Chapter 8

Phonetic Plans for Words and Connected Speech

The first stage of the formulating process, grammatical encoding, is followed by a second stage in which a representation of the utterance's form is generated. It takes successive fragments of surface structure as they become available as input; it produces, incrementally, the form specifications that the Articulator will have to realize, the speaker's phonetic plan. In going from a surface string of lemmas to a phonetic plan for connected speech, the speaker generates a variety of intermediary representations. Phonological encoding is not as simple as retrieving stored phonetic plans for successive words and concatenating them. Rather, the phonetic plan is a rhythmic (re-)syllabification of a string of segments. Each word's segments and basic rhythm are somehow stored in the form lexicon. When these word patterns are concatenated, new patterns arise. Segments may get lost or added, particularly at word boundaries. Syllables may be formed that cross word boundaries. Word accents are shifted to create a more regular rhythm for the larger string as a whole, and so on. Many of these operations serve to create more easily *pronounceable* patterns. And pronounceable is what the input to the Articulator should be. The present chapter reviews some of the form specifications involved in the generation of phonetic plans. How these target representations are built by the speaker will be the subject of chapters 9 and 10.

The review will be done in two steps. Section 8.1 is concerned with the form of words. Words, as we saw in chapter 6, have morphological and phonological structure. This structure is specified in the speaker's form lexicon, or can be composed from smaller lexical elements in the course of speaking. The process by which the speaker retrieves this form information from the lexicon and uses it to create a phonetic plan for the word suggests the existence of a multi-level organization of word-form properties. Section 8.2 deals with connected speech. The phonological encoding of surface structure often requires a phonetic plan that spans several words. Such plans for connected speech have form properties of their own—segmental, rhythmic, and melodic.

Before we proceed to these discussions of lexical and supralexical form, some remarks about the phenomenological status of the generated representations are in order. These representations are, as Linell (1979) suggested, plans for phonetic acts. As such, they have a distinctly articulatory status. And these phonetic plans are, to a substantial degree, consciously accessible. They can, in particular, be present to the speaker as images of the target sound patterns; this is colloquially known as *internal* or *subvocal* speech. It is a remarkable fact that there can be internal speech without overt articulation. The conscious accessibility of phonetic plans makes it, apparently, possible for a speaker to decide freely on whether they should be articulated or not. This differs markedly from the phenomenological status of surface structure. Grammatical relations are not consciously accessible to linguistically untrained speakers, and probably not even to trained linguists when they are in the course of speaking. Surface structure cannot be monitored directly; it is automatically molded into a phonetic plan. It is, however, an empirical issue which aspects of the phonetic plan can be monitored and which cannot. Surely, not every single aspect of the form representation is accessible to the speaker.

Speakers do not always generate sound images (internal speech) for themselves when they speak, and there are probably important individual differences in this respect (McNeill 1987). The accessibility of the phonetic plan is not dependent on the presence of internal speech. In the following, the neutral terms *form representation* and *phonetic plan* will be used to indicate the output of the present component, leaving undecided whether this output is available as internal speech.

8.1 Plans for Words

A word is internally structured at two levels, the morphological and the phonological. At the morphological level a word is composed of one or more *morphemes*, such as roots, prefixes, and suffixes. At the phonological level a word consists of *syllables* and of *segments*, such as consonants and vowels. These two levels of organization entertain complex interrelations. In the following we will first consider some elementary properties of the morphological level of organization. We will then turn to the phonological level, which is organized at different "tiers." We will finish with a few remarks about the interrelationships between morphology and phonology.

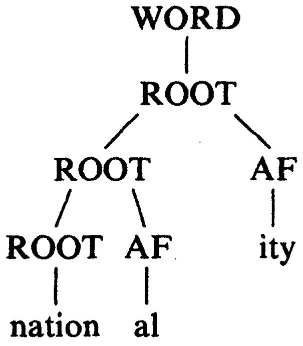
8.1.1 Morphology

Morphological structure is reminiscent of surface structure. A word is a hierarchical organization of meaningful words, roots, and affixes, just as a sentence is a hierarchical organization of meaningful phrases and clauses. The morphological constituents of words combine to make *derivations* (as in *danger-ous*, *loud-ness*, *in-sufficient*), *compounds* (as in *sun-shine*, *dry-clean*), and *inflections* (as in *walk-ed*, *car-s*). In English, affixes are either prefixes (like *in-*) or suffixes (like *-ness*). Other languages, such as Arabic, also have infixes (affixes that are put into a root). In the following we will, however, restrict ourselves to English derivation, compounding, and inflection.

Derivations

Derivations in English can arise through the addition of affixes (prefixes or suffixes) to *words* or to *roots*. Each word that does not carry a word affix is itself a root (e.g., *car*, *sunshine*); hence, many roots are words themselves. But there are also roots that are not words because they cannot stand alone (e.g., *ceive* in *de-ceive*). Let us begin with the affixation of roots.

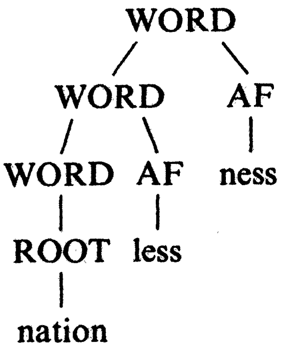
A root can combine with an affix to make a new root. Take the noun root *nation*. It can take the root affix *-al*, which makes it into an adjective root: *national*. This adjective root can, in turn, take an affix *-ity*, which makes it into a noun root: *nationality*. The whole construction, of course, is a word. The tree structure is this:



Other root affixes (called *class I affixes* by Siegel [1974] and *primary affixes* by Kiparsky [1982]) are *-ous*, *-ive*, *-ate*, *-ory*, and *-ify*.

It is characteristic of root affixation that it can affect the sound form of the root: The first vowel in *nation* is different from the first vowel in *national*, and in going from *national* to *nationality* the word accent shifts from the first to the third syllable.

