-gestural temporal coordination patterns differences between /lk/ and /lt/, resulting in an overall tighter gestural cohesion in the former cluster compared to the latter.

We can therefore conclude that the inter-gestural coordination between vocalic nucleus and coda /l/ is insensitive to coda complexity, but undergoes anticipatory spatial and temporal coordination effects when a dorsal consonant follows, that is, when the subsequent consonant is articulated in a tongue region not directly involved in the articulation of the /l/. Though such anticipatory spatial effect does not contribute in a statistically significant way to the vowel-liquid temporal relation, it does regulate the internal temporal organisation within the coda cluster. This temporal-spatial interaction at the articulatory level thus lends support to the claim that inter-gestural temporal coordination and coarticulation (*i.e.,* spatial coordination) are two aspects of a same phenomenon, *i.e.,* gestural overlap.

9.4 Conclusion and outlook

This study has examined the inter-gestural temporal coordination and coarticulation of singleton consonants and clusters with various segmental conditions in Italian. In spite of the abundant literature on coordination patterns in onset clusters and its relation with the articulatory properties of the consonant(s) that participate(s) in the coupling relation (*e.g.,* Marin, 2011 for Romanian; Hermes, 2012 for Italian /sC/ clusters; Pouplier, 2015 for Polish), very limited evidence exists on postvocalic consonant clusters (*e.g.*, Marin and Pouplier, 2010 for American English coda clusters; Pouplier, 2014 for German complex coda). This dearth of data was one of the motivations of this dissertation, which intendes to answer the following questions: a) how do Italian speakers organise the postvocalic consonant gesture(s) and their relation to the preceding vocalic nucleus; b) whether Italian speakers adopt different articulatory strategies for complex codas with respect to singleton codas; c) whether the differences in phonotatics (/VsC/ *vs*. /VsCV/) reflected different degrees of within-cluster inter-gestural overlap. By looking at three consonants varying for both place and manner of articulation, we could provide a relatively exhaustive set of answers to these questions.

In order to answer these questions, the movements of the overall tongue configurations and the linguo-palatal contacts during the production of the stimuli were recorded via the *SynchroLing* system composed of an ultrasound tongue imaging machine (UTI) and an electropalatography (EPG). The raw data captured by the *SynchroLing* system were further elaborated in the environment of the AAA software and interpreted within the framework of Articulatory Phonology in terms of inter-gestural coordination.

It has been hypothesised in Chapter 4 that temporal coordination among articulatory gestures and their spatial accommodation (or coarticulation) are two aspects of one and the same phenomenon — gestural overlap. In this regard, this dissertation constitutes an attempt at showing in concrete terms, by means of synchronous multi-level phonetic data, that the details of the spatial accommodation occurring among consecutive articulatory gestures are intrinsically coupled with their temporal phasing in the speech stream.

We have already anticipated in Chapter 5 that articulatory movements can be characterised as other physiological movements by the presence of two basic coordination patterns, in-phase and anti-phase, corresponding in the speech domain to complete and minimal gestural overlap, respectively. As shown in the previous chapters, syllable structures, systematically conditioned by the phonotactics of languages, are the primary results of the physiological constraints which operates in the individual speech systems: the onset-nucleus timing is shaped by the simultaneous inter-gestural relation (in-phase), whereas the sequential relation (anti-phase) could determine the nucleus-coda inter-gestural timing; and in some languages onset consonants show a complex coordination pattern governed by the C-centre effect, while in others such C-centre effect is absent. The in-phase relation is demonstrably more stable than the anti-phase one (*e.g.*, rate-induced resyllabification of VC to CV reported first by Stetson 1951); it is assumed to be biologically preferred because of its correspondence to the natural jaw aperture-closure cycle; fundamental works on speech errors also favour the idea that there is a link between in-phase coordination and the jaw cycle (Goldstein *et al.*, 2007)[46]. However, establishing an efficient relationship between physiological constraints and linguistic diversity still needs much phonological work, especially considering that many studies on Slavic languages[47] continuously provide examples of much more pervasive variation than previously assumed on the basis of physiological preferences (Pouplier and Beňuš, 2011 on Slovak; Pouplier, 2015 on Polish). This kind of evidence suggests the important role of the articulatory properties of segments in predicting the temporal phasing among speech gestures. Furthermore, traditional as well as more recent studies on coarticulation (Bladon and Al-Bamerni, 1976; Recasens, 1999, 2006, 2014; Farnetani, 1989; Farnetani and Recasens, 1999, 2010; Recasens *et al.*, 1995, 1997) reveal that the degree of articulatory constraints — or coarticulatory resistance — plays an equally important role in shaping inter-gestural overlap in spatial terms, what has been traditionally understood as coarticulation.

