COARTICULATION

Veno Volenec
Faculty of Humanities and Social Sciences, University of Zagreb
vvolenec@ffzg.hr

Abstract
One of the central issues in contemporary phonetics is that of coarticulation. Speech segments do not and cannot occur isolated in real speech. As elements of a multisegmental utterance in continuous speech, all segments have neighbors which exert a certain degree of influence upon them. The vast array of articulatory adaptations that occurs as a result of the influence of one phonetic structure on another is what constitutes coarticulation in its broadest sense. Because of its ubiquitous nature in continuous speech, coarticulation has been and continues to be a central research area in experimental and theoretical articulatory phonetics. Naturally then, any integral theory or model of speech production, and to a lesser degree speech perception, must account for coarticulation. In this chapter we provide a general introduction to coarticulation, to the techniques used to measure it, and to the major theoretical and experimental contributions in explaining the nature and principles that govern it. Depending on the theoretical approach, the definition and scope of coarticulation may vary significantly, thus we aim to review and assess some of the more influential models of coarticulation, which constitute a fundamental part of contemporary phonetics. Theories and models of coarticulation also have significant consequences on phonological theory, especially the phonology-phonetics interface. Modelling coarticulation is central to understanding how the categorical, timeless and context-independent units suggested by phonology and the gradual, dynamic and context-dependent characteristics of continuous speech are related to each other. Therefore, in this chapter we also explore the place of coarticulation in the phonology/phonetics dichotomy, relating coarticulation to what may in essence be viewed as its phonological counterpart – assimilation.

Key words: phonetics, coarticulation, speech production, continuous speech, articulation

1. Introduction
Speech is a dynamic and continuously varying process. In normal speech, sounds do not occur in isolation, but rather as integral elements of a multisegmental utterance, ineluctably exerting influence on other adjacent elements within the utterance. Thus what is mentally a separable unit, call it a ‘segment’ or a ‘phoneme’, in real speech is realized as a continuous stream of articulatory movements. Each of these segments projects an articulatory requirement, a ‘target’ to be reached by the articulators. But in actuality these ideal targets are not necessarily reached, especially not in a uniform manner, for various reasons, perhaps most notably because one target is drawn off by having to aim quickly (within milliseconds) at the next one. The succession of these articulatory movements is seamlessly streamlined: during the production one articulatory movement (related to one segment), another movement (related to an adjacent segment) begins. This ubiquitous property of speech has an important corollary, namely that at any given time during the articulation of speech sounds, the vocal tract may be influenced by more than one articulatory movement (Farnetani – Recasens 2013: 316). So what mentally (or even orthographically) may be considered discrete elements, in speech occurs as the overlapping of articulatory movements. This phenomenon is one of the core properties of

---

speech and is usually referred to as ‘coarticulation’. Its precise definition, however, is a matter of debate and ongoing research in phonetics, and, as will be shown (§3), vastly depends on the theoretical framework adopted.

Coarticulation is a universal property of speech, but only in the sense that at least some coarticulatory effects are present in all (studied) languages of the world (Laver 1994: 376). However, the types of coarticulatory effects, their scope, direction, timing and causes all vary significantly from language to language, and from speaker to speaker (Perkel 1986: 261). Nevertheless, it is possible to reduce this great variability to a manageable number of categories by classifying coarticulatory effects along certain general criteria (§2).

As a feature of the articulatory apparatus, coarticulation may be measured (§4). Various measuring and imaging techniques are at our disposal, such as electropalatography, ultrasound, magnetic resonance imaging, computed tomography, electromyography etc., all with certain advantages and drawbacks. These techniques may be applied in coarticulatory research in order to improve understanding of articulatory organization during speech and to gain insight into various movement strategies present in coarticulation.

Theories and models of coarticulation have significant consequences on phonological theory, especially the phonology-phonetics interface. Modelling coarticulation is central to understanding how the categorical, timeless and context-independent units suggested by phonology and the gradient, dynamic and context-dependent characteristics of continuous speech are related to each other (Harrington et al. 2013: 244). This relation is particularly evident in the relationship between coarticulation and what may in essence be viewed as its phonological counterpart – assimilation (§5).

We shall explore these various topics pertaining to coarticulation in turn.

2. General Account of Coarticulation

There are many different and more or less compatible views of what ‘coarticulation’ is. While there is by no means total consensus, most definitions of coarticulation include either ideas about articulatory movements pertaining to different segments being somehow changed, adapted or blended, or ideas about different articulatory movements being relatively independent but overlapping in time.

The first set of ideas, where articulatory conflict is usually handled by averaging out the goals of the competing articulations (e.g. Kelso et al. 1986), may be considered an ‘adaptive’ approach to coarticulation, due to the fact that the form of a particular articulatory movement may be changed under the influence of adjacent movements. An explanation of coarticulation in such terms, therefore, may take this form:

Coarticulation is (…) to be regarded as a process whereby the properties of a segment are altered due to the influences exerted on it by neighboring segments (Hammarberg 1976: 576)

In co-articulation, there is a discernible degree of accommodation between articulatory features of a given string of adjacent segments (Laver 1994: 151)

Adjacent segments can show an articulatory ‘feature-sharing’ process at work as part of their accommodation to their occurrence in the particular context (op. cit.: 376)

The coordinatory phenomenon of accommodatory spreading of articulatory features has come to be known as co-articulation (op. cit.: 379)
In co-articulation, we can say that a setting finds its expression as an adjustment of the featural properties of one segment towards greater phonetics similarity to those of an adjacent segment (op. cit.: 397).

The second set of principles, where articulatory conflict is viewed as temporal overlapping of relatively independent articulatory movements, may be considered an ‘invariant’ approach to coarticulation, because in such a view articulatory movements remain stable, albeit somewhat masked by the onsets and offsets of adjacent articulations. A definition of coarticulation related to such an approach thus may take these forms:

Coarticulation refers to the fact that at any given point during an utterance, the influences of gestures associated with several adjacent or near-adjacent segments can generally be observed in the acoustic or articulatory patterns of speech (Fowler – Saltzman 1993: 173)

‘Coarticulation’ refers to the overlap of articulatory gestures associated with separate speech segments, and by extension to its acoustic effects (Keating 2003: 330)

(…) the movements of different articulators for the production of successive phonetic segments overlap in time and interact with one another: as a consequence, the vocal tract configuration at any point in time is influenced by more than one segment. This is what the term ‘coarticulation’ describes (Farnetani – Recasens 2013: 316)

In real speech articulations transition continuously from one to the next, with movements constantly overlapping, and sometimes conflicting. These various transitional processes fall under the rubric of coarticulation (Gick et al. 2013: 217).

These two approaches are complementary rather than mutually exclusive: accommodation between articulatory characteristics of segments may be viewed as a result of the overlapping of articulatory gestures needed to produce them, where ‘articulatory gesture’ is defined as “an actively controlled movement toward a presumed target configuration” (Recasens et al. 1997: 544). Some phoneticians use the term ‘coarticulation’ in a narrower and rather literal sense of simultaneous movement of two different articulators – as ‘co-production’, so to say (see §3.4 for elaboration). Under such a definition, for example, the lip rounding produced during the articulation of a consonant, anticipating the rounding of a following vowel (as in saw or sue), is coarticulation, but the adjustment of the tongue position for a velar consonant, anticipating the tongue posture for the following vowel, is not. This second kind of phenomenon may then be termed ‘adaptation’ or ‘accommodation’ (Clark et al. 2006: 86). Here we will adopt a wide perspective and consider both types of processes as coarticulation. Whatever the point of view may be, coarticulatory effects may be classified in terms of their direction, scope and articulators involved.