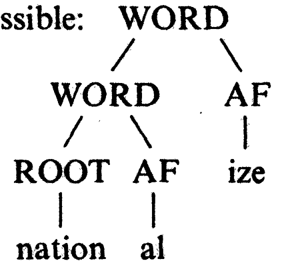
This is different for word affixation, which is far less intrusive on the head word. Take, for instance, the word affixes *-less* and *-ness*. The word *nation* combines with *-less* to make the adjective *nationless*, and the latter word combines with *-ness* to make a noun. As a tree diagram it looks like this:



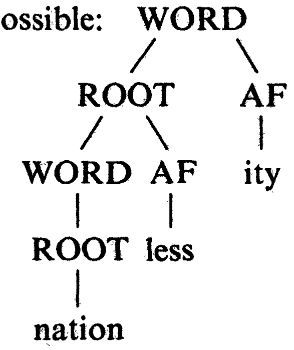
In this case the addition of affixes does not affect the vowel character or the stress pattern of the head word. Other word suffixes (alternatively called Class II or secondary suffixes) are *-er*, *-y*, *-ize*, and *-ish.*

Word and root affixes can combine in a word, but then the word affixes are always outside the root affixes. *Nation-al-ize* is possible, but *nation-less-ity* is impossible. As tree representations:

Possible:



Impossible:

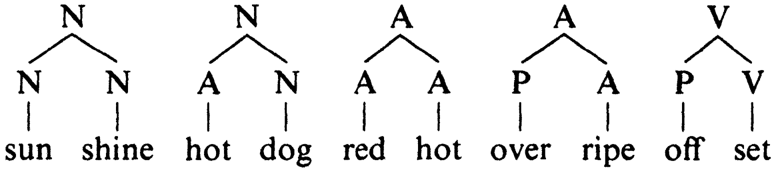


Selkirk (1982, 1984a), whose analysis we are following here, explains this very general phenomenon in English by means of the simple requirement that in a morphological tree a ROOT node may not dominate a WORD node, as is the case in the impossible tree above.

Finally, there is derivational prefixing, and the same regularity appears there. Word prefixes are *ex-* and *non-*; root prefixes are *in-* and *de-*. In mixed cases, word prefixes come to the left of root prefixes (as in *non-de-scending*). There are only a few affixes that can be both word and root affixes, e.g. *un-* and *-ment.* In *governmental*, *-ment* is a root affix, because it is followed by another root affix (*-al*); in *punishment*, it is a word affix because it follows a word affix (*-ish*). For an explanation of these and other exceptions see Selkirk 1982.

Compounds

All English compounds are made up of words from the major lexical categories (nouns, verbs, adjectives, and prepositions), and most combinations of them are possible—for example:



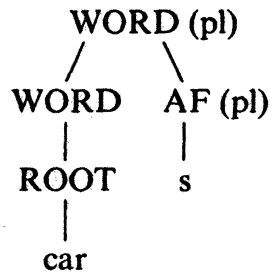
Examples of impossible combinations are NV and VA. Most compounds have a head; it is the rightmost element that has the same syntactic category as the whole compound. In all the examples above, it is the rightmost element (*shine*, *dog*, etc.). The other element usually modifies the meaning of the head, but not all compounds are transparent in this respect. The AN compound *hot dog* is less transparent than the NN compound *sheepdog*.

Also, one can form more complex compounds by combining words that are morphologically complex themselves, or by adding affixes. Examples are *black-board chalk*, *pass-ing note*, and *far-fetch-ed-ness*. The reader may take pleasure in drawing morphological trees for these and other cases.

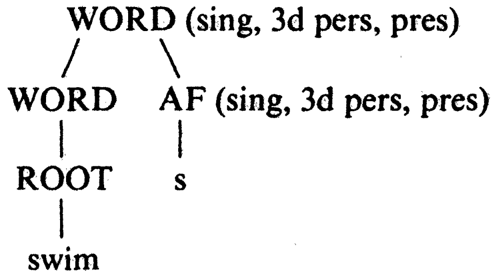
Is compounding the conjoining of roots, or the conjoining of words? Quite probably it is the conjoining of words. If it is the conjoining of words, the resulting compound must be a word as well, because a word node cannot be dominated by a root node. If the compound is itself a word, it cannot take class I or root affixes; it can take only class II or word affixes. That seems to be generally the case. It is, for instance, all right to say *laid-backness*, but not *laid-backity*. It should further be observed that affixed words can be compounded, as in *fighter-bomber*. These, in turn, can be affixed by word affixes (*fighter-bomberless*), and so on. In other words, compounding and class II affixation seem to occur at the same "level." However, in English there may be a higher level of affixation: the level of regular inflection.

Inflections

In English the regular inflections are very simple in structure. They consist of a noun or verb stem plus an affix, and the affix is always a suffix (i.e., it follows the stem word). Take the word form that corresponds to the lemma *car* if it has the diacritic parameter "plural". It has the following morphological structure:



Similarly, the verb form *swims* has this structure:



The three diacritic parameters specified in the lemma for *swim* (in this example, "singular", "3rd person", and "present tense") are together realized in the suffix *-s*. If the number would have been plural, there would have been no suffix.

Halle and Mohanan (1985) argue (contrary to Selkirk 1982) that regular inflection is not on a par with normal word affixation. A major argument is that inflections can be added to all kinds of words, whether they have affixes or not and whether they are compounds or not. But as soon an inflection is added, no further suffixes can be adjoined (as in, say, *swimsness*). Another argument is that it is hard to use an inflected word as the first word in a compound (as in *swimswater*). However, there may be semantic reasons for these facts, and there are, moreover, many exceptions (e.g., *swimming pool*). Whether there is a genuine third morphological stratum in English is a matter of dispute.

The inflectional structure of English is meager. Other languages, such as German, are far richer in inflectional morphology. But German and English have in common that they project different diacritic features of the lemma on the same suffix. The -s in *he thinks* expresses third person, singular, and present tense at the same time. Thus, English is called an *inflectional* language. So-called *agglutinative* languages distribute diacritic features over successive affixes, each one realizing another feature. Turkish, Finnish, Hungarian, and Japanese are languages of this kind. Speakers of agglutinative languages regularly produce new words when they speak, as was discussed in chapter 6. However, an adult speaker of an inflectional language has used almost all the derivational and inflectional forms before, and can readily retrieve them from the mental lexicon. The exceptional new words are mostly compounds. Still, the fact that almost all morphologically complex words are stored in the speaker's lexicon does not mean that the internal structure of a word has become irrelevant in its production. Speech errors and other production phenomena tell us that a word's morphology does play a role in its retrieval during speech.

The entity to which an inflection is added is traditionally called a *stem*. (In the examples above, the roots *car* and *swim* are stems.) So, we now have the notions *word*, *root*, and *stem*. Each word contains (or consists of) at least one root. Many roots can be words themselves; some cannot (such as *-ceive*). If a word is inflected, the stem is whatever the inflectional affixation applies to. Let us now turn to the phonology of words.