Taking into account the complexity of these relationships among multiple physical and categorial constraints in order to explain cross-language diversity as well as within-language variation is among the most urgent challenges that current, substance-based phonological theorising must face. Future research inspired by the same gestural hypothesis and based on the interaction between the "macroscopic" and the "microscopic" levels of articulation (see above, chapter 2, for a definition of the two levels), with the fundamental contribution of the DAC model, will have to be directed

---

[46] The speech errors are, in fact, controlled by the cognitive activity at a higher level. Not all the lapsus could be simply attributed to the accommodation to a gestural-based syllable structure.

[47] Another classic counter example comes from Tashlhity Berber in which onsets are restricted to singleton consonants, and all consonants can function as syllable nucleus (Goldstein *et al.*, 2007).

at least towards the following three aspects: first, to further verify inter-gestural coordiantion in both temporal and spatial dimensions for more consonant clusters in Italian, especially in intervocalic position; second, to explore whether the prosodic domains other than the syllable (*e.g.,* foot, phonological word, *etc.*) correspond to different gestural realisations at the articulatory level (*e.g.,* enlarged or shrunk gestural magnitude; increased or decreased inter-gestural overlap, *etc.*), since the boundary-lengthening effects (at both boundary-inital and boundary-final) have been recognised for larger prosodic domains – intermediate phrase and intonational phrase – in the articulatory dimension; third, to explore whether the rhythmic structure (syllable-timing *vs*. stress-timing) of the languages (as traditionally measured in terms of segments' acoustic duration) has an articulatory basis or at least an impact on the phenomenology of language-specific coarticulatory and coordinatory patterns at the gestural level.

Bibliography

Arvaniti, C., Ladd, D. R., and Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26:3-25.

Atterer, M. and Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics,* 32(2):177-97.

Avesani, C., Vayra, M., and Zmarich, C. (2009). Coordinazione vocale-consonante e prominenza accentuale in italiano. La sfida della Articulatory Phonology. In Ferrari, G., Benatti, R., and Mosca, M., (Eds.), *Linguistica e modelli tecnologici di ricerca, Atti del XL Congresso Internazionale della Società di Linguistica Italiana, Vercelli, 21-23 settembre 2006.* Roma: Bulzoni.

Beckman, M. and Pierrehumbert, J. B. (1989). Intonational structure in English and Japanese. *Phonology Yearbook,* 3:255–309.

Bender, B. (1999). *Marshallese Grammar*, University of Hawaii.

Bertinetto, P. M. (1989). Reflections on the dichotomy stress- vs. syllable-timing. *Revue de phonétique appliquée,* 91(3):99-130.

Bertinetto, P. M. (1999a). La sillabificazione dei nessi /sc/ in italiano: un'eccezione alla tendenza "universale"? In Beninca, P., Mioni, A., and Vanelli, L., (Eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia, Atti del XXXI congresso della SLI*. Roma: Bulzoni. 71-96.

Bertinetto, P. M. (1999b). Psycholinguistic evidence for syllable geometry: Italian and beyond. In Rennison, J. and Kühmhammer, K., (Eds.), *Phonologica 1996. Syllables!?*, The Hague: Holland Academic Graphics. 1-28.

Bertinetto, P. M. (2004). On the undecidable syllabification of /sC/ clusters in Italian: Converging experimental evidence. *Rivista di Linguistica,* 16(2):349-72.

Bertinetto, P. M. (2010). The Syllable. Fragments of a Puzzle.

Bladon, R.A.W. and Al-Bamerni, A. (1976) Coarticulation resistance in English /t/. *Journal of Phonetics,* 4:137-50.

Bladon, R.A.W. and Nolan, F. (1977). A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics,* 5:185-93.

Boersma, P. and Weenink, D. (2012). Praat: doing phonetic by computer, http://www.praat.org/.

Bombien, L., Mooshammer, C., Hoole, P., and Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics,* 38(3):388-403.

Browman, C. P. and Goldstein, L. (1985). Dynamic modelling of phonetic structure. In Fromkin, V. (Ed.), *Phonetic linguistics.* New York: Academic Press.

