

8

Connected speech and coarticulation

We normally speak by producing a continuous, connected stream of sounds, except when we pause. In most languages we can find occasional cases where a speaker makes a single segment in isolation with no sound preceding or following it (in English, for example, we can say 'ah' /ɑ:/ if we make an interesting discovery, or 'sh' /ʃ/ if we want to call for quiet), but such cases are rare. The usual situation is for segments to fit closely together with each other. We have seen that phonetics tends to look on speech as a sequence of segments. However, to imagine these segments as discrete and independent of each other would be quite wrong. In every language we find that segments have a strong effect on other segments which are close to them. The process by which a segment is modified by its neighbours is called **assimilation**, and the description of assimilation has been a part of phonetic description for a long time. As we will see later, much recent phonetic research in this area refers to **coarticulation** instead, and we will discuss whether there is any significant difference between these terms. Another phenomenon of connected speech is **elision**, the process by which sounds that would be pronounced in slow, careful speech seem to disappear.

Assimilation

Let us look at some examples of assimilation. In French, a word-final voiceless consonant will often become voiced if followed by a voiced segment. For example, the word 'avec' on its own is pronounced /avek/, but when it is followed by a word beginning with a voiced consonant such as /v/ in 'vous' /vu/, we usually hear

/aveg/. So the phrase 'avec vous' is often pronounced /aveg vu/. In English, we also find assimilations of voice, but it is more common to find them in the form of *loss of voice*, or **devoicing**. If the word 'have' occurs in final position, its final consonant /v/ will usually have some voicing, but when that /v/ is followed by a voiceless consonant it normally becomes completely voiceless; thus 'I have to' is likely to have the pronunciation /aɪ hæf tu/.

Assimilation, then, is concerned with one sound becoming phonetically similar to an adjacent sound. The examples given so far are of anticipation, where a sound is influenced by the sound which follows it; another term frequently used for this type is *regressive* assimilation. We also find cases where the assimilation can be called *progressive*: here, not surprisingly, the process is for a sound to take on characteristics from a sound which precedes it. In general, this effect is less frequently found, though it is difficult to explain why this should be so. Historically, it must have been effective in English in order to produce the different pronunciations of the -s ending: the plural of 'cat' /kæt/ is 'cats' /kæts/ with a final /s/; the plural of 'dog' /dɒg/ is 'dogs' /dɒgz/ with /z/. The voicing of the suffix is conditioned by the voicing of the preceding final consonant.

Assimilations are traditionally classified into three main types, though as we shall see this classification is not completely adequate.

- 1 One type is assimilation of voice (we have seen examples of this taken from French and English); this may take the form of a voiced segment becoming voiceless as a consequence of being adjacent to a voiceless segment; alternatively, a voiceless segment may become voiced.
- 2 Another type is assimilation of place: this refers to changes in the place of articulation of a segment (usually a consonant). A well-known case is that of English word-final alveolar consonants such as /t, d, n/: if a word ending in one of these consonants is followed by a word whose initial consonant has a different place of articulation, the word-final alveolar consonant is likely to change so that it has the same place of articulation. Thus the word 'that' /ðæt/ may be followed by 'boy' /bɔɪ/ and become /ðæp/ (thus 'that boy' /ðæp bɔɪ/), or it may be followed by 'girl' and become /ðæk/ (thus 'that girl' /ðæk ɡɜ:l/).

A third type is assimilation of manner: here one sound changes the manner of its articulation to become similar in manner to a neighbouring sound. Clear examples of this type are not easy to find; generally, they involve a change from a 'stronger' consonant (one making a more substantial obstruction to the flow of air) to a 'weaker' one, and are typical of rapid speech. An English example could be a rapid pronunciation of 'Get some of that soap', where instead of the expected /get sʌm əv ðæt səʊp/ the speaker says /ges sʌm ə ðæs səʊp/, with /s/ replacing /t/ in two words.

We should now consider what the reason is for these processes. We must remember that in most cases several articulators are involved in making a speech sound, and that they are not capable of moving instantaneously. In the example of French consonant voicing, the final consonant is intrinsically voiceless, but in the example given, it is preceded by a fully voiced vowel, and is followed by a voiced consonant. To produce a voiceless consonant usually requires the opening of the vocal folds to prevent voicing from happening. If the vocal folds are instead left in the position appropriate for the voicing of the vowel context, the result is likely to be that the consonant is produced with voicing, and we can suppose that this is why the consonant becomes voiced. This argument suggests that when we find assimilation, we can usually find an explanation based on what we know about how the relevant sounds are produced.