When it comes to direction, two types of coarticulation are distinguished, anticipatory and perseverative coarticulation. ‘Anticipatory coarticulation’ occurs when the articulation of a particular sound is affected by that of a later-occurring sound. Accordingly, in anticipatory coarticulation the articulatory gesture related to a particular sound begins during the production of one or more articulatory gestures that precede it. Its direction is sometimes (e.g. Ashby 2011: 152) depicted as ‘L ← R’, where ‘L’ and ‘R’ represent adjacent speech sounds, and ‘←’ means ‘is influencing’. Anticipatory coarticulation is also known under various different terms, such as ‘right-to-left’, ‘regressive’ or ‘forward’ coarticulation (Kent – Minifie 1977: 117; Lubker 1981: 129; Reetz – Jongman 2009: 39). An example can be found in the behavior of velar plosives [k] and [g] in English. Velar plosives become fronted when preceding front vowels, as in the word ski. Therefore, in anticipation of the upcoming front vowel [i], which requires a fronted tongue body, the tongue in [k] contacts the velum at a substantially more forward location. Conversely, when a velar plosive is adjacent to a back vowel such as [u], it becomes retracted, as in cool. The same holds for the velar plosive [g], in words such as geek or forgive.
where the tongue body is fronted, and *goo or ghoul*, where the tongue body is retracted. In a narrow phonetic transcription, such fronting and retracting of the tongue body may be noted with corresponding diacritic marks: [skːiː] and [kʰwːuːl]. The latter example shows another kind of anticipatory effect, namely labial coarticulation, indicated by the diacritic ‘w’ in transcription. During production of the plosive [k], which is usually unrounded, the lips nevertheless make a rounding gesture, in anticipation of the upcoming articulatory gesture characteristic for the vowel [u]. Vowels can also show effects of anticipatory coarticulation. Thus a vowel preceding a nasal consonant becomes nasalized because the velar port starts opening during the vowel in anticipation the upcoming nasal. This effect can be observed in a myriad of examples such as *seen* [sǐːn], *soon* [sʰwːn] etc. It is important to point out that a single sound may be affected by several distinct coarticulatory effects at the same time, depending on the totality of articulatory properties of the affecting sound: [kʷ] in *cool* shows two distinct coarticulatory effects, retraction and labialization, where retraction stems from the vowel’s backness property and labialization from its roundedness.

In ‘perseverative coarticulation’ the prolonged influence of a certain segment is exercised on one or more segments which follow it in a stream of speech. Accordingly, in perseverative coarticulation the articulatory gesture related to a particular sound is still being produced while the gestures of following sounds are being initiated. Therefore, in preservative coarticulation a simple ‘L → R’ relation holds. This type of coarticulation is also known as ‘left-to-right’, ‘carry-over’, ‘progressive’, ‘retention’ or ‘backward’ coarticulation (Kent – Minifie 1977: 117; Lubker 1981: 128; Reetz – Jongman 2009: 39). The devoicing of approximants following voiceless aspirated plosives can be viewed as an instance of perseverative coarticulation, with the voicelessness of the initial plosive carrying over to the following approximant, as in *please* [pʰliːz]. The aforementioned cases of anticipatory tongue body fronting and retracting work in an opposite, perseverative, direction as well, as in *seek* [sɪːk] and *arc* [ɑːk], respectively. It is important to mention that the designation of coarticulatory effects as anticipatory or perseverative depends strongly on the theoretical premise that at some level before the production of speech there is an underlying linear abstract segmental representation, what is usually called a ‘phonological representation’, where discrete precedence relations hold (Hardcastle 2006: 503; see Harris 2007: §6.2 for discussion, and Cairns – Raimy 2011 for a review of precedence relations in phonological representations). Otherwise, the directionality of such effects would be untenable, and perhaps even the entire concept of coarticulation depends on this theoretical position: “An intuitive concept of ‘segment’ underlies our recognition that there is a phenomenon of coarticulation requiring explanation” (Fowler 1980: 114).

Coarticulatory effects may also be classified in terms of segmental classes that influence each other, namely consonants (C) and vowels (V). Thus the so-called ‘V-to-C coarticulation’ refers to the influence of a vowel on a consonant, in either direction (Recasens 1982: 114; 2002: 2828), as shown in the example *cool* [kʰwːuːl], where the vowel [u] influences the consonant [k]. Another type is ‘C-to-V coarticulation’, where the consonant exerts influence on an adjacent vowel, like in cases of vowel nasalization. A particularly interesting case is the transconsonantal ‘V-to-V coarticulation’, initially described by Öhman (1966, 1967). Öhman found that in sequences V₁CV₂, where C is a stop consonant, the values of second formant transitions of V₁C and CV₂ depend not only on C and the adjacent V, but also on the identity of the transconsonantal vowel. Such a finding indicates that vowels in these cases are produced continuously, that is, the production of VCV utterances does not involve three linearly ordered successive articulatory gestures but rather a longer diphthong-like gesture on which a shorter consonantal gesture is superimposed. Not surprisingly then, cases of vowel harmony are sometimes also described as transconsonantal V-to-V coarticulation (e.g. Gafoğlu 1999: §4.1).
The fourth possibility is the ‘C-to-C coarticulation’, in which one consonant influences another. Hardcastle (1985: 253) provides electropalatographic evidence of C-to-C coarticulation in the word weakling, where the articulatory gestures of consonants [k] and [l] are overlapped during the period of about 28 ms. During that time the tongue is in contact with both the velum (related to [k]) and the alveolar ridge (related to [l]).

Different types of coarticulatory effects can also be related to different articulators involved in their production. Thus Farnetani and Recasens (2013: 316–317) recognize four different articulators involved in different types of coarticulation: lips, tongue, velum and larynx. Coarticulatory effects pertaining to perturbation of the lips give rise to ‘labial coarticulation’ (Farnetani 1999). Lip rounding (marked with ‘*’)) and lip spreading (unmarked in transcription), as in zoom [z*ːuːm] and keel [kʰːiːl] respectively, fall under this rubric. Tongue displacements along the high/low and front/back dimensions are the essence of ‘lingual coarticulation’ (Recasens 1999). The mentioned case from Hardcastle (1985), where the back tongue body position required for the articulation of [k] is, for a brief time, overlaid by the front tongue body position of [l], is an example of lingual coarticulation. ‘Velar coarticulation’, sometimes dubbed ‘velopharyngeal’ (e.g. Chafcouloff – Marchal 1999: 69), consists of nasality effects caused by the lowering of the velum. Nasalization of vowels adjacent to nasal consonants is a type velopharyngeal coarticulation and it works both in anticipatory (e.g. sang [sɛŋ]) and perseverative (e.g. can you [kʰənjʊ]) directions. Cases in which coarticulatory effects are related to vocal fold abduction and adduction, especially to the property of voicing, are known as ‘laryngeal’ or simply ‘voicing coarticulation’ (Hoole et al. 1999). The perseverative effect of approximant devoicing, as in please [pʰjʰiːz], is a type of laryngeal coarticulation.

