Introduction

WILLIAM J. HARDCASTLE and NIGEL HEWLETT

Continuous speech is characterized by great variability in the articulatory and acoustic properties of segments. Sound segments are highly sensitive to context and show considerable influence from neighbouring segments. Such contextual effects are described as being the result of overlapping articulation or *coarticulation*. Coarticulation has been the object of much recent research in speech science and related disciplines.

Coarticulation was the focus of a large-scale research project ACCOR ('Articulatory–acoustic correlations in coarticulatory processes: a cross-language investigation') funded by the EU under the ESPRIT framework. ACCOR ran from 1992–1995 and brought together researchers from different parts of Europe in a unique concerted attempt to tackle the problem of coarticulation from a number of different theoretical perspectives and using a variety of different research methodologies. A cross-language approach was used to differentiate between those aspects of the phenomenon which could be attributed to universal features (due to factors such as inherent characteristics of the speech producing mechanism) and those which are language-specific and which could therefore be related to the phonological rules of the particular language.

The project aimed at a detailed description of the complex coordination between the main phonological systems underlying speech production and the resulting acoustic output. The articulatory dimensions under study were: the respiratory system (producing a flow of air), the laryngeal system (modifying the airflow by the valving mechanism of the vocal folds) and the complex system of supraglottal structures in the mouth and nose, such as the tongue, lip, jaw and soft palate shaping the vocal tract into different resonating cavities. It was possible in ACCOR to investigate specific articulatory processes in the seven languages of the project (English, French, German, Italian, Irish Gaelic, Swedish, Catalan) with a view to determining how these processes differed according to the different phonological systems. Another aim was to examine the functions of the different motor sub-systems with respect to coarticulation.

This book grew out of the ACCOR project. Most of the authors were original members of the ACCOR consortium and the theoretical perspectives and methodology largely represent the approach used by them.

The book is divided into four sections. Part I contains two chapters which provide the background and history of coarticulation in phonetic theory. Kühnert and Nolan make play with the ambiguity of the word 'origin' in their title, *The origin of coarticulation*, and discuss the origin of the concept of coarticulation in the history of phonetics and the origin of coarticulation in the process of human speech production. Farnetani and Recasens provide a detailed review of alternative contemporary models of coarticulation within phonetic theory in their chapter *Coarticulation models in recent speech production theories.*

The five chapters of Part II are concerned with research findings on coarticulation, and each is focused on a particular aspect of the speech production mechanism. The spreading of nasality from a consonant to a neighbouring vowel was one of the earliest coarticulatory phenomena to be observed and studied systematically. *Velopharyngeal coarticulation*, by Chafcouloff and Marchal, reports on the evidence concerning nasal coarticulation and its theoretical implications. The tongue is a complex, mobile organ which plays a major articulatory role in the production of all vowel sounds and the majority of consonants. In his chapter on *Lingual coarticulation,* Recasens provides a review of empirical findings and theoretical issues and points to important outstanding questions concerning the control processes for lingual coarticulation, such as the question of whether different areas of the tongue (tip, front, back) should be modelled as being quasi-independent of each other. The chapter on *Laryngeal coarticulation*, by Hoole, Gobl and Ní Chasaide, has two parts. Laryngeal vibration has an on/off function in speech and Philip Hoole discusses *Coarticulatory investigations of the devoicing gesture* in the first section of the chapter. Mode of laryngeal vibration is also subject to coarticulatory influence and it is this aspect which is explored in the second section, *Voice source variation in the vowel as a function of consonantal context*, by Gobl and Ní Chasaide. Lip movements, which are relatively accessible to visual inspection, have provided a rich source of data on coarticulation in speech. Farnetani's chapter on *Labial coarticulation* reviews the findings from research in different languages, informed by a description of the muscles involved in the control of lip movement, and explores the implications for competing theoretical models. As Farnetani points out, movements of the lower lip and the jaw are interdependent to a considerable extent. Fletcher and Harrington's chapter, *Lip and jaw coarticulation*, focuses particularly on the coarticulatory influences, involving these structures, upon vowels. They discuss the influence of an adjacent consonant and the influence of another vowel in an adjacent syllable.

We can be reasonably certain that coarticulation is a universal characteristic of human speech production. How particular coarticulatory processes compare across different languages and whether or to what extent coarticulatory processes impinge upon phonological representations are issues explored in Part III. Manuel's chapter, *Relating language-particular coarticulation patterns to other language-particular facts*, focuses on the results of research into non-Indo-European languages. Beckman, in her chapter *Implications for phonological theory,* situates the findings on coarticulation within the issue of the relationship between phonetics and phonology as levels of representation in linguistic theory.