An important question arises at this point, which concerns the role of the phoneme in assimilation processes. Much of the earlier writing on assimilation has suggested that assimilatory changes generally involve a change from one phoneme to another; for example, the example 'I have to' is expressed as showing a change from /v/ to /f/; 'that girl' is supposed to show final /t/ changing to /k/ in /ðæk ɡɜ:l/. Does this mean that all assimilations involve phonemic change of this sort? The answer must be 'no' — we can observe many cases in which there is a clear assimilation that does not involve phonemic change. An easy process to observe is the position of the lips. In a vowel such as English /i:/ (as in 'see'), the lips are spread, as for a smile. In a vowel such as English /ɔ:/ (as in 'saw'), the lips are rounded and pushed forward. This spreading

and rounding of the lips is quite a slow process, and it often happens that preceding and following sounds are also affected by it, even when they belong to a different word. Thus, the /s/ at the end of 'this' will tend to have spread lips in the phrase 'this evening' (where it precedes /i:/) and rounded lips in the phrase 'this autumn' (where it precedes /ɔ:/). The effect is even more noticeable within a word: for example, the two /s/ sounds in 'see-saw', which precede /i:/ and /ɔ:/ respectively, usually have very different lip-shapes. You can easily observe this effect in a mirror. The difference between rounded and non-rounded /s/ is not phonemic in English.

Can we always find an articulatory explanation for assimilation? These explanations seem to assume that we are basically lazy, and do as little work as possible—this is sometimes called the 'principle of least effort', and it does seem to explain a lot of human activity (or lack of it) in a very simple way. A good example is **nasalization**, particularly of vowels, and to understand this process we need to look at the activity of the **soft palate** or velum. When we produce a nasal consonant such as [m] or [n], the soft palate must be lowered to allow air to escape through the nasal cavity; however, for most vowels the velum is raised, preventing the escape of air by this route. In the English sentence 'I know' /aɪ nəʊ/ we would expect that if each segment were produced independently of its neighbours the soft palate would first rise for /aɪ/, then be lowered for /n/, then raised again for /əʊ/. But speech research has shown that the soft palate moves slowly and begins to make its movement some time before the completion of that movement is needed—in other words, we can see *anticipation* in its activity. As a result, the diphthong preceding [n] will be nasalized. We can see a more extreme example in a word like 'morning' /mɔːnɪŋ/ where all the vowels are next to nasal consonants, and the soft palate is often left in the lowered position for the whole word, producing nasalization of each of the vowels. In some languages, the difference between nasalized and non-nasalized vowels is phonemic, but this is not the case in English.

We have seen, then, that the picture of assimilation as a process which causes phonemic change is not adequate. The next point to make is that the simple idea of one sound influencing one neighbour is also unsatisfactory. Let us begin with an example where

there is a regular process of a sound being changed only when it is *both* preceded *and* followed by an appropriate neighbour. In Tokyo Japanese, the vowels /i/ and /u/ regularly change into voiceless segments if they occur between voiceless consonants. Thus in the word 'futon' (the word for a type of bed), the /u/ vowel of the first syllable becomes a voiceless vowel, or simply a short burst of fricative noise, since the /u/ is preceded by the voiceless consonant /f/ and followed by the voiceless consonant /t/.

Coarticulation

The more deeply we look into the complexity of assimilatory processes, the more we need to move away from simple ideas like phoneme change and a single influencing neighbouring segment. This subject is of the most profound importance for understanding how speech is produced. If we want to follow recent experimental and theoretical work in this area that might help us to understand these processes, we must move on to the area of study known as **coarticulation**. In this field, the terms used in assimilation studies that were introduced above (*regressive* and *progressive*) are not usually used; it is more common to encounter the terms *anticipatory* and *perseverative* used (respectively) instead. Alternatively, terms which are easier to use but show a bias towards left-to-right alphabetic systems are *right-to-left* and *left-to-right*, respectively.