Browman, C. P. and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook,* 3:219-52.

Browman, C. P. and Goldstein, L. (1988). Some notes on syllable in Articulatory Phonology. *Phonetica,* 45 (2-4):140-50.

Browman, C. P. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology,* 6:201-51.

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

Browman, C. P. and Goldstein, L. (1991a). Gestural structures: distinctiveness, phonological processes, and historical change. In Mattingly, I. and Studdert-

Kennedy, M., (Eds.), *Modularity and the Motor Theory of Speech Perception*. Erlbaum, New Jersey. 313-38.

Browman, C. P. and Goldstein, L. (1991b). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics,* 18:411-24.

Browman, C. P. and Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica,* 49:155-80.

Browman, C. P. and Goldstein, L. (1993). Dynamics and articulatory phonology. *Haskins Laboratories Status Reports on Speech Research,* 113:51-62.

Browman, C. P. and Goldstein, L. (1995). Dynamics and Articulatory Phonology. In Port, R. F. and Van Gelder, T., (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge, MA: The M.I.T. Press. 175-93.

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

Bruner, J., Geng, C., Sotiropoulou, S., and Gafos, A. (2014). Timing of German onset and word boundary clusters. *Journal of Laboratory Phonology,* 5 (4):403-54.

Byrd, D. (1994). *Articulatory timing in English consonant sequences.* PhD dissertation, UCLA.

Byrd, D. (1995). C-Centers revisited. *Phonetica,* 52 (4):285-306.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics,* 24 (2):209-44.

Byrd, D. (2006). Relating prosody and dynamic events: Commentary in the papers by Cho, Navas, and Smiljanić. In *Laboratory Phonology,* 8: 549-61. Berlin, New York: Walter de Gruyter.

Byrd, D. and Saltzman, E. (1998). Intergestural dynamics of multiple phrasal boundaries. *Journal of Phonetics,* 26:173-99.

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modelling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics,* 31:149-80.

Celata, C., Vietti, A. and Spreafico, L. (2017). An articulatory account of rhotic variation in Tuscan Italian: synchronized UTI and EPG data. In Gibson, M. and Gil, J. (Eds.), *Romance Phonetics and Phonology*. Oxford: Oxford University Press.

Chen, C., Ricci, I., Bertini, C., Falahati, R., and Celata, C. (2015). An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian: A preliminary study. In *Abstract Book of Ultrafest VII, Hong Kong, University of Hong Kong, 8-10 December,* pp. 49-51.

Chen, C., Celata, C., and Ricci, I. (2017). An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian. In Bertini, C., Celata, C., Lenoci, G., Merluzzi, C., and Ricci, I., (Eds.), *Fattori sociali e biologici nella variazione fonetica / Social and biological factors in speech variations*, pp. 145-65.

Cho, T. (2001). *Effects of Prosody on Articulation in English.* PhD dissertation, UCLA.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics,* 32:141-76.

De Jong, K. (1992). Acoustic and articulatory correlates of p-center perception. *UCLA Working Papers in Phonetics,* 81:66-75.

De Jong, K. (2001). Effects of Syllable Affiliation and Consonant Voicing on Temporal Adjustment in a Repetitive Speech Production Task. *Journal of Speech, Language, and Hearing Research,* 44:826-40.

Dressler, W. G. and Dziubalska-Kolaczyk, K. (1994). Evidence of syllabic writing against the phonological syllable. In Dressler, W. G., Prinzhorn, M. and Rennison, J. R., (Eds.), *Phonologica* 1992. Torino: Rosenberg and Sellier. 65-76.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton.

Farnetani, E. (1986). Lingual and velar coarticulatory movements in the production of /n/ in Italian: some preliminary data. *Quaderni del Centro di Studio per le Ricerche di Fonetica, Padova, Consiglio nazionale delle Ricerche,* 5:285-307.

Farnetani, E. (1989). An articulatory study of voicing in Italian by means of dynamic palatography. *Speech Research 89, Budapest,* pp. 395-98.

Farnetani, E. and Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 6:17-34.

Farnetani, E. and Recasens, D. (1993). Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech* 36:279-302.

Farnetani, E. and Recasens, D. (2006). Coarticulation models in recent speech production theories, In Hardcastle, W. J. and Hewlett, N., (Eds.), *Coarticualtion. Theory, Data and Techniques*. Cambridge: Cambridge University Press. 31-65.