Contemporary phonetics is a mainly experimental science and the development of explanatory models depends crucially on data provided by a number of rather different techniques, from electromyography to acoustic analysis. The details of these techniques may be of less interest to beginning students or to those whose interest in coarticulation lies mainly with its implications for phonological theory. However a discussion of techniques is highly relevant for those intending to undertake empirical research in the area. Part IV contains seven chapters each of which is concerned either with a particular experimental technique or with techniques for investigating coarticulation involving a particular organ of the speech production system. The techniques covered reflect the interests of those involved in the ACCOR project and no claim is made for a completely comprehensive coverage. For techniques not covered here, for example strain gauge and optoelectronic transducers for measuring lip/jaw movements, and electrolaryngography, the reader is referred to other sources such as Abberton and Fourcin (1997), Baken (1998) and Stone (1996).

In practice of course, a particular experimental technique may anyway apply solely, or primarily, to one part of the production mechanism. This is the case with *Palatography* for example, as described by Gibbon and Nicolaidis. Palatography is now a well-established technique for recording the location of, and extent of, lingual contact with the hard palate during speech and it has gathered a comparatively large body of research literature. Direct imaging is probably the ideal method for examining articulatory movements – or it would be, given a system that was sufficiently comprehensive, convenient and risk-free. Stone's chapter, *Imaging techniques,* provides an overview of the advantages and disadvantages of the techniques that are currently available. Hoole and Nguyen's chapter describes *Electromagnetic articulography*, a relatively recent but very promising technique for tracking tongue (and lip) position over time during speech. *Electromyography*, described by Hardcastle, is distinctive in that it is targeted at a stage further back (so to speak) from articulation in speech production, namely that of the neural innervation of articulatory events. Variants of the technique are described in detail together with a discussion of their advantages and disadvantages. Chafcouloff's *Transducers for investigating velopharyngeal function* reviews a variety of techniques available for gaining evidence on the control of velopharyngeal valving. *Techniques for investigating laryngeal articulation*, by Hoole, Gobl and Ní Chasaide, is divided into two sections, which mirror the division of chapter 5 by the same authors. It is the operation of vocal fold abduction in the signalling of a voiceless segment that is the topic of the first section, by Hoole, while the second section, by Gobl and Ní Chasaide, describes techniques for investigating vocal fold vibration itself. The technique with the most extensive history in experimental phonetic research is, of course, that of acoustic analysis and descriptions of acoustic analysis are already plentiful in the literature. Recasens' chapter on *Acoustic analysis* therefore takes a rather different approach from that of the other authors. After briefly surveying variant techniques of analysis and display he discusses applications of acoustic analysis to each of the classes of speech sounds, with special reference to the phenomenon of coarticulation.

Cambridge University Press
978-0-521-44027-1 - Coarticulation: Theory, Data and Techniques
Edited by William J. Hardcastle and Nigel Hewlett
Excerpt
More information

Part I
Theories and models

1

The origin of coarticulation

BARBARA KÜHNERT and FRANCIS NOLAN

What is coarticulation, and why does it exist?

The title of this chapter is deliberately ambiguous. *Origin* refers both to the history of the scientific concept of coarticulation and to the question of what causes the phenomena in speech which are known as coarticulation. The history of the concept will be dealt with later, while the reasons why there are phenomena in speech which we can characterize as coarticulation are dealt with explicitly below, as well as implicitly in the discussion of the history of coarticulation. There is even a third sense of 'origin' which is dealt with briefly in this chapter, namely the way in which coarticulation develops as a child learns to speak.

Coarticulation, very broadly, refers to the fact that a phonological segment is not realized identically in all environments, but often apparently varies to become more like an adjacent or nearby segment. The English phoneme /k/, for instance, will be articulated further forward on the palate before a front vowel ([k̟iː] 'key') and further back before a back vowel ([k̠ɔː] 'caw'); and will have a lip position influenced by the following vowel (in particular, with some rounding before the rounded vowel in [k̠ʷɔː] 'caw'). As here, some instances of coarticulation are available to impressionistic observation and constitute an important part of what has traditionally been thought of as allophonic variation. In many other instances, however, the kind of variation which a segment undergoes only becomes apparent from quantitative instrumental investigation, either of the acoustic signal or of speech production itself.