8.1.2 Tier-Representation in Word Phonology

A word consists of one or more syllables, and each syllable consists of one or more slots that can contain phonetic material, such as consonants or vowels. The final phonetic plan for a word represents how the phonetic material is distributed over these so-called *timing slots*. The axis of this representation is nothing but the row of timing slots. It is called the *skeletal tier* (or, alternatively, the *timing tier*). But these slots are partitioned in small chains. There is, first of all, their grouping into successive syllables. And syllables can consist of further constituents, such as onsets and rimes. This constituent organization of the slots at the skeletal tier is represented at the so-called *syllable tier*.

The phonetic content of the slots is of two kinds: *quality* (also, confusingly, called *melody* in the phonological literature) and *prosody*. The quality of phonetic content is determined by such features as whether the speech sound should be voiced or unvoiced, nasal or non-nasal, plosive or non-plosive. These qualitative features are specified at the *segment tier*. The prosodic plan, on the other hand, specifies the metrical and the intonational pattern of the intended utterance. They are specified at the *metrical* and *intonational* tiers, respectively.

It is not precisely known how the different tiers relate to one another (see Clements 1985 for various hypotheses), but figure 8.1 can be a starting point for the following discussion. The tier organization is like a set of pages glued together at the edges. Each page edge represents a tier, and each page connects two tiers. The skeletal tier connects via the right page to the segmental tier and via the left page to the syllable tier. The syllable tier, in turn, connects directly to the metrical and to the intonational tier. Later I will suggest that the segment tier at the right is the spine for a further set

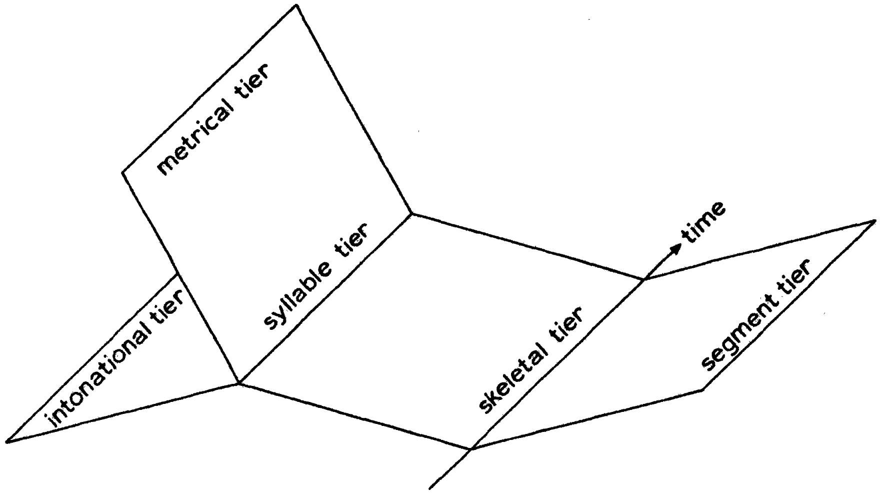


Figure 8.1
An initial representation of the coupling of tiers in an articulatory plan.

of pages, containing the *feature tiers*. Each page connects two tiers, and it displays the "association lines" between the tiers, i.e., the precise connections between the units of the two tiers involved. The following subsections will give a summary discussion of the tiers and their interconnections.

8.1.3 The Skeletal Tier

The skeletal tier is a sequence of slots, and these slots can contain phonetic content. A filled slot is a phonetic segment or a *phone*. It is not the case that a speaker pronounces one phone, then the next phone, and so on. Still, his phonetic plan consists of a sequence of phones. Each phone, however, will be realized by the articulator as constituting part of a larger articulatory gesture. The smallest such gesture is the articulation of a whole syllable.

The simplest representation of the skeletal tier is just

X X X X . . . ,

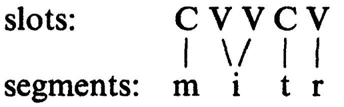
the sequence of timing slots to be filled. Indeed, some phonologists leave it at that (see, e.g., Halle and Mohanan 1985). Others, however, distinguish between slots that should be filled with *sonorous* content and slots that should be filled with *nonsonorous* materials. Sonorous speech sounds are relatively loud. The most sonorous ones are vowels, such as [a] and [ɔ]. Least sonorous are voiceless stops, such as [p] and [k]. Others are in between; [r], [l], [m], and [n], for instance, are in the middle range of sonority (Selkirk 1984b). A syllable always has a high-sonorous peak segment, which can be preceded or followed by other less sonorous sounds. Many phonologists indicate at the skeletal tier which slots should contain syllabic peaks; i.e., "syllabicity" is represented at the skeletal tier (Hayes 1986). Such slots are indicated by V. This, of course, derives from "vowel," but it means only "high-sonorous segment of the syllable." All other slots are indicated by C. This derives from "consonant," but it means only "low-sonorous segment of the syllable."

The syllable peak may extend over two slots, as in long vowels, but it cannot have two peaks separated by a less sonorous element. In fact, there is a general rule that, within a syllable, sonority decreases from the peak to the syllable boundaries (Selkirk's [1984b] "Sonority Sequencing Generalization"). The peak is the top of a "sonority hill," which slopes down toward the boundaries without secondary sonority peaks. One must, probably, allow for a slightly weaker version of this rule: that sonority is *nonincreasing* away from the peak. There may be juxtaposed segments of *equal* sonority.

Here we will adopt the CV notation for the skeletal tier (Clements and Keyser 1983). A sequence of slots could, for instance, look like

C V V C V.

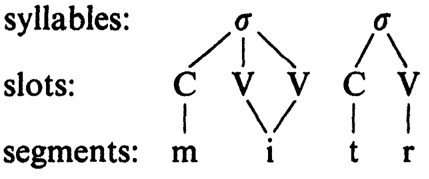
This represents a sequence of two syllables, because there are two peaks: a long one and a short one. This particular tier could be filled by materials for the word *meter*. The first syllable has [i] as peak, a long vowel that takes two slots. It is preceded by the less sonorous [m], which fills the initial C slot. The second syllable has [r] as its peak; it is more sonorous than the [t] that precedes it. The segments, therefore, relate to the slots as follows.



In the following, phones or strings of phones will be put in square brackets. The phonemes of which phones in the phonetic plan are a realization are given between slashes. So, [m] is the phone in the phonetic plan that, in some syllabic environment, realizes the phoneme /m/.