The name and the concept of coarticulation have been around since the 1930s, but it remains difficult to explain or define what coarticulation is. We have seen that traditional descriptions of assimilation appear to have concentrated on cases where a change of phoneme results from the assimilation process, or at least on cases where a clearly detectable change takes place which could be represented in phonetic transcription with a different symbol. We find a number of differences from this point of view when we study coarticulation, and I will begin by summarizing these briefly. First, the most important point is that we are interested in *all* aspects of the working together of different articulators, even if the result is difficult or impossible to detect by ear; this is because our primary interest in coarticulation phenomena is in finding a way to explain how the brain and the central nervous system control the muscles which move the articulators (the **neuromuscular**

control of the articulators), rather than in describing the pronunciation of a particular language. Second, it has been demonstrated by many experimental studies that coarticulation has effects which extend much further than just from one segment to another, so coarticulation studies have to assume a more widely spread effect of segments on each other. Third, we take it to be a basic principle that coarticulation is something that can be explained in physical terms, and is not arbitrary. We will now look at these three characteristics of coarticulation in more detail.

When we talk about the brain controlling the production of speech, we make the assumption that there has to be a conversion from an *abstract* form of the utterance we are going to produce to a *physical* form that can be observed and measured. Part of the theoretical study of speech consists of making theories about what the abstract form might be like. In a very simple view, the brain would have the task of deciding what is to be said, and then assembling something resembling a phonemic transcription of it which would be stored somewhere in the brain. We know that the brain has a specific area which has the job of sending commands to the many muscles in the body, including those of the vocal tract, and it is therefore assumed that the instructions to produce each phoneme are passed to this area and converted into signals which cause the articulators to move and produce speech. While the instructions are being executed, various processes cause a partial merging together of the phonemes with the result that assimilation or coarticulation takes place. This view of the process has been likened by one writer to having a conveyor belt carrying eggs (the phonemes) passing between rollers which break the eggs and mix them together. The task of the brain is then to 'unscramble' the eggs and recognize each phoneme so that understanding of the message by the listener is possible. This picture of the speech production process is a very simple one, but while it is probably true to the facts to some extent, it is inadequate in so many ways that it must be drastically modified or rejected. The most important problem is that of time and timing: when we speak, our control of the time taken by each sound we make is very accurate. Yet we know that the task of synchronizing the movements of articulators is very complex: one problem is that the tracts of nerve fibres which carry the commands are of different

lengths and work at different speeds. In the case of a consonant which involves a movement of the tongue tip and at the same time a movement of the vocal folds in the larynx to produce voicing, the impulses will reach the articulators in the mouth some time before they reach the larynx, yet the brain manages to arrange things in such a way that the commands all take effect at the right time. Another problem for the brain to deal with is the different mass of the various articulators: some articulators (for example, the tip of the tongue and the vocal folds) are light and mobile, while others (the tongue body, the soft palate) are relatively heavy and difficult to move quickly. Thus the problem of inertia has an effect on the timing and overlapping found in connected speech.

Coarticulatory effects often extend further than just from one sound to its neighbour. For example, in the word 'screws' /skru:z/, lip-rounding is often found extending over the whole word; it is actually required for the pronunciation of /u:/ and, for most English speakers, for the /r/ too, but it seems that the command to round the lips is sent to the articulators in time for the initial /s/ to be rounded, and this command will remain in effect after the end of the /u:/ so as to produce lip-rounding in the final /z/. This is not just an English characteristic: similar sound-sequences in French have been observed to form in the same way. The French word 'structural' contains two rounded /y/ vowels, and the lip-rounding may, again, be seen on the initial /s/ and through the word up to the beginning of the /a/ vowel. We have already seen how the vowels in the English word 'morning' /mɔ:nɪŋ/ will tend to be nasalized as a result of the lowering of the soft palate for the nasal consonants. All languages appear to exhibit some degree of coarticulatory nasalization of vowels adjacent to nasal consonants.