It is essential to the concept of coarticulation that at some level there be invariant, discrete units underlying the variable and continuous activity of speech production. If this were not the case, and, for instance, the mentally stored representation giving rise to a production of the word 'caw' were a fully detailed articulatory plan, then when that word was spoken (in isolation at least) there would be no question of a process of coarticulation – the word would simply correspond to a set of instructions for the time-varying activity of the articulators, and sub-word segments would not exist in any sense, and could therefore not undergo 'coarticulation'.

There are, however, good reasons to assume that the 'componentiality' which characterizes language (sentences made up of words, words made up of morphemes, and so on) extends down at least to the level of phoneme-sized segments. From the point of view of storing and accessing the mental lexicon, it would be massively less efficient if every entry were represented by its own idiosyncratic articulatory (or indeed auditory) properties, rather than in terms of some kind of phonemic code – a finite set of symbols abstracted from phonetic behaviour. Ironically, studies of coarticulation itself, based on the premise of phoneme-sized segments at some level of representation, lend independent support to the premise. For instance such studies have conspicuously not shown, to take a hypothetical example, that the onset of lip rounding is consistently different between the words 'caw' and 'caught' or that the degree of velar fronting is consistently different in each of the words 'key', 'Keith' and 'keen'. If each word were represented holistically and independently, this result would have to be put down to coincidence. Repeated across the lexicon, the coincidence would be huge and extraordinary. On the other hand this kind of regularity across words is predicted by a view in which these sets of words are represented with common sequences of abstract elements, for instance /k/, /ɔː/ in the case of the first set and /k/, /iː/ in the case of the second, and in which the articulatory realization of those sequences is governed by regular principles of integration – that is, by principles of coarticulation.

Accepting the role of segments we may then ask a complementary question: why, if a linguistic system operates in terms of discrete and invariant units (let us say phonemes), are these units not realized discretely and invariantly in speech? After all, there is a medium in which this does happen. When we use a typewriter, each letter is realized on the paper separately from the preceding and following ones, and is realized identically (as near as makes no difference) every time it is typed. English orthography has of course strayed somewhat from a phonemic analysis, but all alphabetic writing systems are in essence a way of representing the phonemic code visually. Why does the speech mechanism not behave like an acoustic typewriter?

One reason is perhaps that we do not have a separate vocal tract for each phoneme, in the way that an old-fashioned typewriter has a separate 'hammer' to produce each letter. Instead, a single vocal tract has to alter its shape to satisfy the requirements of all the sounds in a sequence. The vocal tract is governed by the laws of physics and the constraints of physiology, but (also unlike the typewriter) it is producing its communicative artefact in 'real time'. It cannot move instantaneously from one target configuration to the next. Rather than giving one phoneme an invariant articulation, and then performing a separate and time-consuming transition to the next, it steers a graceful and rapid course through the sequence. The result of this is coarticulation. It is perhaps rather like a slalom skier, whose 'target' is to be to the left and to the right of successive posts, and who minimally satisfies this target with his skis as they zig-zag down the hill, but whose body pursues a more direct course from the top of the hill to the bottom. In the written medium, it is not typing but handwriting which provides the closer analogy to speech. Examine the occurrences of a given letter in any fluent handwriting, and its realizations will vary. Maybe the tail of the 'y' will make a closed loop if a letter follows, but not when it is at the end of a word, and so on. The more fluent the handwriting, the less possible it is to pick out discrete letters, and concomitantly the more each letter's shape will be a product of its environment.

It would be misleading to think of coarticulation in speech as if it were an imperfection in the way language is realized. Speech and language have evolved under the influence of the constraints of the vocal mechanism, and there is no reason to suppose that the relationship between language and the vocal mechanism is not a satisfactory one. The phenomenon of coarticulation may in fact bring advantages beyond the efficient integration of the realizations of successive phonological units.

In particular, the fact that the influence of a segment often extends well beyond its own boundaries means that information about that segment is available to perception longer than would be the case if all cues were confined inside its boundaries. As pointed out by early perceptual theories, the possibility of 'parallel processing' of information for more than one phoneme probably allows speech to be perceived more rapidly than would otherwise be feasible. The possibility that the origin of coarticulation lies not only in the requirements of the articulatory mechanism, but in those of our perceptual system, cannot be discounted.

To recapitulate: the concept of coarticulation entails the hypotheses that at some level speakers make use of a representation in terms of abstract phonological segments, and that there are regular principles governing the articulatory integration of those segments in speech.