Coarticulation is a phenomenon that does not only include the interactions between strictly adjacent segments, but also interactions between segments farther apart. By counting either the number of segments through which coarticulatory effects are spread, or by measuring the duration of such effects, we can determine the ‘scope’ or ‘span’ of coarticulation. A number of empirical findings related to coarticulatory scope exist. Öhman (1966), investigating V-to-V coarticulation in VCV clusters in Swedish, Russian and English, found that coarticulatory effects are independent of syllable boundaries. Similarly, Lisker (1978: 133) states that “lip-rounding and nasalization are segmental features of English that refuse to be contained within their ‘proper’ segmental boundaries, as these are commonly placed”. Amerman et al. (1970) note that speakers may anticipate an open vowel by beginning to lower the jaw during preceding consonants. Likewise, according to Benguerel and Cowan (1974) lip protrusion may be evident several consonants in advance of the rounded vowel for which it is required. They went on to show that in French labial coarticulation can extend up to 6 segments in anticipatory direction (Benguerel – Cowan 1974). Not surprisingly then, Amerman and Daniloff (1977) have shown that when a speaker articulates a CCV sequence, the tongue body may begin to move towards the vowel even during the first consonant. Laver (1994: 381) concludes that “the segmental domain of co-articulatory settings tends to vary with the type of setting involved”. In other words, the coarticulatory span of influence is hierarchically related to the type of coarticulation involved. Lingual coarticulation can be transsegmental, as shown by Amerman and Daniloff (1977), but is usually contained within the domain of the syllable. Velar coarticulation can cross both syllable and word boundaries: Moll and Daniloff (1971) showed that in English the velum can lower in anticipation of a nasal consonant several segments earlier, across such boundaries, provided that the influenced segments are all sonorants. Labial coarticulation seems to have the longest potential scope. Apart from Benguerel’s and Cowan’s (1974) demonstration of anticipatory labial coarticulation in French, Lubker et al. (1975), using electromyography, showed that in Swedish lip-rounding can start up to 600 ms ahead of a rounded vowel.
Causes for coarticulatory effects are various and not completely understood. Ashby (2011: 151) surmises that the function of coarticulation is “to create a continuum – a stream of speech without gaps”. One usual and often repeated cause for coarticulatory effects, especially for its perseverative direction, is the inertness of the articulatory organs (Daniloff – Hammarberg 1973; Hammarberg 1976). So in perseverative coarticulation the offset of an articulatory gesture extends in time, impinging on the upcoming gesture, because the articulator itself is too inert to move fast enough to complete the gesture entirely, before the next one is initiated. Also, factors such as the anatomical connections between articulators, the elasticity of articulator tissues, and aerodynamic factors are sometimes cited (e.g. Ohala 1981: 112) as further constraining the actions of the vocal apparatus. However, such explanations clearly do not hold for anticipatory coarticulation. Studies of Daniloff and Moll (1968) on labial coarticulation and of Moll and Daniloff (1971) on velar coarticulation revealed that the lowering of the velum in anticipation of a nasal consonant and the rounding of lips in anticipation of a rounded vowel can start up to four segments before the influencing one. These patterns clearly indicate that anticipatory coarticulation cannot be the product of inertia. Rather, the initiation of an articulatory gesture ahead of its time is explained by several different claims, some of which are experimentally confirmed. For one, the anticipation of articulatory gestures may be a production strategy that enables the articulatory organs to move fast enough (despite their inherent inertia) to produce segments and syllables at the rate of normal or faster-paced speech. So for example Daniloff (1973) claims that the tongue tip – the fastest of the articulators controlled by muscles – can perform only about 8 closures per second. Nevertheless, we are able to produce from 12 to 18 segments per second in normal speech, and up to 13 syllables in maximally fast speech (Škarić 1991: 298). Secondly, anticipatory coarticulation may be viewed as part of a general human tendency towards economy of movement, which is in constant conflict with communicative needs, or as Lindbloom (1990: 403) puts it: “Speakers can, and typically do, tune their performance according to communicative and situational demands, controlling the interplay between production-oriented factors on the one hand, and output-oriented constraints on the other”. In this sense, anticipation of articulatory gestures arises in order to minimize the movement of the articulators. Third, at the cognitive level, anticipatory coarticulation can be seen as evidence of the tendency for the brain to execute movements that have been planned ahead of time (Hardcastle 2006: 503; but see Farneti – Recasens 2013: 341 for a brief discussion). Fourth, there may also be some perceptual motivation for anticipatory coarticulation. Thus, as a result of anticipatory coarticulation, acoustic information on an upcoming segment is available to the listener before that segment is articulated, and this prior information may facilitate more accurate perception than would be the case if all acoustic cues were confined within the temporal boundaries of that segment (Kühnert – Nolan 1999: 9; Fowler 2005: 200; Harrington et al. 2013: 246–247).

3. Theories and Models of Coarticulation

3.1 Background

The basic concept of coarticulation was recognized at least as early as Henry Sweet’s A Handbook of Phonetics (1877). There Sweet (1877: 56) acknowledges the fact that, although it is useful to study sounds in isolation (analysis), in real speech sounds do not occur in isolation, hence the need to study speech as a “stream of incessant change” (synthesis):

We have hitherto considered sounds from purely analytical point of view, that is, each sound has been considered by itself, as if it were a fixed, isolated element. But in language sounds are combined together to form sentences, and many sounds only occur in certain fixed combinations. Hence the necessity for synthesis as well as analysis. Analysis regards each sound
as a fixed stationary point, synthesis as a momentary point in a stream of incessant change. Synthesis looks mainly at the beginning and end of each sound, as the points where it is linked on to other sounds, while analysis concerns itself only with the middle of the fully developed sound (Sweet 1977: 56).

However, it wasn’t until the 1930s that the term ‘coarticulation’ was devised, presumably in Menzerath’s and De Lacerda’s book Koartikulation, Steuerung und Lautabgrenzung (1933), and it took further 30 years for the term to become widespread in phonetics, after the publication of influential works such as Lehiste (1964), Öhman (1966, 1967) and Fant (1968). The pioneering studies also include – as accounted by Kühnert and Nolan (1999: 11–16), and Farnetani and Recasens (2013: §2.1) – the so-called ‘overlapping innervation theory’ by Joos (1948), where he proposed that phonetic segments influence each other in such a way that the neural command for each segment is an invariant wave “that waxes and wanes smoothly” and that “waves for successive phones overlap in time” (Joos 1948: 109).

During 1960s, in light of considerable experimental evidence, Lindblom (1963a; 1963b) devised his influential ‘target undershoot model’, where he posited that a string of phonetic segments is realized by a string of commands issued at a very fast rate; since the articulators’ maximum speed is restricted by biomechanical properties, they usually do not complete a given response before the onset of the next command. Thus the articulators often fail to reach their targets (i.e. they ‘undershoot’) in attempt to respond to more than one articulatory command simultaneously, which is in Lindblom’s view the essence of coarticulation. Articulatory timing was further elaborated by Kozhevnikov and Christovich (1965), within the framework of what Bladon and Al-Bamerni (1976) later dubbed ‘the articulatory syllable model’. There, based on a battery of elaborate tests on Russian speakers, the temporal extent of anticipatory (labial) coarticulation was taken as an indication of the size of the articulatory syllable: all the articulatory actions connected with one articulatory syllable are supposed to start at its beginning. As was already mentioned, in the 1960s Öhman (1966, 1967) developed his ‘VCV model of coarticulation’. His most important finding was that the in V1CV2 sequences vowels influence each other, so that from V1 to V2 one diphthongal articulatory gesture is realized, upon which a consonantal gesture is superimposed. Lindblom’s and Öhman’s models have an important commonality: both presuppose that at the level of motor commands, the instructions to the articulators are invariant. However, the difference is that for Öhman the presence of coarticulation results from the co-occurrence of vocalic and consonantal instructions, whereas for Lindblom it results from economy constraints tending to minimize articulator displacements.

An important insight into coarticulation of that time was Bladon's and Al-Bamerni's (1976) concept of ‘coarticulatory resistance’. They hypothesized that each allophone has a different value of coarticulatory resistance to which a speech production mechanism has access. Thus coarticulatory effects are allowed to spread until they are inhibited by a high resistance specification of some segment. The resistance values themselves were supposed to be determined by universal, language-specific and speaker-specific factors. However, Kent and Minifie (1977: 120) were critical of such explanations, remarking that coarticulatory resistance values seem to be based on many different poorly understood and quite general factors.