8.1.4 The Syllable Tier

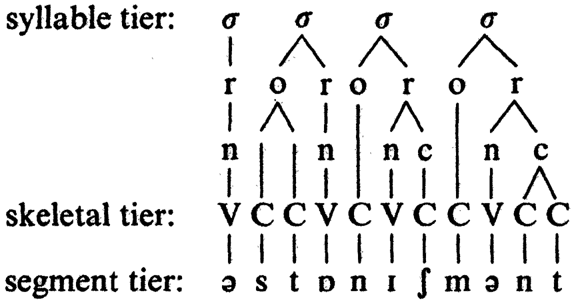
The slots at the skeletal tier are grouped in larger *syllable units*. These are represented at the syllable tier. The simplest representation would be one in which the slots are merely connected to successive syllable nodes. The syllabic representation for *meter* would then be the following.



What kinds of strings can be syllables? All languages have CV or CVV syllables, most have V syllables, many have CVC syllables, and some have VC syllables. English has them all. It is, moreover, possible in English to make strings of two Vs, and of two or three Cs within a syllable (as in the monosyllabic word *scraped*).

Such strings of V and C slots within a syllable can be further partitioned into so-called *syllable constituents*. A syllable's main constituents are *onset* and *rime*. Each syllable has a rime. The rime begins with the syllable peak. If the syllable is just a V, that V is the rime. If further consonants follow the peak, they also belong to the rime. So *art* is a word, it is a syllable, and it is a rime. A rime is naturally partitioned into a *nucleus* and a *coda*. The nucleus contains the peak slot(s), the coda the remaining C slot(s). The syllable *art* has /a/ as nucleus and /rt/ as coda.

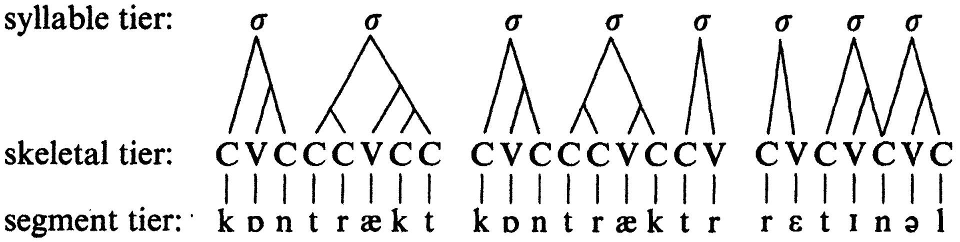
The onset of a syllable is the string of Cs preceding the peak. In *meter*; /m/ and /t/ are syllable onsets. Onsets can also be clusters of low-sonorous elements, such as /skr/ in the monosyllabic word *script*. If no C precedes the V, the onset is empty. Some of the main syllable types appear in the four-syllable word *astonishment*:


syllable tier: σ σ σ σ
r o r o r o r
n n n c n c
skeletal tier: V C C V C V C C V C C
segment tier: ə s t ɒ n ɪ ʃ m ə n t

The first syllable, /ə/, fills a single slot. Its peak is, at the same time, its nucleus, its rime, and the whole syllable. The second syllable, /stɒ/, has an onset, /st/, and it is further branching over two slots containing /s/ and /t/. The rime is nonbranching; it consists of the nucleus /ɒ/. The next syllable, /nɪʃ/, has a nonbranching onset, /n/, but a branching rime, consisting of a nucleus /ɪ/ and a coda /ʃ/. A coda can also be branching. This is the case in the fourth syllable, /mənt/. Here the onset /m/ and the nucleus /ə/ are nonbranching, but the coda fills two C slots: with /n/ and /t/, respectively, where /n/ is more sonorous than /t/. The way in which a syllable branches is, as we will shortly see, a determinant of whether it will receive stress at the metrical tier.

In English multisyllabic words, each V is part of the peak of a different syllable. Compare the words in example 1: *contract*, *contractor*, and *retinal*.

(1)



Contract has two V peaks and hence two syllables; *contractor* and *nominal* have three of each. Since V here is the peak of a syllable, how are the consonants distributed over adjacent syllables? The main rule is that a V attracts as many of the consonants preceding it as it can. So, instead of *contr-act* or *cont-ract*, we have *con-tract*. The vowel of the second syllable takes on both /t/ and /r/. This rule is known as *maximization of onset*. Only the "leftovers" are for the coda of the preceding syllable. In *contract*, /n/ is the leftover C; it is subsumed under the previous syllable (*con*). Why is /n/ not also attracted into the initial consonant cluster of the second syllable? A major reason is that the phone [n] (the realization of /n/) is more sonorous than [t] (the realization of /t/). As a consequence, the resulting syllable, [ntrækt], would have two sonority peaks: [n] and [a]. This contradicts the Sonority Sequencing Generalization. There may, in addition, be language-specific restrictions on clusters. For each language, there are so-called *phonotactic rules* that specify possible and impossible clusters in syllable onset and syllable offset.

The second syllable of *contract* is *tract*. But the second syllable of *contractor* is *trac*. This is, again, due to maximization of onset. The third syllable of *contractor* doesn't like to begin with V; it tries to assemble syllable-onset consonants. It takes /t/; however, it cannot take /k/, since /kt/ is phonotactically impossible as a syllable onset in English (though not in other languages). The result is thus *con-trac-tor*. This example shows how the addition of a suffix can lead to *resyllabification*, the reassignment of elements at syllable boundaries.

Though the maximization of onset applies quite generally, there are also exceptions. In *nationlessness* there is a syllable break between *less* and *ness*; still, /sn/ is a possible syllable-initial cluster in English. In subsection 8.1.1 it was discussed that word affixes are not "sound-intrusive." Here we see that they also tend to preserve their own syllable identity. This is different for root affixes that easily attract consonants from the root (as in *plaint* → *plain-tive*).

The word *retinal* in example 1 demonstrates still another important feature of syllabification. Maximization of onset would predict the syllabification *re-ti-nal.* Still, the consonant /n/, which is syllable-initial, also participates in the foregoing syllable, so that the word contains as syllables *re*, *tin*, and *nal*. This phenomenon is called *ambisyllabicity*. Certain consonants in intervocalic position (i.e., in VCV configurations, where both Vs are unstressed) can make a *liaison* between adjacent syllables. For American English, the flapping /t/ of *retinal* is probably also ambisyllabic.