Farnetani, E. and Recasens, D. (2010). Coarticulation and connected speech processes. In Hardcastle, W. J., Laver, J. and Gibbon, F. E., (Eds.), *The Handbook of Phonetic Sciences*, 2nd edition. Oxford: Wiley-Blackwell. 316-52.

Fontdevila, J., Pallarès, M. D., and Recasens, D. (1994). The contact index method of electropalatographic data reduction. *Journal of Phonetics*, 22:141-54.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics,* 29:109-35.

Fougeron, C. and Jun, S. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics,* 26:45-69.

Fowler, C. (1980). "Perceptual centers" in speech production and perception. *Perception and Psycholinguistics,* 25:375-88.

Fowler, C. and Saltzman, E. (1993). Coordination and coarticulation in speech production, *Language and Speech,* 36:171-95.

Fowler, C. and Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, 43(1):1-41

Fowler, C., Rubin, P., Remez, R., and Turvey, M. (1980). Implications for speech production of a general theory of action. In Butterworth, B., (Ed.), *Language Production*. New York: Academic Press. 372-420.

Gafo, A., Hoole, P., Roon, K., and Zeroual, C. (2010). Variation in overlap and phonological grammar in Moroccan Arabics clusters. *Laboratory Phonology,* X. pp. 657-98.

Gao, M. (2008). *Mandarin Tones: An Articulatory Phonology Account.* PhD dissertation, Yale University.

Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics* / 中国语音学报, 2:43-50.

Gick, B., Wilson, I., and Derrick, D. (2013). *Articulatory Phonetics.* Maldon/Oxford: Wiley-Blackwell.

Giles, S. and Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica,* 31:206-27.

Goldstein, L. (1983). Vowel shifts and articulatory-acoustic relations. In Cohen, A. and van den Broeke, M. P. R., (Eds.), *Abstracts of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications. *267-73.*

Goldstein, L. (2011). Back to the past tense in English. In Gutiérrez-Bravo, R., Mikkelsen, L., and Potsdam, E., (Eds.), *Representing language: Essays in honour of Judith Aissen*. Santa Cruz, CA: Linguistic Research Center. 69-88.

Goldstein, L. and Fowler, C. (2003). Articulatory phonology: a phonology for public language use. In Meyer, A. and Schiller, N., (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* New York: Mouton. 159-207.

Goldstein, L., Byrd, D., and Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Aarbib, M., (Ed.), *From action to language: The mirror neuron system*. Cambridge, U.K.: Cambridge University Press. 215-49.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (2006). Dynamic action units slip in speech production errors. *Cognition*, 103(3):386-412.

Goldstein, L., Chitoran, I., and Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidences from Georgian and Tashlhihyt Berber. In Trouvain, W. and Barrry, W., (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences,* pp. 241-44.

Goldstein, L., Nam, H., Saltzman, E., and Chitoran, I. (2008). Coupled oscillator model of speech timing and syllable structure. In *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers, Beijing.*

Goldstein, L. and Pouplier, M. (2014). The temporal organisation of speech. In Goldbrick, M., Ferreira, V., and Miozzo, M., (Eds.), *The Oxford Handbook of Language Production.* Oxford: Oxford University Press. 210-27.

Gussenhoven, C. (2004). *The phonology of tones and intonation.* Cambridge: Cambridge University Press.

Gussenhoven, C. and Jacobs, H. (2005). *Understanding Phonology,* 2nd Edition. London and New York: Routledge.

Haken, H., Kelso, J., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51:347-56.

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry,* 13:227-76.

Hayes, B. (1985). *A Metrical Theory of Stress Rules.* New York: Garland.

Hermes, A. (2012). *Articulatory Coordination and Syllable Structure in Italian.* Peter Lang Edition.

Hermes, A., Grice, M., Mücke, D., and Niemann, H. (2008). Articulation indicators of syllable affiliation in word initial consonant clusters in Italian. In *Proceedings of the 8th International Seminar of Speech Production, Strasbourg,* pp. 433-36.

Hermes, A., Grice, M., Mücke, D., and Niemann, H. (2012). Articulatory coordination and the syllabification of word initial consonant clusters in Italian. In Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C., and Kühnert, B., (Eds.), *Consonant Clusters and Structural Complexity*, pp. 157-76. Berlin: Mouton de Gruyter.