The third point is that, while studies of assimilation have tended to concentrate on clearly observable aspects of the pronunciation of a particular language, studies of coarticulatory processes are more likely to be looking for effects which are found (not necessarily in exactly the same form) in all languages because they are due to mechanical and biological limits on what the articulators can do in a given amount of time. To return to assimilation for a moment, we can observe in most accents of British English a rather surprising limit on regressive voicing assimilation. In the example 'I have to', given above, it was said that the /v/ is likely to lose any voicing

it might have had, if it is followed by the voiceless consonant /t/. However, it is very unusual to find an English accent which permits regressive assimilation of voicing of the opposite type, that is, a final voiceless consonant becoming voiced as a result of being followed by a voiced consonant. Although this type of assimilation is common in many languages (see for example the French example given above of 'avec vous'), it is not found in English. The phrase 'nice voice' /naɪs vɔɪs/ will therefore not be pronounced with the /s/ changed to /z/, though French learners of English quite commonly do make this change and say /naɪz vɔɪs/, which sounds foreign to English ears. This effect is difficult to explain in terms of coarticulation. If we explain the French change of voiceless to voiced consonant in 'avec vous' as the result of the /k/ being influenced by the vocal fold activity of the neighbouring voiced consonants, how can we account for the fact that this does not happen in the case of English speakers, who are physically the same as the French speakers? The usual answer is said to lie in the difference between the phonetics and the phonology of a language: phonetically speaking, we are all built in much the same way and are subject to the same restrictions on what we can do in producing speech sounds. Phonologically speaking, however, each language has its own private set of rules, which makes it possible for each language to permit or prevent particular coarticulatory processes from taking place. In Spanish, the phonemes /b/, /d/, /g/ are normally pronounced as voiced plosives at the beginning of a word, but as voiced fricatives [β], [ð], [ɣ] between vowels. To Spanish speakers this seems a perfectly natural process which makes the phonemes in question more similar to vowels. In many other languages, this change from plosive to fricative does not happen at all. In others, it can be observed, but it is much less easy to detect. An example from English (BBC accent) would be the phrase 'getting better' /getɪŋ betə/, which in rapid speech is often pronounced with incomplete closure for the /t/ consonants. This results in weak fricatives being produced instead of plosives, though English speakers and listeners usually do not notice this; in the English spoken in southern Ireland, the effect is much more noticeable and is often detected by English people who know nothing about phonetics. There are thus language-specific,

phonological constraints on how much coarticulation, and what type of coarticulation, will be found in a particular language.

Elision

To conclude this account of connected speech and coarticulation, we should look briefly at **elision**. Like assimilation, this is a topic which has had its place in the description of the pronunciation of languages for a very long time. The name refers to the disappearance of one or more sounds in connected speech which would be present in a word pronounced in isolation; the effect is also found when we compare rapid speech with slow, careful speech. If we take as an example the English sentence 'She looked particularly interesting', we could expect the pronunciation in slow, careful speech to be /ʃi lʊkt pətɪkjələli ɪntərəstɪŋ/ (which contains 27 phonemes); in rapid conversational speech, however, I might say /ʃi lʊk pətɪkli ɪntrstɪŋ/ (which contains 20). Where have the seven missing segments gone? The /t/ at the end of 'looked' has been left out because, we may assume, producing three voiceless plosives is hard work, and in English the middle one would not be pronounced audibly in any case. The other elisions are of syllables containing the 'schwa' vowel /ə/, which is so weak that it is usually one of the first items to disappear when speech is produced at higher speed. So the two syllables /jələ/ in 'particularly' are left out, as are the two schwa vowels before and after the /r/ in 'interesting'. As with assimilation, languages differ in which elisions, and how many, they allow, but all languages show some tendency in this direction. From the point of view of coarticulation studies, elision is not a separate process from assimilation. It is simply an extreme result of coarticulation whereby two sounds are articulated so closely in time to each other that a sound or sounds between them are completely obscured. The rapid speech version of 'looked particularly' in the above example is transcribed with no /t/ at the end of 'looked'; however, if we use laboratory instruments to observe what the tongue is doing, we often find that where it sounds as though a /t/ has disappeared, the tongue still makes a partial attempt to articulate a /t/, even though this is impossible to hear. Similarly, in the case of Japanese vowel devoicing, in rapid

speech the vowel sometimes seems to disappear altogether; again, however, if we observe the contact between the tongue and the palate carefully, using laboratory instruments, we can see that the shape of the contact is different according to whether the missing vowel is (or was) /i/ or /u/. It would not be correct, therefore, to say that this is a case of a vowel phoneme being completely lost or deleted; it is more accurate to say that as a result of coarticulation, the neighbouring consonants have occupied all the time available and have overlapped on to the vowel.

There is a lot that we still do not understand about the changes that take place when we change from slow, careful speech to rapid, conversational speech. So much research is being carried out on this subject at the present time, however, that our knowledge is growing rapidly.