8.1.5 The Segment Tier

A speaker's phonetic plan represents which phones go in successive timing slots. The sequence of phones in a syllable specifies the articulatory gesture to be made by the speaker in order to realize that syllable. The number of different phones in the world's languages is fairly limited. Whatever its exact size, it is small enough that an International Phonetic Alphabet (IPA) could be designed by which the phonetic segments in different languages can be roughly transcribed. The IPA, given in appendix A, is used in this book to represent phonetic segments. The phone sequence *cat* will be transcribed as [kæt], *nation* as [neɪʃən], and so on.

Phones are not indivisible wholes. They represent various features of the articulatory gesture to be made. To utter a speech sound is a fairly complicated motor activity involving maneuvers of different parts of the respiratory system, the larynx, and the vocal tract. The sound may require inhalation or exhalation, it may or may not require voicing (the periodic vibration of the vocal folds), and it will require a range of vocal-tract specifications, such as positions or movements of the velum, the tongue, and the lips. A phone is an abstract representation of such gestural components. Their precise execution is not specified in the phonetic plan; that is the task of the articulatory component.

Which features, then, are specified at the segmental level? I will not give an exhaustive listing, but I will mention some major classes (see also section 11.1.3). At the respiratory level, a speaker might make a distinction between inhaling and exhaling sounds. But English and most (but not all) other languages have exhaling speech sounds only. At the laryngeal level the most important feature is *voicing*. All spoken languages make a contrast between voiced and unvoiced speech sounds. In English, [d] and [t] differ in just this feature, and so do the segments [b] and [p] and the segments [g] and [k]. At the supralaryngeal level (i.e., the level of the vocal tract), two main classes of features can be distinguished: *place* features and *manner* features.

The place features indicate where in the vocal tract the speech sound is to be made. Some examples of place features are *labial* (versus *nonlabial*), which specifies whether the lips should form a constriction, as in [m], [p], or [b]; *coronal*, which requires the constriction to be made by the tongue blade, as in [t] or [O]; and *posterior* (versus *anterior*), where the primary constriction is behind the alveolar ridge, as in [g] or [h].

The manner features specify a variety of further aspects of the articulation beyond its place. Among them are *nasal* (versus *non-nasal* or *oral*), which specifies whether the velum should be lowered so that the nasal cavity will resonate with the speech sound, as in [m] and [ŋ]; *rounded* (versus *unrounded*), meaning that the sound is to be made with protruding lips, as in [u]; and *strident* or *fricative*, which involves the generation of a spirantal noise at the place of constriction, as in [s] and [f]. The above-mentioned feature of voicing is also often considered to be a manner feature. Several of these manners and places of articulation will be further discussed in chapter 11.

A feature may or may not be *distinctive* within a language. It is distinctive if the distinction between two words in a language hinges on that feature. Voicing is distinctive in English, because it opposes words such as *bill* and *pill*, *tell* and *dell*. The manner feature *nasal* is distinctive because it discriminates the words *man* and *ban*, and *name* and *dame*, and so on. Other features are not distinctive. For English this is so for the *aspiration* feature. Compare the words *pot* and *spot*. The [p] of *pot* is pronounced such that a little puff of air follows the release; this is called aspiration. The word *spot* is pronounced without p-aspiration. But the aspiration feature is nondistinctive in English, because there are no two words whose difference hinges on that feature. There are, for instance, no two different words *p'ot* and *pot* in English, where the former is aspirated and the latter is not and where different things are meant. Whether or not /p/ should be aspirated depends only on its syllabic enviroment. They are two context-dependent variants of /p/. Variants of a speech sound differing only in nondistinctive features are called *allophones*; [p'] and [p] are allophones in English. They are allophonic realizations of the same *phoneme*, /p/. A phoneme is a segment as far as specified for its distinctive features only. So, *pot* and *spot* contain the same phoneme /p/, but realized as different (allo)phones.

It is an important question whether the more abstract notion of "phoneme" is psychologically relevant for a theory of the speaker, or whether the phone is the only relevant level of form representation. The next chapter will discuss evidence from speech errors that strongly supports the notion of the phoneme as a processing unit of phonetic planning. When the speaker intends to say *spot-and-kill* but instead happens to say *skot-and-pill*, the *phonemes* /p/ and /k/ exchange, not the allophones; the /p/ will be aspirated in *pill*, whereas it was not in *spot*. By and large it seems to be the case that the form specifications for words in the mental lexicon are *phonemic*, like /spɒt/ and /pɪl/. The speaker uses these representations to retrieve the corresponding syllabic gestures. The phonetic plans for these syllabic gestures are phonetic, i.e., in terms of allophones (e.g., [spɒt] and [pʻɪl]. The next chapter will discuss how the phonemic codes in the lexicon are used to retrieve the phonetic specifications for the syllables.

The number of distinctive features for a language is quite small, probably around 20. Each of them is a more or less independently controllable aspect of articulation. It was remarked above that a speaker does not pronounce one phone, and then the next one, and so on. Rather, he plans whole articulatory gestures of at least syllabic size. Such a gesture normally involves not only feature change but also feature maintenance. Successive syllabic slots may maintain the values of certain specific features while changing the values of others. In the word *manner*, for instance, the distinctive feature of voicing persists over the whole word. Or take nasality. The French pronunciation of the town name *Nancy* involves a sequence of [n] and [ã], where both [n] and [ã] are nasal; nasality persists over two phones, the whole first syllable of the word. It has therefore been suggested that the segment tier be used as a spine to which a whole set of *feature tiers* are linked, like the pages of a book. At least, one could make a laryngeal tier and a supralaryngeal tier, and subdivide the latter into a place tier and a manner tier. This was suggested by Clements (1985). Alternatively, one could write a separate tier or line for each distinctive feature, as is done for the voices in a musical score. We will not pursue this here. The important point is that we avoid the trap of a strictly segmental picture of articulatory planning. Articulatory gestures can span several timing units, and should be represented as such.

8.1.6 The Metrical Tier

Speaking is a rhythmic process. A speaker organizes his utterance in patterns of stressed and unstressed syllables, and can assign various degrees of stress or accent to different syllables. The production of speech is, in this respect, not unlike the production of music. Rhythmical organization is probably the most universal property of music. There is always some regular organization of beats in groups of two, three, or more, and the metrical organization is almost always hierarchical to some degree (Longuet-Higgins 1987). There is, of course, no fixed time unit in speech like a measure in music, but there is surely some hierarchical organization of beats.