Hermes, A., Ridouane, R., Mücke, D., and Grice, M. (2011). Kinmatics of syllable structure in Tashlhiyt Berber: The case of vocalic and consonantal nuclei. In Laurie, Y. and Steiner, I., (Eds.), *Proceedings of the 9th International Seminar on Speech Production*, pp. 401-08.

Hermes, A., Mücke, D., and Grice, M. (2013). Gestural coordination of Italian word-initial clusters: The case of "impure s". *Phonology,* 30(01):1-25.

Hermes, A., Mücke, D., and Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics,* 64:127-44.

Hsieh, F. Y. (2011). A gestural account of Mandarin tone 3 variation. In *Proceedings of the 17th International Congress of Phonetic Sciences,* pp. 890-93.

Hoequist, C. (1983a). Durational correlates of linguistic rhythm categories. *Phonetica,* 40:19-31.

Hoequist, C. (1983b). Syllable duration in stress-, syllable- and more-timed languages. *Phonetica,* 40:203-37.

Hogg, R. and McCully, C. B. (1987). *Metrical Phonology: A Coursebook.* Cambridge: Cambridge University Press.

Hoole, P. and Pouplier, M. (2015). Interarticulatory coordination: Speech sounds. In Redford, M. A., (Ed.), *The Handbook of Speech Production.* London/New York: Wiley-Blackwell. 133-157.

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Science,* 31:213-38.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., and Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics,* 44:62-82.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics,* 40(3):390-402.

Keating, P. A. (1985). CV phonology, experimental phonetics, and coarticulation, *UCLA Working Papers in Phonetics*, 62:1-13.

Keating, P. A. (1988). The window model of coarticulation: articulatory evidence, *UCLA Working Papers in Phonetics*, 69:3-29.

Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (1999) Domain-initial strengthening in four languages, *UCLA Working Papers in Phonetics,* 97:139-51.

Kelso, J. A. S. (1984). Phase transitions and critical behaviour in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology.*

Kelso, J. A. S. and Tuller, B. (1984). A dynamic basis for action system. In Gazzaniga, M. S., (Ed.), *Handbooks of cognitive neuroscience,* pp. 321-56. New York: Plenum.

Kelso, J. A. S., Saltzman, E., and Tuller, B. (1986). The dynamical perspective on speech production: data and theory, *Journal of Phonetics*, 14:29-59.

Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns; syllables, sentences, stress and speaking rate, in Huffman, M. K. and Krakow, R. A., (Eds.), *Phonetics and Phonology, Nasals, Nasalisation and the Velum V,* San Diego, CA: Academic Press, pp. 87-116.

Krakow, R. A. (1999). Physiological organisations of syllables: a review. *Journal of Phonetics,* 27:23-54.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of The Royal Society B Biological Sciences,* 369: 20130397.

Krivokapić, J. (2016). Articulatory phonetics and prosody. *Lecture presented in Aix Summer School on Prosody 2016.*

Kügler, P. N. and Turvey, M. T. (1987). *Resources for ecological psychology. Information, natural law, and the self-assembly of rhythmic movement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Labov, W., Yaeger, M., and Steiner, R. (1972). *A quantitative study of sound change in progress: Volume 1. Report on National Science Foundation Contract NSF-GS-3287.* Philadelphia, PA: University of Pennsylvania.

Ladd, D. R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica,* 18(1):19-38.

Ladd, D. R. (2008). *Intonational Phonology*, 2nd Edition. Cambridge: Cambridge University Press.

Ladd, D. R. (2014). *Simultaneous Structure in Phonology.* Oxford: Oxford University Press.

Ladd, D. R., Faulkner, D. and Schepman, A. (1999). Contrast "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America,* 106:1543-54.

Levelt, W. J. M. (1989). *ACL-MIT Press series in natural-language processing. Speaking: From intention to articulation.* Cambridge, MA: M.I.T. Press.

Maddieson, I. (1984). *Pattern of Sounds.* Cambridge: Cambridge University Press.

Manuel, S. and Krakow, R. A. (1984). Universal and language particular aspect of vowel-to-vowel coarticulation, *Haskins Laboratories Status Report on Speech Research,* SR 77/78:69-78.

Marin, S. (2011). Organisation of complex onsets in Romanian. In Laprie, Y., and Steiner, I., (Eds.), *Proceedings of the 9th International Semina on Speech Production, Montréal, June 20-23, 2011.* pp. 179-86.