---

2 They wrote of this principle as follows: “Antagonistic vocal tract adjustments apart, coarticulation is inhibited only by ‘coarticulation resistance’ (CR) at some point in the succession of speech events. Each extrinsic allophone (and indeed each boundary condition) is assigned a value for CR by rules which may in some instances be language particular and in others quasi-universal. The CR value could be represented as a numerical coefficient attaching to a phonetic feature, say (3 CR), along the lines proposed by Chomsky & Halle (1968) for all other phonetic specifications in the phonetic system” (Bladon – Al-Bamerni 1976: 149).
A significantly different approach in the study of coarticulation was undertaken in the 1970s within a more phonologically oriented framework, namely within generative phonology of that time and its theory of distinctive features (Chomsky – Halle 1968). These early ‘feature-based’ or ‘feature-sharing’ models of coarticulation, such as Daniloff and Hammarberg (1973) and Hammarberg (1976), incorporated coarticulation within the phonological component of the grammar, despite Chomsky’s and Halle’s (1968: 295) earlier explicit banishment of coarticulation from the grammar. The basic idea of such feature-based models was that the phonetic representation, which is the end-point of the phonological derivation and the feature-spreading process, and which serves as an input into the speech mechanism, necessarily includes both articulatory and coarticulatory specifications. Because the various features are not always specified for all given input units, the model is equipped with a look-ahead operator that scans the forthcoming units to determine the next specified value of a feature. For this reason, such models are dubbed ‘look-ahead models of coarticulation’. Coarticulation then, following instructions from the grammar, smooths out the differences between adjacent sounds so that the transitions between them are minimized. Such approaches, however, ran into a myriad of contradictory data, most notably its failure to predict the extensive nature of perseverative coarticulation in VCVCV sequences (Recasens 1989), and its inadequacy in dealing with contradictory underlying feature specifications for two different segments (Sussman – Westbury 1981). Therefore, feature-based models were forsaken for a time, until their revival in the late 1980s and 1990s.

From these developments three approaches to coarticulation emerged in more recent times. On one front Lindblom (1983; 1989; 1990) continued to develop the view in which coarticulation is a matter of speech economy and adaptive variability (§3.2); on the other front Keating (1985; 1988; 1990a; 1990b), in her ‘window model’, revived the feature-based approach to Coarticulation (§3.3); the third approach, developed jointly by Fowler (1977; 1980; 1985), Brownman and Goldstein (1986; 1989; 1992), Recasens (2002) and others, sees coarticulation as a coproduction process to which articulatory gestures are the input (§3.4). These three approaches, roughly speaking, comprise the contemporary theories and models of coarticulation.3

3.2 Speech Economy Theory

Speech, the optimal form of human communication, seems to be torn between two antagonistic tendencies: the successful relaying of information and the economy of effort in doing so. As a consequence of this simple truism, speech production is never an invariant process; it is always the resolution of the interaction between the listener-oriented requirements of successful communication and the speaker-oriented requirement of speech economy. Lindblom (1983; 1989) noticed this ever-changing equilibrium and termed it ‘adaptive variability’: depending on situation and intent, speakers adapt their speech production to meet either of the requirements to a higher degree. So when a situation demands a high degree of phonetic precision (and thus more perceptual contrast), speakers are able to over-articulate; when this is not the case (e.g. in casual and informal speech), speakers turn to economy principles and under-articulate. This continuum of possibilities is termed ‘hyper-’ and ‘hypo-speech’, the first being the tendency to over-articulate, and the second to under-articulate (Lindblom 1990). Coarticulation plays a

---

3 The distinction between a theory and a model, within coarticulatory research, is explained clearly by Farnetani and Recasens (1999: 31): „The aim of coarticulation theories is to explain coarticulation, i.e. account for its origin, nature and function, while coarticulation models are expected to predict the details of the process bridging the invariant and discrete units of representation to articulation and acoustics. Coarticulation theories are also expected to explain how listeners overcome coarticulatory variability and recover the underlying message“.
central role in such a theory because going from hyper-speech to hypo-speech the degree of perceptual contrast decreases, while the degree of coarticulation increases. Therefore, within the theory of speech economy coarticulation is seen as an instantiation of the hypo-speech principle: the tendency to economize articulatory movements and to propagate low-cost motor behavior.

Within this theoretical framework, a revised version of Lindblom’s target undershoot model was embedded. In the original work (Lindblom 1963a), coarticulatory effects were seen as resulting from the inability of the motor system to respond to articulatory commands issued in very short temporal intervals. The faster the speech production, the greater the extent of undershoot and reduction. The revised model (Moon – Lindblom 1994) was enriched with considerations of speech style. In this view, speech rate and coarticulatory effects are still proportionately related, but it is also shown that a deliberately clear style of speaking inhibits the degree of undershoot, despite the relatively high speech rate. This finding is consistent with the speech economy theory in that it presupposes that a clear speech style leads to a higher degree of perceptual contrast (leaning towards hyper-speech) and therefore entails less pronounced coarticulatory effects, while spontaneous speech style (leaning towards hypo-speech) adheres to economy principles and therefore entails more coarticulation. All in all, a low-cost strategy, characterized by prominent coarticulatory effects, is the norm in natural speech (Lindblom 1975).

An important property of these findings is that coarticulation can be measured in statistical terms, namely by the so-called ‘locus equations’. A locus equation can be defined as “a straight-line regression fit to coordinates formed by plotting onsets of F2 transitions in relation to their coarticulated F2 midpoint ‘target frequencies’” (Sussman et al. 1993: 1256). In other words, each production of a CV syllable yields one datapoint consisting of an F2 value measured at vowel onset and an F2 value measured in the middle of the vowel. When all these points are plotted in an F2(onset) by F2(middle) space, a best-fitting line is drawn through this cloud of points such that each point is as close to this line as possible (Figure 1). The equation corresponds to a linear function ‘F2(onset) = k F2(middle) + c’, where ‘k’ is the slope and ‘c’ is the intercept of the regression line. The more that a consonant is influenced by a vowel, the less the formant transitions converge to a common locus and the greater the slope in the plane of vowel onset frequency by vowel target frequency. Thus in coarticulatory research, locus equations can be considered as indicators of the degree of coarticulation between a consonant and a following sonorant (Reetz – Jongman 2009: 204).
Krull (1987; 1989) used locus equations to quantify coarticulatory effects in CV syllables. She showed that labial-consonant loci\(^4\) undergo larger coarticulatory effects than dental-consonant loci and that coarticulation is larger in spontaneous speech than in words read in isolation. Moreover, Krull (1989) found for CVC syllables that prevocalic stop consonants undergo stronger anticipatory vowel effects than postvocalic consonants undergo carryover vowel effects.

### 3.3 Window Model of Coarticulation

As reviewed in §3.1, feature-based models held that coarticulation is a part of the grammar, presumably a part of its phonological component. However, it was difficult to reconcile the fact that phonology, as conceived at that time (1970s), operates with discrete and time-independent elements with the fact that coarticulatory effects entail continuous processes distributed in space and time. In the second half of the 1980s, Keating (1985; 1988; 1990b) proposed a new feature-based model of coarticulation, called the ‘window model’. She agreed with the view that phonological rules cannot account for the gradual nature of coarticulation, but contested the assumption that coarticulatory effects are solely a result of extra-linguistic, biomechanical properties of the speech apparatus. Instead, in Keating’s view, coarticulation is governed by phonetic rules of the grammar.

The standard rule-based approach to phonology, as laid out in Chomsky and Halle (1968), takes an underlying representation, either from the lexicon or from the output of the syntactic module, applies to it a set of ordered re-write rules, and thus derives the output representation. Both the underlying (phonological) and the output (phonetic) representation are specified in terms of binary features. Keating supplements this architecture with intermediate

---

\(^4\) A 'locus frequency', or simply 'locus', may generally be defined as the “apparent point of origin of the formant for each place of articulation” (Ladefoged – Johnson 2010: 199), the point of origin of the consonantal formants usually depending on the adjacent vowels.
phonetic rules, presumably governing coarticulation, and localizes them after the domain of application of phonological rules.