At the word level of speech, multisyllabic words can also be said to have a metrical pattern, and this pattern is represented on the metrical tier. The metrical tier is associated to the syllable tier. Stress is a sound property of syllables. It is, among other things, reflected in the duration of the spoken syllable. When a syllable is stressed, it is longer than when it is not stressed. The syllable *for* is longer in *formal* than in *forlorn*. From this point of view, the metrical tier might be represented as a string of syllable-linked musical notes of different lengths, like this:

for mal
for lorn

This, however, would be too suggestive. Syllable duration is not the only way in which stress variations are realized. Stress is also, in part, realized by variation in amplitude or intensity and by pitch movement. It needs a more abstract representation than a purely temporal one. There is, moreover, a relation to vowel quality. The initial phoneme /ɔ/ is realized differently in *forlorn* and in *formal*. When a vowel is unstressed in English, its quality may differ from the stressed allophone. In short, stress is an abstract category, like "phoneme". It may be realized in different ways, depending on the language and on the speaker.

An attractive way to represent a word's stress pattern on the metrical tier is by means of a so-called *metrical grid* (Prince 1983; Selkirk 1984a). It provides each syllable with one or more beats. Take the word *California*. It has a beat pattern like this:

```
       x
x      x
x  x   x   x
Ca li for nia
```

The main word stress or *word accent* is on the third syllable (*for*); there is a secondary stress on the first syllable (*ca*), and the remaining two syllables have the minimum amount of stress. We assume that, in most cases, a word's basic stress pattern is simply available to a speaker. But it is surely interesting that, when confronted with an unknown printed word, a native speaker will immediately assign a stress pattern that conforms to the rules of the language. For instance, *pefinarca* will most likely receive the same stress pattern as *California*. Speakers apply stress-assignment rules unawares; the rules are part of their knowledge of the language.

This is not the place to review the metrical rules of English (see especially Selkirk 1984a), but some major tendencies in stress assignment can be mentioned.

A first, quite general observation is that the rime of a syllable is important in stress attraction, whereas the onset plays no role in it. When a syllable has a branching rime—i.e., when the rime is anything more than a single V, like *for* in *California*—it is called a *heavy* syllable, and it will get an extra beat in its column. Syllables with just a V rime are called *light*; they do not get an extra beat for heaviness.

There is, second, a strong tendency toward stress *alternation*. Speakers dislike a sequence of two stressed syllables. If this is imminent, for instance because there is a sequence of two heavy syllables, one of them will lose a beat. On the other hand, speakers also dislike sequences of unstressed syllables, especially more than two of them. A speaker will add beats in order to break such patterns. The resulting alternating stress pattern is present in *California*, but also in most other polysyllabic words.

Third, affixes play a special role in stress assignment. Word-initial and word-final affixes are never stressed in English. Affixes can receive stress only in nonextreme positions, and only when they are root affixes. This is, for instance, the case for *al* in *nationality*.

Fourth, function words (auxiliaries, pronouns, determiners, and so on) tend to be destressed as if they were affixes. Phonologically, they are not really words at all.

Fifth, when a lemma has been assigned pitch accent in the generation of surface structure, at least one extra beat is given to the syllable that has word accent. For example, if a speaker answers the question *Were you born in Oregon?* with *No, in California*, the metrical grid for *California* will be

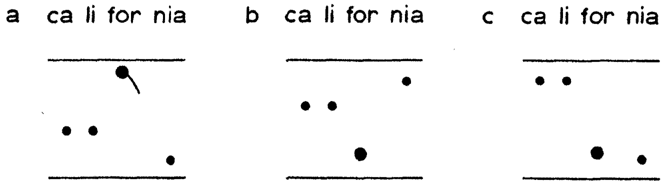
```
      x
      x
x     x
x  x  x   x
Ca li for nia
```

So, even if the stress pattern of *California* is stored in the speaker's word form lexicon, it can be adapted to accommodate the diacritical feature of pitch accent. Other accomodations of a word's basic metrical pattern can result from stress clashes with neighboring words (see subsections 8.2.2 and 10.2.1).

8.1.7 The Intonation Tier

So far we have seen that the phonetic plan an English speaker constructs for a word derives largely from information stored in the mental lexicon. The phonemic constitution of morphemes and words is stored in the lexicon; so, quite probably, is the (allo)phonic composition of the language's syllables. To generate the phonetic plan for a word, the speaker uses the phonemic code as a key to retrieve the phonetic syllabic code. (This process is the subject of chapter 9.) But some aspects of the word's phonetic form are not stored in the lexicon and have to be constructed time and again. For English and many other languages, this is true of a word's intonation pattern. (This pattern should be carefully distinguished from the word's stress pattern; the same stress pattern can go with very different intonation patterns.) Here are some possible intonation contours for *California* (for the notation system, see subsection 8.2.3):

(2)



It is not hard to think up contexts in which these different contours would occur naturally. Version a can occur in the simple assertive utterance *I live in California*; version b can be used in a question, such as in *Do you live in California?*; and version c can occur in the emphatic *California again!*. The three intonational contours perform three different functions. But they are not lexically stored as three different versions of *California*; rather, they derive from the intonational phrase in which the word participates.

In all three phrases, *California* has pitch accent, and this is in all contours realized by an obstruction of pitch, either up or down, to the syllable that carries the word accent (i.e., *for*). The intonational shape of a word is largely determined by whether it has pitch accent or not and by where it occurs in the intonational phrase. More will be said on this in subsection 8.2.3.

8.1.8 Morphophonemic Relations

So far we have dealt with the morphological structure of words, and with their phonological structure. Clearly, the two are intimately related. The phonetic material that appears in the syllabic slots of the phonetic plan originates from the phonemic properties of a word's morphemes. But syllable boundaries do not always respect morpheme boundaries, and a lot of rearrangement takes place in going from the morpheme level to the syllable level. En route, phonemes may be added or deleted, and feature values may change. All this is governed by *morphophonemic rules.* These rules will not be discussed here in any detail; we will just consider the different regular phonetic shapes the plural morpheme can take in English:

pit → pits	(addition of /-s/)
toad → toads	(addition of /-z/)
edge → edges	(addition of /-ɪz/)

Here we see the addition of different segments, and resyllabification (in the case of /-ɪz/). Such variants of the same morpheme are called *allomorphs.* The morphophonemic relations between singular and plural are known to native speakers of English, who easily apply them to nonsense words: *rit → rits, tood → toods, losh → loshes.* This means that the morphophonemic relation between singular and plural is not only stored for each noun in the speaker's lexicon, but is also abstractly stored as a set of rules that can be productively applied to new words. These rules are rather simple; they refer only to the final segment of the noun (is it voiced, coronal, strident?), and they apply in just the same way to possessives (*the pit's taste, the toad's legs, the edge's sharpness*).