Marin, S. (2013). The temporal organisation of complex onsets and codas in Romanian: A gestural approach. *Journal of Phontics,* 41:211-27.

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

Marin, S. and Pouplier, M. (2014). Articulatory synergies in the temporal organisation of liquid clusters in Romanian. *Journal of Phonetics,* 42:24-36.

Mücke, D., Nam, H., Hermes, A., and Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. In Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C., and Kühnert, B., (Eds.), *Consonant Clusters and Structural Complexity*, pp. 205-30. Berlin: Mouton de Gruyter.

Mücke, D., Grice, M., and Cho, T. (2014). More than a magic moment - Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics,* 44:1-7.

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). Compensatory shortening in monosyllables of spoken English, *Journal of Phonetics,* 20:225-39.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and mosaic structure. In Cole, J. and Hualde, J. I., (Eds.), *Laboratory Phonology 9.* Berlin: Walter de Gruyter. 483-506.

Nam, H., Goldstein, L., and Saltzman, E. (2009). Self-organisation of syllable structures: A coupled oscillator model. In Pellegrino, F., Marsico, E., Chitoran, I., and Coupé, C., (Eds.), *Approaches to phonological complexity.* Berlin: Walter de Gruyter. 297-328.

Nespor, M. (1993). *Fonologia.* Bologna: Il Mulino.

Nespor, M. and Vogel, I. (1982). Prosodic domains of external sandhi rules. In van Der Hulst, H., and Smith, N., (Eds.), *The structure of Phonological Representations.* Part I. Dordrecht: Foris Publications. 225-55.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology.* Dordrecht: Foris Publications.

Ohala, J. L. (1997). Aerodynamics of phonology. In *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL], Seoul, Linguistic Society of Korea,* pp. 92-97.

Ohala, J. L. (1999). The relation between phonetics and phonology. In Hardcastle, W. J., Laver, J. and Gibbon, F. E., (Eds.), *The Handbook of Phonetic Sciences*, 1st edition. London: Wiley-Blackwell.

Ohala, J. L. and Kawasaki-Fukumori, H. (1997). Alternatives to the sonotiry hierarchy for explaining segmental sequential constraints. In Eliasson, S. and Jahr, E. H., (Eds.), *Language and its Ecology: Essays in memory of Einar Haugen.* Berlin: de Gruyter Mouton. 343-65.

Ohala, J. L. and Solé, M. J. (2010). Turbulence and phonology. In Fuchs, S., Toda, M., and Zygis, M., (Eds.), *Turbulent Sounds: An Interdisciplinary Guide.* Berlin/New York: De Gruyter Mouton. 37-101.

Öhman, S. (1966). Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America,* 39:151-68.

Öhman, S. (1967). Numerical model of coarticualtion, *Journal of the Acoustical Society of America,* 41:310-20.

Pastätter, M. and Pouplier, M. (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. In *Proceedings of the 10th International Seminar on Speech Production, Cologne, Germany, 5-8 May 2014.*

Pastätter, M. and Pouplier, M. (2015). Onset-vowel timing of coarticulation resistance evidence from articulatory data. In The Scottish Consortium for ICPhS, (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK.*

Pierrehumbert, J. B. (1990). Phonological and phonetic representation, *Journal of Phonetics,* 18:375-94.

Pierrehumbert, J. B. and Beckman, M. (1988). *Japanese Tone Structure,* Cambridge, MA: M.I.T. Press.

Pouplier, M. (2011). The atoms of phonological representations. In van Oostendorp, M., Rice, K., Hume, B., and Ewen, C., (Eds.), *The Blackwell Companion to Phonology,* Volume I. Willey-Blackwell. 107-29.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In Fuchs, S., Weihrich, M., Pape, P., and Perrier, D., (Eds.), *Speech Planning and Dynamics*. Peter Lang. 63-96.

Pouplier, M. (2015). Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity. In Vayra, M., Avesani, C., and Tamburini, F., (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destruttutrazione della struttura sonora del linguaggio / Language acquisition and language loss. Acquisition, change and disorders of the language sound structure,* pp. 41-57.

Pouplier, M. and Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, 33(1):47-75.

Pouplier, M. and Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2(2):243-73.

Pouplier, M., Chen, L., Goldstein, L., and Byrd, D. (1999). Kinematic evidence for the existence of gradient speech errors. *Journal of the Acoustical Society of America,* 106(4).

Prieto, P., van Santen, J., and Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics,* 23:429-51.