The input to the phonetic rule domain is a potentially underspecified phonological representation. Thus every feature of every segment may be assigned either ‘+’, ‘−’ or ‘0’ as its value. The phonetic component of the grammar then interprets these features by assigning a ‘window’ to every feature of each segment. Windows serve as a conversion medium from categorical phonological elements to gradual phonetic elements. These windows represent all the possible physical (temporal and spatial) parameters a segment can take, i.e. the maximal range of variability within a segment. Specified features (‘+’ or ‘−’) are associated with narrow windows and allow for little contextual variation; unspecified features (‘0’) are associated with wide windows and allow for more variation. The exact width of a window is also determined for each language from information on the maximum amount of contextual variability observed in speech. This allows for all intermediate degrees between maximally narrow and maximally wide windows.

Windows pertaining to different segments are connected by interpolation functions, called ‘contours’. Contours represent actual articulator trajectories over time and are governed by principles of efficiency and minimal effort (and in that respect are in line with speech economy theory). In principle, the narrower the window for any feature, the greater the coarticulatory resistance. Narrow and displaced windows require more curvature of the contour, which results in less coarticulation (Figure 2; b). When the windows are wider and aligned, however, a relatively straight line can be interpolated, which suggests more coarticulation (Figure 2; a).

![Figure 2](image-url)

**Figure 2** Window model of coarticulation.

In Figure 2, example (a) shows a VCV sequence, [apa], and its specification for the feature [low]. The consonant [p] is unspecified for that feature in the phonological representation (first row), and remains unspecified up to the phonetic level (second row). Because of its underspecification, the segment [p] is assigned a wide window, allowing greater contextual variability. Both vowels around [p] are specified as [+low], so they get a narrower window. Since [p] allows for greater variability, the most economical way to connect the three windows is by drawing a straight line through them. In articulatory terms, this translates to minimal adjustment of the relevant articulator (in this case the tongue body) during the production of [apa], and therefore to maximal coarticulation. Conversely, in example (b), the phonetic rule assigns a negative value (‘−’) to the phonologically unspecified segment [s]. Because the segment is now phonetically specified, it gets a narrow window, and since the value is negative,
the window is located at the bottom, misaligned with adjacent vowels. The most economical line that connects the three windows requires a lot of curvature. This translates to significant adjustment of the articulator during the production of [asa], and therefore minimal coarticulation.

Most of the research done within the window model of coarticulation aims to account for coarticulatory cross-language differences. Since in this model coarticulation is accounted for within the grammar, the differences in coarticulation between languages may originate either from phonology or from phonetics. So for example Cohn (1993), using the window model, reports different velar coarticulatory effects in English, where nasality is not phonemic, than in French, where nasality is phonemic. In nasalized English vowels, the shapes of the contours describe smooth trajectories from the (phonologically) non-nasal vowel to the nasal adjacent segment, suggesting the presence of velar coarticulatory effects that have a phonetic origin. However, vowel nasality is inherently phonological in French and is therefore categorical and not a result of phonetic coarticulatory effects. The plateau-like shape of the contour describing articulatory movement in French corroborates this claim.

Keating’s model was not met with universal acclaim. Browman and Goldstein (1993: 53) criticized the disparity between phonological and phonetic (under)specification and the general lack of constrained interaction between those two domains. Manuel (1987; 1990) argued that the existence of coarticulatory cross-language differences does not automatically imply that coarticulation must be accounted for within the grammar, by showing that language-specific coarticulation patterns may be completely unrelated to phonological constraints. Furthermore, Liker et al. (2008: 185) showed that the window model does not correctly predict the effects of V-to-V lingual coarticulation, since different segments with the same specification for the feature [high] under experimental ultrasound investigation showed different coarticulatory resistance in the same intervocalic position.

3.4 Coproduction Theory

The main tenet of the ‘coproduction theory’ is that the basic phonological units and basic phonetic units are not categorically different (i.e. abstract and discrete vs. concrete and continuous), but instead virtually identical: in all cases those units are articulatory gestures. In other words, the assertion is that what a speaker implicitly knows about speech sounds (phonology) is not different from what a speaker uses during speech production (phonetics). From this hypothesis then follows that coarticulation is any occurrence of the overlapping of articulatory gestures. Such a theory had been put forward in works such as Fowler (1977; 1980; 1983; 1985), Bell-Berti and Harris (1981) and Fowler and Saltzman (1993).

The coproduction theory is motivated by the fact that the distinction between abstract and timeless phonological units and concrete phonetic units that are organized in time relies on the idea of ‘extrinsic timing’, which precludes the possibility of any coherent explanation of coarticulation. Extrinsic timing here is intended to mean that the externalization (via speech) of the phonological component of the grammar, which operates in time-independent units like segments, depends on some kind of ‘translation’ of these timeless units into a temporally organized set of (presumably completely different) units of speech, say gestures. However, since in this view time is excluded from the phonological representation, the spoken utterance is given temporal coherence externally, from the process of actualization, which seems to suggest a significant dissociation between the phonological planning level (timeless) and its execution (timed) (Fowler 1980: 113, 117). The paradox that this leads to is that anticipatory coarticulation obviously includes some kind of planning and temporal distribution of
articulatory gestures, but at the planning level time just does not exist, since phonological representations are timeless. To overcome this adversity, Fowler (1980: 122) proposes a theory of ‘internal timing’, where phonological representations also include the temporal dimension. Thus the proposed units of phonology are ‘articulatory gestures’: planned and linguistically significant actions of the structures of the vocal tract, serially ordered, context-independent and dynamical (Fowler – Saltzman 1993: 172). This theory became the core idea behind Articulatory Phonology of Browman and Goldstein (1986; 1989; 1992), which propounds the claim that “phonology is a set of relations among physically real events, a characterization of the systems and patterns that these events, the gestures, enter into” (Browman – Goldstein 1992: 23), where gestures “are basic units of contrast among lexical items as well as units of articulatory action” (ibid.) that “can be used to capture both categorical and gradient information” (ibid.).

The coproduction theory adopts the basic insights of internal timing and Articulatory Phonology, namely that the input to any model of coarticulation are the dynamical and invariant articulatory gestures. It is thus proposed that gestures are not modified when actualized in speech. The intrinsic temporal structure of the gestures allows them to overlap in time, so that they are not altered by the context, but rather ‘coproduced’ with it (Figure 3). ‘Coproduction’ then entails the temporal overlap of articulatory gestures and serves as an explanatory theory of coarticulation.

![Coarticulation as coproduction of articulatory gestures](image)

As can be seen from Figure 3, both anticipatory and perseverative effects are accounted for by the same principle of gestural overlap. The prominence of a gesture increases and decreases in time and so does its influence on the vocal tract (and by extension on the acoustic signal). The dashed vertical lines delimit the intervals in which a certain gesture is most prominent. During the waning of gesture A, the production of gesture B is initiated. The time from the initiation of gesture B to the time where it becomes more prominent than gesture A represents the scope of anticipatory coarticulation, and the time between where gesture C becomes more prominent than gesture B and the time where the realization of gesture B is finished represents the scope of perseverative coarticulation. Logically then, the bigger the gestural overlap, the larger the coarticulatory effects, while the degree of gestural overlap is determined at the level of cognitive planning.