Given a coarticulatory standpoint, one way to conceptualize part of the variation in the realization of /k/ in 'caw' and 'key' above is to think in terms of the velar stop having a 'target' place of articulation, which is then modified to facilitate the integration of /k/ with the tongue movement for the following vowel segment. From this perspective coarticulation involves a spatial or configurational modification of the affected segment. Alternatively we can break away from thinking in terms of spatial targets for successive segments, and regard coarticulation as the spreading of a property from one segment to a nearby one. For instance if we concentrate on lip activity in the example above, we noted that the lip rounding always associated with the vowel of 'caw' is also present on the consonant preceding it: [k̠ʷɔː]. In 'key' there is no lip rounding on the velar. From this alternative (temporal rather than spatial) perspective, what matters is when articulatory movements begin and end relative to each other. The rounding of [ɔː] has begun during, or even at the start of, the [k].[1]

It might appear from this example that the spatial/temporal distinction depends on whether or not a property involved in coarticulation is crucial to the identity of the affected segment. A velar stop involves a raising of the tongue dorsum, and it is merely the precise location of that raising which is affected by a following [iː]. On the other hand lip activity is not required for a velar stop, and so, in the word 'caw', the lip movement can be anticipated. It is unlikely, however, that a consistent division can be sustained between 'crucial' properties and other properties of a segment. It may be that the absence of lip rounding on the /k/ of 'key' is just as crucial to the perception of this word as the presence of lip rounding on 'caw' (cf. the 'trough' of rounding found on the fricative in /usu/ sequences – see later). So a simplistic linking of spatial coarticulation to crucial properties, and temporal coarticulation to inessential properties, is not valid.

In fact the very distinction between spatial and temporal coarticulation breaks down as soon as we take a more abstract view of articulation. Recent models of speech production (see later) hypothesize that each segment is associated with an abstract control structure which is in tune with the mechanical properties of the vocal tract, and which defines that segment in terms of dynamic activity of the articulators. In such a view the distinction between space and time becomes less clear. The control structure for [k] would overlap that for [iː] in time in the phonetic plan of an utterance of 'key', but the competing demands of the two would result in a spatial compromise in the resultant articulation. A current hope, therefore, is that a definition of segments not in terms of superficially observable articulatory movements and positions, but in terms of more abstract articulatory control structures, may lead to a more general and unified description of the variety of coarticulatory phenomena.

This section has summarized the origin, in the nature of language and its vocal realization, of the phenomena which are conceived of as coarticulation. We now turn to the origin of the concept of coarticulation in the history of phonetics and its widespread adoption as the basis of a research paradigm.

The historical perspective

The early history

The term 'coarticulation' dates from the 1930s when Menzerath and de Lacerda published *Koartikulation, Steuerung und Lautabgrenzung* (1933). However, the fact that speech sounds influence each other and vary, often substantially, with changes in the adjacent phonetic context had already been known for centuries, while the demonstration that the stream of speech cannot be divided into separate segments corresponding to 'sounds' (or 'letters') coincided with the establishment of experimental phonetics as an independent discipline.

Before experimental techniques were introduced in the study of speech sounds the main tools were 'direct observation' and introspection. Brücke's (1856) and Bell's (1867) insights for German and English, respectively, which laid the foundations for academic phonetics, were based upon such subjective observations. Not surprisingly, early phoneticians shared the assumption that alphabetical letters have corresponding physical realizations in the form of single sounds. The leading idea at the time was that every sound has a static positional (steady state) phase and that different sounds are connected by short transitional glides. The concept of such transitional glides ('Übergangslaute'), which allow the stream of speech to be continuous, was formulated most explicitly by Sievers (1876). For instance, he described the production of a syllable such as 'al' in such a way that there exists neither a pure 'a'-sound nor a pure 'l'-sound during the linking movement of the tongue but a continuous series of transitional sounds which as a whole were referred to as a 'glide'. For an overview of early phonetics, see Tillmann (1994).

There were, however, some indications in the early literature that the classical view might not capture the whole story. Sievers (1876) himself acknowledged the possibility that, in certain sound combinations, articulators which are not involved in the current sound production might anticipate their upcoming configuration as long as it does not compete with the requirements of the present sound. Examples are the rounding of the lips during the production of /k/ in /ku/ or the preparation of the tongue position for the vowel during the consonant in syllables such as /mi/. And from a more theoretical perspective, Paul (1898: 48) wrote: 'A genuine dissection of the word into its elements is not only very difficult, it is almost impossible. The word does not correspond to a sequence of a specific number of independent sounds, each of which could be represented by a sign of the alphabet, but it is, in fact, always a continuous row of an infinite number of sounds . . .'[2]

Historically, phonetics moved towards experimental research during the last quarter of the nineteenth century. 'Kymography' allowed the mechanical