Proctor, M. (2011). Towards a gestural characterisation of liquids: Evidence from Spanish and Russian. *Laboratory Phonology,* 2:451-85.

Recasens, D. (1996). An articulatory-perceptual account of vocalization and elision of dark /l/ in the Romance languages. *Language and Speech*, 39:63-89.

Recasens, D. (1999). Lingual coarticulation. In Hardcastle, W. J., and Hewlett, N., *Coarticualtion. Theory, Data and Techniques*. Cambridge: Cambridge University Press. 80-104.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America*, 111:2828-41.

Recasens, D. (2004a). The effect of syllable position in consonant reduction: evidence from Catalan consonant clusters. *Journal of Phonetics*, 32:435-53.

Recasens, D. (2004b). Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control. *Clinical Linguistics and Phonetics,* 18:593-603.

Recasens, D. (2006). Integrating coarticulation, assimilation and blending into a model of articulatory constraints. In Goldstein, L., Whalen, D. H., and Best, C. T., (Eds.), *Laboratory Phonology 8: Varieties of phonological competence.* Berlin/New York: Mouton de Gruyter. 611-34.

Recasens, D. (2007). Patterns of VCV coarticulatory direction according to the DAC model. In Prieto, P., Mascaró, J., and Solé, M. J., (Eds.), *Segmental and Prosodic Issues in Romance Phonology.* Amsterdam: John Benjamins. 25-40.

Recasens, D. (2012a). A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowel. *Journal of the Acoustical Society of America,* 132:412-20.

Recasens, D. (2012b). A phonetic interpretation of sound changes affecting dark /l/ in Romance. In Solé, M. J., and Recasens, D., (Eds.), *The Initiation of Sound Change. Production, Perception, and Social Factors.* Amsterdam: John Benjamins. 57-76.

Recasens, D. (2014). *Coarticulation and Sound Change in Romance.* Amsterdam: John Benjamins.

Recasens, D. (2016). What is and what is not an articulatory gesture in speech production: The case of lateral, rhotic and (alveolo)palatal consonants. Gradus. *Revista Brasileira de Fonologia de Laboratório,* 1:24-42.

Recasens, D. and Farnetani, E. (1990). Articulatory and acoustic properties of different allophones of /l/ in American English, Catalan and Italian. In *Proceedings of the 1990 International Conference on Spoken Language Processing, Kobe,* pp. 961-64.

Recasens, D., Fontdevila, J., and Pallarès, M. D. (1995). Velarisation degree and coarticulatory resistance for [l] in Catalan and German, *Journal of Phonetics, 23:37-52.*

Recasens, D., Pallarès, M. D., and Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America,* 102:544-61.

Recasens, D., Farnetani, E., Fontdevila, J., and Pallarès, M. D. (1993). An electropalatographic study of alveolar and palatal consonants in Catalan and Italian. L*anguage and Speech,* 36:213-34.

Recasens, D., Pallarès, M. D., and Fontdevila, J. (1998). An electropalatographic and acoustic study of temporal coarticulation for Catalan dark /l/ and German clear /l/. *Phonetica*, 55:53-79.

Recasens, D. and Pallarès, M. D. (2001). Coarticulation, assimilation and blending in Catalan consonant clusters, *Journal of Phonetics*, 29:273-301.

Recasens, D. and Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan. *Journal of Phonetics*, 34:295-318.

Recasens, D. and 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,* 125:2288-98.

Recasens, D. and Espinosa, A. (2010). Lingual kinematics and coarticulation for alveolopalatal and velar consonants in Catalan. *Journal of the Acoustical Society of America*, 127:3154-65.

Recasens, D. and Rodriquez, C. (2015). An investigation of lingual coarticulation resistance using ultrasound. In *Abstract Book of Ultrafest VII, Hong Kong, University of Hong Kong, 8-10 December,* pp. 41-43.

Recasens, D. and Rodriguez, C. (2016). A study of coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, 59:58-75.

Romero, J. and Riera, M. (2015). (Eds.)' introduction. In Romero, J., and Riera, M., (Eds.), *The phonetics–phonology interface: Representations and methodologies,* pp. ix-xxi. Amsterdam: John Benjamins.

Saltzman, E. (1986). Task dynamic coordination fo the speech articulators: A preliminary model. *Experimental Brain Research Series,* 15:129-44.

Saltzman, E. and Kelso, J. (1987). Skilled actions: A task dynamic approach. *Psychological Review,* 94:84-106.