The most prominent model of coarticulation developed within the coproduction theory is the ‘degree of articulatory constraint (DAC) model’ (Recasens et al. 1997; Recasens –
Pallarès 2001; Recasens 2002; 2014), which aims to predict the scope and direction of lingual coarticulation by taking into account the demands imposed on the tongue dorsum during speech production. Different degrees of articulatory constraint, called ‘DAC values’, are attributed to consonants and vowels based on the involvement of the tongue dorsum in forming a constriction in the vocal tract during their production (Recasens et al. 1997: 545). A minimal DAC value (= 1) is attributed to segments which do not require any movement of the tongue dorsum: e.g. bilabials [p, b, m], labiodentals [f, v], vowel [a]. An intermediate DAC value (= 2) is assigned to segments for the production of which the tongue dorsum is not directly involved in closure or constriction formation but is somewhat constrained by movements of the primary articulator: e.g. alveolars [t, d, n, s], low vowel [a] etc. A maximal DAC value (= 3) is assigned to segments primarily articulated by the tongue dorsum: e.g. alveolarpalatal [ʃ, ʒ], palatals [j, ɲ, ʎ], velars [k, g, x], high vowel [i]. In other words, the more prominent and precise the requirements on the tongue dorsum during the formation of the constriction, the more the segments are constrained, while less constrained segments impose less strict requirements on the tongue dorsum. The concepts of ‘coarticulatory resistance’ and ‘coarticulatory dominance’ follow logically from this: a more constrained segment (higher DAC value) is more resistant to coarticulatory effects and exerts more prominent coarticulatory effects on adjacent phonetic segments, while a less constrained segment (lower DAC value) is less sensitive and less dominant in this respect (Recasens – Pallares 2001: 274). A great deal of experimental data has been gathered using the DAC model, such as investigation of the relationship between coarticulation, assimilation and blending (Recasens – Pallarès 2001), studies of coarticulatory direction (Recasens 2002), scope of coarticulation (Recasens et al. 1997), coarticulatory resistance (Fowler – Brancazio 2000; Zharkova 2007), the relation of coarticulation and sound change in Romance languages (Recasens 2014) etc. An obvious limitation of the model is its applicability to effects regarding only lingual coarticulation.

4. Measuring Coarticulation

Precise data on various coarticulatory effects can be gathered by measuring certain parts of the vocal apparatus during speech. These measurements may be either ‘direct’ or ‘indirect’. Direct measurements come from instruments that are in contact with the measured structure (e.g. electropalatography), while indirect measurements come from instruments that are remote from the structures of interest (e.g. imaging techniques) (Stone 2013: 9). While many different techniques may be used for obtaining data relevant to coarticulation, three procedures are used frequently: imaging techniques (especially ultrasound), electropalatography and electromyography.

4.1 Imaging Techniques

Imaging techniques register internal movements without directly contacting the structures. An obvious advantage of such techniques is that they are noninvasive and unobtrusive, not impinging upon normal speech production. Also, they provide an image of the entire structure (such as the tongue), rather than single points on the structure. The four common imaging methods in research of speech production are X-ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound. Of these, the ultrasound has proven to be the most productive method in research of coarticulation.

In a lateral X-ray image, an X-ray beam (a ray of light with the wavelength ranging from 0.01 to 10 nanometers) is projected on one side of the head. The rays then pass through all the tissue in the head and are recorded as an image on a plate on the other side. In X-ray, soft tissue
structures like the tongue are often obscured by harder structures like the jaw and teeth. Another, more significant drawback is that potential radiation hazards from long-term exposure to X-ray beams preclude the possibility of more extensive measuring. Although X-ray has played an important part in research of speech production (e.g. Kent 1972), it was seldom used in coarticulatory research due to its very low temporal resolution. A notable case is Wood (1996), who used X-ray to elucidate the differences between palatalization and coarticulation in Bulgarian, and to show that gestural programming for this case of palatalization is better described as coproduction of gestures than as feature spreading.

Computed tomography (CT) also uses X-rays, but in a significantly different manner: it records ‘slices’ of tissue as thin as 0.5 mm. A CT scan can image soft tissue more clearly than X-ray because the scanner rotates around the body taking multiple images, from different angles, of a single section of the body, and then creates a composite image. However, a CT still uses X-ray beams, so radiation exposure is still an issue. The temporal resolution is somewhat faster – about 15 images per seconds – but still not fast enough to provide for real time speech research, therefore it is also rarely an instrument of choice to phoneticians.

Magnetic Resonance Imaging (MRI) uses a magnetic field and radio waves to image a section of the tissue (see Stone 2013: 14–21 for a more detailed account of how MRI works). Numerous MRI procedures are available, such as high-resolution MRI, cine MRI, functional MRI etc., all using the same hardware, only in different ways. Although some phonetic research had been done using MRI (e.g. Story et al. 1996), very few investigations into coarticulation have actually used experimental MRI techniques (Stone 1999: 255), most likely due to its slow response time and low visual resolution. However, Kim (2012), using cine MRI, interpreted Korean palatalization as a gradual phonetic phenomenon, i.e. a type of V-to-V coarticulation, rather than a common phonological process involving consonants and front vowels.

Ultrasound produces an image by using reflective properties of sound waves (Stone 1999: 252). A sound wave (1 MHz frequency or higher) is emitted from the ultrasound’s transducer; it travels through soft tissue, and reflects back when it reaches an interface with a structure of different density, like bone or air. The reflected echoes are received by the transducer, processed by computer and displayed as a video image. By placing the probe under the chin (stabilizing it either by hand or by an immobilization kit), a clear image of the tongue body (contour of its outer rim) is acquired (Figure 4), which is appropriate for phonetic research. The sampling rate of fastest ultrasounds is about 90 Hz, adequate for most lingual processes, but to slow for the study of real-time laryngeal phenomena. Thus ultrasound imaging is usually constrained to research into lingual coarticulation. For example, Liker et al. (2008) used ultrasound to test the predictive capabilities of the window model versus the DAC model of coarticulation in accounting for lingual V-to-V coarticulation. Zharkova and Hewlett (2009) used ultrasound to measure lingual coarticulation in English and showed that the tongue contour during [t] adapts to the influence of the neighboring vowels approximately three times more than the tongue contour during [a] adapts to the influence of the neighboring consonants. Of all the imaging techniques used in research of coarticulation, the ultrasound seems to be the most productive one.
'Electropalatography' (EPG) is a technique used for recording the contact that the tongue makes with the hard palate during continuous speech (Gibbon – Nicholaidis 1999: 229). EPG is an invasive technique insomuch that the measuring equipment enters the human body, but it is completely painless and impinges upon normal speech production minimally. It requires the construction of an artificial palate, molded to fit precisely the speaker’s hard palate, that is layered with numerous individual electrodes (i.e. sensors) that register the tongue-palate contact. The electrodes conduct this signal through the wires that lead out of the mouth and to a computer which generates multiple images of the contact. There are approx. 60 scanning electrodes on an artificial palate (the exact number depends on the manufacturer) (Wrench 2007: 3), with the sampling rate up to 200 Hz (Stone 2013: 29). Because of its relatively fast sampling rate and high spatial resolution, EPG has a long and rich history in phonetic research, starting with Hardcastle (1972), and was easily extended to research of coarticulation.

*Figure 5* shows an EPG image of the pronunciation of the word *actor*. The sampling rate is 100 Hz, therefore the interval between individual frames is 10 ms. The frames are ordered in rows from left to right, with the alveolar region at the top and the velar region at the bottom. Tongue-palate contact is indicated by zeroes. The articulation of [k] starts at frame 289, where a complete closure at the velum can be observed. However, while the velar closure is still maintained, an alveolar closure related to [t] is also formed, around frame 301. Through frames 301 and 302 both closures are maintained for about 20 ms. Then from frame 303 the velar
closure starts to release, while only the alveolar closure is maintained, up to frame 316. From there on, there is a rapid decrease in tongue-palate contact during production of schwa. Coarticulatory effects are most prominent during frames 301 and 302, where both a velar and an alveolar articulatory gesture are maintained for a brief time. However, anticipatory effects can also be observed: during complete velar closure, the onset of the alveolar gesture is building up, starting from around frame 294. After the start of the velar release (frame 303), the velar gesture gradually starts to wane, but still exerts its influence on the vocal tract for several frames, which implies perseverative coarticulation as well.