This concludes our review of the speaker's phonetic plan for single words. But the final phonetic shape of a word also depends on the context in which it appears. Words assimilate to their environment in connected speech. This may affect their segmental composition, but also their metrical properties. Moreover, words participate in the melody of the utterance.

8.2 Plans for Connected Speech

The phonetic plan for connected speech can also be represented as a layered structure. In fact, the tiers are no different than the word-level tiers. On the skeletal tier there is a sequence of high and low sonorant slots; the phones in these slots are represented at the segment tier. Their partitioning into syllables and syllable constituents is represented at the syllable tier. The stress pattern over syllables is represented at the metrical tier, and the utterance melody appears at the intonation tier. Here we will be concerned only with those aspects of the plan that make it different from a mere concatenation of word plans. There are, in fact, differences at the level of segments and syllables, at the metrical level, and at the intonation level.

8.2.1 Segments and Syllables

The phonetic plan for connected speech may involve a special choice of allomorphs, cliticization, and resyllabification. Let us consider some examples of these phenomena.

A special choice of allomorphs occurs in English *auxiliary reduction*. Auxiliaries like *have* and *is* have reduced forms, which are chosen in certain connected speech situations. Rather than saying *I have bought it*, speakers will opt for *I've bought it*. Similarly, they will say *Dick's running* instead of *Dick is running*. In these cases the reduced allomorph, /-v/ or /-z/, cannot stand alone as a word, because it is not syllabic. Rather, a new word-like unit is formed by gluing the allomorph to the preceding word to form *I've* and *Dick's*. This is called *cliticization*; /-v/ and /-z/ are *enclitics* to the preceding word. The new string is called a *phonological word*. Little is known about the lexical status of such cliticizations. Is *I've* a stored lexical element in the speaker's mental lexicon? It probably is, because of its frequent usage. But it is less likely that *Dick's* is a readily stored item with the meaning of *Dick is* (except if the speaker happens to be married to Dick and is always talking about him). At any rate, a succession of two different lexical items in surface structure may be realized as a single word-like unit in the phonetic plan for connected speech (see Kaisse 1985 and Nespor and Vogel 1986 for detailed analyses of these matters).

When words are juxtaposed in connected speech, they may be subject to resyllabification at their junctures. In fact, segments may appear which are absent when the word is spoken in isolation. Let us take an example from British English: /r/ deletion. The /r/ is not pronounced in syllable-final position, so *car* is pronounced as [ka], and *care* as [kɛ:]. But it reappears in the word when in syllable-initial position, as in *carry* or in *caring*: [kæ-ri], [kɛy-rɪŋ]. This is a word-internal phenomenon, and one wonders whether it also occurs across words in connected speech. It does. When the British English speaker says *the car is running* without cliticizing the *is*, the sequence *car is* becomes resyllabified as [kɑ-rɪz]; the /r/ of *car* is now syllable-initial and thus becomes pronounced. Similar phenomena occur in French *liaison*, as in *nous-avons*, where the /s/ reappears in syllable-initial position.

The segmental and syllabic structure of a phonetic plan for connected speech is not a mere concatenation of plans for isolated words. There can be considerable restructuring, in particular by the use of allomorphs, by cliticization, and by resyllabification. These and other restructurings are especially apparent in fast speech, to which we will return in chapter 10.

8.2.2 Metrical Structure

Words and word-like units are grouped into smaller or larger prosodic units. The main such unit in English and many other languages is the *intonational phrase.* As the term indicates, it is a unit of intonation, and as such it will be discussed in the next subsection. Take example 3.

(3) //The detective /1 remembered //2 that the station /3 could be entered /4 from the other side as well //5

The speaker of this sentence could deliver it as a sequence of two intonational phrases, *The detective remembered* and *that the station could be entered from the other side as well*—i.e., the two finite (and basic) clauses of the sentence. This partitioning is indicated by double slashes (//). But intonational phrases also have an internal metrical structure. Many authors assume the existence of so-called *phonological phrases* as metrical building blocks of intonational phrases (see especially Nespor and Vogel 1986). According to this view, each intonational phrase consists of one or more phonological phrases. These are metrical units which relate to surface structure in the following way (for English, and leaving some qualifying details aside): The first phonological phrase of a sentence begins where the sentence begins and ends right after the first lexical head of an NP, a VP, or an AP. The next phonological phrase begins just there and ends after the next such lexical head, and so recursively; any remaining "tail" after the last lexical head is added to the last phonological phrase. According to this definition, sentence 3 consists of five phonological phrases, with right boundaries at /1, //2, /3, /4, and //5. In this sentence the respective lexical heads-of-phrase are *detective*, *remembered*, *station*, *entered*, and *side*.

The phonological phrase is characterized by a metrical togetherness of adjacent words. Bock (1982), Garrett (1982), and van Wijk (1987) suggest that it is indeed a unit of phonological encoding, an output package for the articulatory component. However, others (in particular, Selkirk [1984a]) do not partition intonational phrases into such all-or-none metrical building blocks. Rather, they assume various *degrees* of metrical togetherness between adjacent words in intonational phrases. There are, on this view, no strict rules for partitioning an intonational phrase into smaller phonological packages, except that the speaker will respect a certain degree of togetherness. In other words, the speaker has certain options that he may or may not use to complete a "package" for the Articulator. Whether he will use an available option depends on various performance factors (to which we will return in chapter 10). The speaker of sentence 3, for instance, could, after planning the word group *the detective*, ignore the option for a break and add *remembered* in the same phonological phrase, thus creating *the detective remembered* as one output package.

There are, then two issues to be considered: Where are the options for a speaker to complete a phonological phrase? What makes the speaker use an available option? The latter issue will be taken up in chapter 10; here we will only consider the former.

There are better and worse options. A very good place to complete a phonological phrase is the end of a sentence (position //5 in the example above), or the end of a clause (positions //2 and //5). Ends of clauses are quite often also ends of intonational phrases. Not a good option is right after a preposition in a prepositional phrase, e.g., after *from* in *from the other side.* Several factors conspire to make a good option. Let us review some of the potential places for a break.

(i) The end of an intonational phrase. This break is obligatory.

(ii) The end of a sentence constituent. A sentence constituent is one that, in the destination hierarchy (see subsection 7.1.3), is delivered to S. This is, normally, the case for the subject and the predicate phrase of a sentence.