Saltzman, E. and Munhall, K. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology,* 1:333-82.

Saltzman, E., Nam, H., Krivokapić, J., and Goldstein, L. (2008). A task-dynamic toolkit for modelling the effects of prosodic structure on articulation. In *Proceedings of the speech prosody 2008 conference,* pp.175-84.

Scobbie, J. M. (2007). Interface and overlap in phonetics and phonology. In Ramchand, G., and Reiss, C., (Eds.), *The Oxford handbook of linguistic interfaces*, pp. 17–52. Oxford: Oxford University Press.

Scobbie, J. M. and Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics,* 38:240-59.

Schmidt, R. C., Carello, C. and Turvey, M. T. (1990). Phase Transitions and Critical Fluctuations in the Visual Coordination of Rhythmic Movements between People. *Journal of Experimental Psychology, Human Perception and Performance,* 16:227-47.

Selkirk, E. (1982). *The Syntax of Words*. Cambridge, MA: M.I.T. Press.

Selkirk, E. (1984). *Phonology and Syntax: The relation between sound and structure.* Cambridge, CA: M.I.T. Press.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In MacNeilage, P. F., (Ed.). *The production of speech,* pp.109-36. New York: Springer.

Shattuck-Hufnagel, S. and Turk, A. E. (1998). The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America,* 103(5).

Shaw, J., Gafos, A., Hoole, P., and Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from patterns of tempora stability in articulation. *Phonology,* 26:187-215.

Silverman, K. and Pierrehumbert, J. (1990). The timing of prenuclear accents in English. In Kingston, J., and Beckman, M. E., (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech,* pp. 72-106. Cambridge: Cambridge University Press.

Solé, M. J., and Recasens, D. (2012). *The Initiation of Sound Change.Production, Perception, and Social Factors.* Amsterdam: John Benjamins.

Spreafico, L., Celata, C., Vietti, A., Bertini, C., and Ricci, I. (2015). An EPG+UTI study of Italian /r/. In The Scottish Consortium for ICPhS, (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK.*

Sproat, R. and Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementatio, *Journal of Phonetics,* 21:291-312.

Stetson, R. (1951). *Motor Phonetics: a Study of Speech Movements in Action (2nd edition).* Amsterdam: North Holland.

Stevens, K. N. (1972). Segments, features, and analysis by synthesis. In Kavanagh, J. F., and Mattingly, I. G., (Eds.), *Language by ear and by eye: The relationship between speech and reading.* Oxford, England: M.I.T. Press.

Stevens, K. N. (1989). On the quantal nature of speech, Journal of Phonetics, 17:3-45.

Sugahara, and Turk, A. E. (2009). Durational correlates of English sublexical constituent structure, Journal of Phonology, 26(3):477-524.

Tabain, M. (2013). Research methods in Speech Production. In Jones, M. J., and Knight, R. A., (Eds.), *The Bloomsbury Companion to Phonetics.* London: Bloosbury. 57-78.

Turk, A. E. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words, *Journal of Phonetics,* 35:445-72.

Turvey, M. (1990). Coordination. *American Psychologist,* 45:938-53.

Vallduvì, E. (1990). *The Informational Component.* PhD dissertation. University of Pennsylvania.

Vékás, D. and Bertinetto, P. M. (1989). Controllo vs. Compensazione: sui due tipi di isocrona. In Caldognetto, E. M., and Benincà, P., (Eds.), *L'interfaccia tra Fonologia e Fonetica,* pp. 155-62.

Vennemann, T. (1988). *Preference Laws for Syllable Structure. And the Explanation of Sound Change with Special Reference to German, Germanic, Italian, and Latin.* Berlin: de Gruyter Mouton.

Waltl, S. and Marin, S. (2010). Temporal organisation of three-consonant onset clusters in American English. *Poster presented at the 12th Conference on Laboratory Phonology,* Albuquerque, New Mexico.

Whalen, D., Noiray, A., and Bolaños, L. (2010). Articulation of Vowel Height in English Front Vowels: An Ultrasound Study. In Sock, R., Fuchs, S., and Laprie, Y., (Eds.), *Proceedings of the 8th International Seminar on Speech Production.*

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

Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Papers, 23.* Lund University.

Zharkova, N. (2007). *An Investigation of Coarticulation Resistance in Speech Production Using Ultrasound*. PhD dissertation. Queen Margaret University.