A great deal of research into coarticulation had been done using electropalatography. Without any attempt at exhaustiveness, we can cite such examples as Recasens’ extensive work on lingual coarticulation (later within the DAC model) and studies of coarticulatory effects in particular languages, such as Hardcastle (1994; 1995) and Byrd (1996) for English, Kohler (1976) for German, Marchal (1988) for French, and Farnetani (1990) for Italian. Recasens (1990) used EPG data to show that vowels differ in coarticulatory resistance from effects exerted by adjacent consonants: consonant-dependent effects were shown to be larger upon those articulatory regions which do not intervene in the formation of a vowel constriction, therefore those lingual regions which are not involved in the formation of a vocalic constriction are left freer to coarticulate than those which are actively involved. Recasens et al. (1998) revealed in VIV sequences more prominent C-to-V effects for the Catalan dark [l] than for German clear [l], more so in the [i] context that in the [a] context, which is in agreement with the existence of high lingual requirements on the formation of two constriction places for dark [l]. Electropalatographic insights into coarticulation remain to be a highly productive research avenue.

It is worth mentioning that EPG and ultrasound show complementary data: EPG shows the place on the palate in contact with the tongue, but without showing which part of the tongue was used to make that contact; ultrasound shows the shape and position of the tongue during
articulation, but without showing its relation to the palate. The main drawback of EPG is that it requires a specific custom-made artificial palate for every different speaker, which makes mass data collecting an expensive endeavor.

4.3 Electromyography

‘Electromyography’ (EMG) is a technique for recording the changes in electrical activity within muscles (Hardcastle 1999: 270). EMG allows for the study of neuromuscular mechanisms that underlie the production of articulatory gestures during speech. The key part of the EMG system are the electrodes which detect ‘muscle action potentials’: the electrical activity accompanying muscle contraction. There are several types of electrodes, but those commonly used in phonetic research are either ‘hooked-wire’ or ‘surface’ electrodes. Hooked-wire electrodes are inserted into the body by a needle, after a local anesthetic has been applied. They record signals from individual ‘motor units’ (a single motor neuron and the group of muscle fibers which it innervates), identified as spikes in the electrical pattern, and are most useful for investigating muscles that are not close to the surface of the articulator, such as the genioglossus muscle of the tongue (Raphael et al. 2011: 288). Surface electrodes are used to study the response of the muscle as a whole, because they record the gross muscle activity from a large number of motor units. They are applied to an articulatory structure either by an adhesive substance or by suction (Harris et al. 1964; Allen – Lubker – Harrison 1972), and are useful for investigating muscles whose fibers are just below the skin, such as the orbicularis oris muscle of the lips.

Besides providing a measure of the strength of muscle contraction, EMG also provides temporal measures of muscle activity, such as the duration of muscle contractions and the relative timing of the activity of different muscles contributing to the movements of the structures used in speech production. The recording process is relatively straightforward: the electrodes register the electric potential and send the signal to a computer where it is displayed. The interpretation of the results is, however, much more complex, especially in terms of the functions of particular muscles and the resultant movements of speech organs: “It is extremely difficult to attribute a particular EMG output to a function in a given muscle and even more difficult to infer movement of a speech organ from such an output” (Hardcastle 1999: 273). A comprehensive introduction to recording and interpreting EMG signals is given by Loeb and Gans (1986).

How is data obtained by EMG relevant to a phonetician? Speech researchers want to know the relationship between the neuromuscular activity needed to produce speech sounds and the actual movement of particular speech organs that produced those sounds. Quite generally, EMG provides insight into the complex control mechanism used in speech production. It has been employed in the study of coarticulatory effects as well (see Farnetani 1999: 154–155 and Hardcastle 1999: 274 for extensive references). For example, Lubker and Gay (1982) used EMG to compare anticipatory labial coarticulation in Swedish and American English. They showed that the rounded vowels of Swedish were systematically produced with more extensive and more precise lip-protrusion movements than the rounded vowel of English, which suggests “that Swedish and American English speakers have learned different motor-programming goals” (Lubker – Gay 1982: 437). Boyce (1990) used EMG to compare patterns of protrusion movements for upper and lower lip for speakers of English and Turkish. They measured EMG activity in the orbicularis oris muscle and showed that Turkish speakers produce “plateau” patterns of movement rather than “trough” patterns, and unimodal rather than bimodal patterns of EMG activity, which suggests that “English and Turkish may have different modes of coarticulatory organization” (Boyce 1990: 2584).
The main drawbacks of hooked-wire electrodes is that they provide certain discomfort for the subjects which may affect the naturalness of their speech, and that specialist knowledge is needed for the insertion procedure, in order to avoid piercing any blood vessels. On the other hand, while surface electrodes are easy to apply and are non-invasive, their use is restricted to the investigation of superficial muscles such as the orbicularis oris, and there are substantial difficulties in attaching them onto moist intraoral surfaces.

5. Coarticulation and assimilation

In phonological literature ‘assimilation’ is usually defined as a phonological alternation in which two sounds become more alike (Lass 1984: 171; Zsiga 2013: 1919). This definition is, however, strikingly similar to any general definition of coarticulation (see §2), thus the question of the relation between these two phenomena naturally arises. Many different answers to this question have been proposed in the literature and here we will review some of them.

One simple explanation is offered by Laver (2001: 176):

(...) given the rate that segments follow each other in a stream of speech, (...) one segment may influence the articulatory characteristics of segments yet to be spoken, or be influenced by those that precede it. When such an influence crosses a word boundary, it is said to result in assimilation; when it is restricted to word-internal action, it is said to show coarticulation.

This explanation, however, is not quite consistent with concrete data. In Croatian, for example, voicing assimilation occurs automatically and invariably both within a single (tonic) word and within a phonological word (a tonic word plus a clitic) as shown in (1a–b), but is blocked between different phonological words (1c).

(1) Croatian voicing assimilation

a. Word internal

/s-broj-i- ti/ → [zbrojiti] 'add up'
/svat-b-a/ → [svadba] 'wedding'
/glad-k-a/ → [glatka] 'smooth' (fem.)
/urab-ts-a/ → [uraptsa] 'sparrow' (gen. pl.)

b. Within a phonological word

/s drv-ima/ → [zdrvima] 'with wood'
/kod kute-e/ → [kotktee] 'at home'
/grax bi/ → [graybi] ‘beans would’
/slog tee/ → [sloktee] ‘syllable will’

c. Between phonological words

/grax zivi/ → [grax zivi] ‘beans live’
/slog pada/ → [slog pada] ‘syllable falls’

Example (1), and many other such examples in different languages (see Zsiga 2013: 1921–1922 for Russian and Korean), shows that assimilation operates both word-externally and across word boundary, without any significant differences. This suggest that using the type of morpheme
juncture as a criterion for distinguishing between assimilation and coarticulation is not more than an arbitrary decision.

Another explanation pertains to different types of transcription:

The general convention in phonetics is to term coarticulation anything that will only show up in narrow phonetic transcription, reserving the term assimilation for changes involving a complete change of target. Often, assimilations are so extreme that they can be recorded in a much broader transcription, a phonemic transcription (Ashby 2011: 152).

In this view, the relation between coarticulation and assimilation is a matter of magnitude of the change. When a change is significant enough to be recorded in a phonological transcription, it is an assimilation; otherwise it is coarticulation and it may only be evident in phonetic transcription. In other words, what counts as one of these processes depends on the availability of symbols to indicate it and on conventional judgements about its auditory or linguistic significance. What is a “significant” change in this case, and what “significant enough” generally means in phonology and phonetics, has no definitive answer (though see below) and directs any coherent contemplation of how this criterion might distinguish between assimilation and coarticulation straight into the fundamental question of the relation between phonology and phonetics. Indeed, the matter of the relation between assimilation and coarticulation penetrates to the very core of long-standing dichotomies such as ‘phonetics and phonology’ (Beckman 1999), ‘gradient and categorical’ (Ernestus 2011), and perhaps ultimately ‘language and speech’, that is ‘langue and parole’ in structuralist terms (de Saussure 1959 [1916]: 9ff), ‘language competence and performance’ in traditional generative terms (Chomsky 1965: 4) and ‘I-language and E-language’ in more recent generative terms (Chomsky 1986: 22; 2013: 35). It has notably been argued that assimilation is a linguistic process pertaining to segmental phonological units, while coarticulation is a speech process pertaining to the gradual nature of phonetic elements (Chomsky and Halle 1968). As an extension of this view, the distinction between coarticulation and assimilation may be delineated on theoretical grounds: phonology, being a certain kind of implicit knowledge, a part of language competence and therefore ontogenetically prior to phonetics, encapsulates assimilation entirely, and thus assimilation can be defined within it as an alternation (an unfaithful mapping of distinctive features) imposed upon an underlying representation and apparent at the level of a derived phonetic representation; conversely, coarticulation, belonging to the phonetic domain, occurs subsequently in language performance (namely, speech) and may be viewed as an extra-linguistic process. Thus assimilation in principle entails a change of at least one distinctive feature, whereby coarticulation entails articulatory adaptations that are grammatically (phonologically) insignificant, in no way pertaining to the system of distinctive features. Also, it is fairly well established that coarticulation to a large extent depends on factors unrelated to language (in the sense of mind-internal linguistic competence, the so-called ‘I-language’), factors such as speaking style (Lindblom – Lindgren 1985) and speech rate (Hardcastle 1985). On the other hand, there seems to be no evidence that assimilation depends on any of these factors. However, this position is by no means universal and some influential theories contest most aspects of such an approach.

Attempts have been made (e.g. Hammarberg 1976; 1982; Keating 1990b) to integrate all that coarticulation stands for into the grammar, presumably into its phonological component, and to describe and explain coarticulation with the same atoms of representation used to model assimilation, namely with distinctive features. Keating (1990b) divides the phonological component of the grammar into phonological and phonetic parts, the former part consisting of phonological rules and representations, and the latter part, applying on the output of the former
part, consisting of phonetic rules and representations. In this view, any kind of feature-sharing process between segments is coarticulation, and the type of rules that govern this feature-sharing process determines whether coarticulation is phonological (i.e. assimilation) or phonetic:

Phonological rules of feature-spreading will produce partial or complete overlapping of segments, including assimilation over long spans. Phonological rules nonetheless have only limited access to segment-internal events. Phonetic rules, on the other hand, can (and more typically will) affect portions of segments, or affect them only slightly, or cause them to vary continuously in quality (Keating 1990b: 452).

Contrary to this position, the theory of Articulatory Phonology (see §3.4) argues that assimilation can be accounted for in terms of differences in articulatory organization and gestural overlap, without invoking any phonological feature change. Browman and Goldstein (1990) showed that a coronal closing gesture is still present in English phrases which sound as though a coronal nasal had become labial: for example, in the phrase seven plus, the [n] is not changed from [coronal] to [labial], but is overlapped by the following [p], and [n] and [p] articulated together sound like [m], as in se[vmp]lus. Browman (1995) argues that apparently categorical assimilations are just endpoints of a gradient distribution and a total case of gestural overlap. In other words, all assimilation is a kind of coarticulation. However, not all researchers agree that assimilation processes are describable in terms of gradual gestural overlap. Ladd and Scobbie (2003: 16) provide evidence that vowel assimilation at word boundaries in Sardinian is indeed categorical and conclude “that gestural overlap is on the whole not a suitable model of most of the assimilatory external sandhi phenomena in Sardinian, and that accounts of gestural overlap in some cases of English external sandhi cannot be carried over into all aspects of post-lexical phonology”.

Fowler (1990: §28.2) proposes a listener-oriented basis for discriminating between coarticulation and assimilation. When a listener perceives a sound change, i.e. when a sound change is “significant enough”; it is grammatically relevant, phonological, and therefore an assimilation. When a listener cannot be made to perceive a sound change, even though it can be proven to exist by instrumental measurement, the change is grammatically irrelevant, and therefore a coarticulation. Fowler (1990: 484) asserts that the “difference in listeners’ perceptions of coarticulation and assimilation reflects a real difference in the phenomena being described”, and remarks (op. cit.: 483) that “many processes identified as coarticulatory will count as nongrammatical, whereas true assimilations are grammatical”.

All in all, the theoretical and practical explorations of assimilation and coarticulation center around the question of whether these are two systematically different phenomena, related to different domains of the language faculty (i.e. to phonology and phonetics), or merely different instantiations of a single processes. In the first view assimilation is a part of phonology, and is therefore a mental process with results observable in the physical phonetic output,7 while coarticulation is a part of phonetics with no relevance for the language system.

---

5 “Phonological representations, which are essentially static and categorical, can be acted on by phonological rules, which change or rearrange the features which comprise segments. The output phonological representations can be acted on by phonetic rules, which interpret these features in time and space. Phonetic rules can thus, for example, assign a segment only a slight amount of some property, or assign an amount that changes over time during the segment” (Keating 1990b: 452).

6 “For example, no matter how hard we concentrate, we cannot make ourselves aware of the acoustic differences between the ‘intrinsic allophones’ (…) of /d/ in /dʌ/ and /dʌ/. (…) In contrast, we can be made aware of the durational differences between the vowels in heat and heed” (Fowler 1990: 483).

7 This position is in agreement with classic generative conceptions of phonology, namely that “all the work in phonology is internal to the mind/brain” (Chomsky 2012: 48).
Alternatively, assimilation may be viewed as a type of coarticulation, perhaps as its endpoint, but with lines between them in practice mostly blurred, if present at all. The matter is far from being resolved and will indubitably motivate extensive further research.

6. Conclusion

It is not possible to delve as deeply as one would like into various aspects of coarticulation covered in this chapter, therefore we have attempted to counterbalance the resulting superficiality by citing some of the most influential literature. An indispensable source of information on coarticulation is provided in Hardcastle and Hewlett (1999) and Farnetani and Recasens (2013), and of course in the references cited therein.

One conclusion that can reasonably be maintained is that coarticulation is a universal characteristic of speech and this chapter has, hopefully, demonstrated that given real life speech – all articulation is coarticulation. We have attempted to provide some general ideas about what the term ‘coarticulation’ denotes and how it can be classified (§2). Various types of coarticulatory effects testify to the incredible complexity of speech production. Because of its ubiquitous nature in continuous speech, coarticulation continues to be a central research area in experimental and theoretical articulatory phonetics. A number of different theories and models attempt to account for the nature and origins of coarticulation and to predict various aspects of its effects (§3). The theory of adaptive variability sees coarticulation as a consequence of speech economy. For models that have emanated from the theory of generative phonology, like the window model, coarticulation is a set of rules that converts categorical elements into gradual elements and serves as a link between language competence and performance. For coproduction theory coarticulation is the temporal overlap of articulatory gestures. The development of all these theories and models of coarticulation depends crucially on data provided by a number of different experimental methods (§4). These various techniques reflect the fact that contemporary phonetics is mainly an experimental science. Research into coarticulation also produces far-reaching implications for phonological theory and for the relationship between the mind-internal knowledge of language and its externalization through speech. These implications are particularly prominent in the complex relationship between assimilation and coarticulation (§5).

The existence of different and usually exclusive approaches to coarticulation and the fact that coarticulation remains a productive field of inquiry in phonetics permits us from arriving at specific and definitive conclusions. However, if one conclusion holds true universally, it is surely that “speaking is coarticulating gestures” (Farnetani – Recasens 2013: 320).
References


Cohn, A. C. (1993), Nasalisation in English: Phonology or phonetics. Phonology, 10, 43–81.


23