(iii) The end of a multi-word phrase. Ends of NPs, VPs, APs, or PPs are good options for a prosodic break.

(iv) After the lexical head of a NP, a VP, or an AP (i.e., the criterion used above to define phonological phrases). A good break point is after the main verb of the verb phrase, or after the main noun in a noun phrase, even if these are not in constituent-final position.

(v) After a content word. It is almost never the case, except in self-repairs, that a speaker breaks *within* a word. And if a speaker breaks after a word, that word is usually a content word. Function words that are not phrase-final are seldom followed by a break. The head of a prepositional phrase, for instance, is usually not followed by a break—except when the speaker has trouble accessing the following head noun.

Selkirk (1984a) represents these options by adding "silent demibeats" to the metrical grid for each of the above factors (I will use the term "silent beats" instead.) Here is one of her examples:

```
   x        x                  x          x
   x        x            x     x          x
   x x  xxx x x      xx  x     x   x   x  x  x xxxxx
Mary       finished     her  Russian    novel
```

The first word, *Mary*, is followed by three silent beats: one because *Mary* is a content word (v), one because it is head of phrase (iv), and one because it completes a sentence constituent (ii). The next word, *finished*, is followed by two silent beats, because it is a content word (v) and because it is head of phrase (iv). The possessive pronoun *her* is not a content word (see v), nor does it meet any other condition for the addition of a silent beat. And so forth. Such a metrical-grid representation should, of course, not be read as a direct representation of pause durations between words. Rather, it represents degrees of rhythmical togetherness of words; *her* and *Russian* are metrically more together than *finished* and *her*, for instance. The best break options are at points of least metrical togetherness. In the present example, that is between *Mary* and *finished*, and between *finished* and *her*. On a phonological phrase analysis, these are precisely the boundaries between the three phonological phrases of this sentence: *Mary*, *finished*, and *her Russian novel*.

Even if the speaker does not take every major break option, he may highlight the phrasal structure of his utterance by other metrical means. He may, for instance, slightly stretch a syllable or insert a small pause where there is a metrical caesura. He can, moreover, add prominence to the last already-prominent word in the phrase. This is called the *nuclear stress rule* (Chomsky and Halle 1968; Selkirk 1984a; subsection 5.2.2 above). Take again the example *Mary finished her Russian novel*. The noun phrase *her Russian novel* has *novel* as its lexical head, its last prominent word. Nuclear stress requires it to be more prominent than all other words in its phrase. Hence, it is given an additional beat. Similarly, the verb phrase *finished her Russian novel* also has *novel* as its last prominent word. With its additional beat, *novel* is also the most prominent word in the latter phrase, in accordance with the nuclear stress rule. Finally, *novel* is also the last word of the whole sentence. Hence, it should also be the most prominent word of the sentence. With its additional beat, it is. The resulting metrical grid looks like this:

```
                               x
  x       x            x       x
  x       x        x   x       x
  x x xxx x x   xx x   x  x  x x x xxxx
Mary     finished her Russian novel
```

Ideally speaking, then, the surface structure of sentences can be highlighted by various metrical means. Whether and to what degree that is actually done in fluent speech is a different matter. What should be kept in mind is that pitch accent will overrule everything else. If the speaker has reason to give pitch accent to *her* (for instance, to make a contrast with *his Russian* *novel*), then *her* will be given greatest prominence and *novel* will be almost stressless.

A final readjustment to be mentioned is one that promotes an optimal alternation of high and low stresses. The speaker will try to avoid stress clashes within a metrical group or a phonological phrase. Take for instance the adjective *abstrAct*, which has its word accent on the second syllable. In the phrase *Abstract Art*, the word accent will appear on the first syllable. Selkirk (1984a), who gives a systematic treatment of these shift phenomena, calls this "beat movement." The beat moves to the penultimate syllable of *abstract* in order to avoid a stress clash with the accent in *art*. The tendency toward alternating stress, which we already observed at the word level, apparently also holds within coherent metrical groups of words. Speakers of English dislike sequences of stressed syllables within a phonological phrase.

This section on the speaker's metrical plan would not be complete without addressing the issue of speaking *rate*. Deese (1984) found a speaking rate of 5–6 syllables per second in normal conversational speech, but speakers can accelerate substantially. There are, now and then, short stretches of high-rate speech with about 8 syllables per second. Deese was able to recognize at least a few functions served by such increases of rate. One had to do with turn-taking. Speakers speed up toward the end of a sentence and into the next one in order to keep the floor. This bridging is done to prevent an interlocutor from taking the next turn at the end of the sentence. Another function is expressive: to say something in a modest, nonassertive way. This probably happened in the following utterance: *There's a very recent paper. I'm trying to think of the author of it.* The second sentence was spoken very rapidly and with flattened intonation. Clearly, such rate parameters must be set in the speaker's phonetic plan.

Though the rhythm of connected speech builds on the metrical properties of the individual words, it has additional features of its own. There is, first, speaking rate, which may serve interactional and expressive functions. There is, second, a grouping of words in short stretches leading up to the lexical heads-of-phrase. Within these small metrical groups there will be a tendency toward alternating stress. These metrical units or phonological phrases may, third, combine to form larger patterns, whose metrical properties can (to some degree) highlight surface-structure relations, especially through nuclear stress assignment. Pitch-accent peaks are quite marked in the metrical pattern of connected speech. The larger coherent metrical patterns are usually called intonational phrases, because they constitute the domain for assigning melody to an utterance.

8.2.3 Intonation

The most characteristic form property of connected speech in intonational languages is its melody. The English speaker's lexicon, we saw, contains no intonation patterns for words, only stress patterns. The intonation of a word depends on which syllable is lexically marked for word accent, on whether the word is focused in surface structure (i.e., whether it should receive pitch accent), and on the sentence melody in which it partakes.

Intonation is, in the very first place, an expressive device. Pitch accent expresses the prominence of a concept, the interest adduced to it by the speaker, or its contrastive role. The melody of an utterance expresses a speaker's emotions and attitudes. It is a main device for transmitting the rhetorical force of an utterance, its weight, its obnoxiousness, its intended friendliness or hostility. It also signals the speaker's intention to continue or to halt, or to give the floor to an interlocutor. The expressive functions of language, though intimately tuned to its referential and predicative functions, are probably rather independently controlled.

When we discuss the speaker's intonational plan we must keep in mind that its roots are special, and that it is less representational and more directly expressive than any other aspect of speech. The present section presents only some of the bare outlines of what a speaker puts into his intonational plan. A full treatment would necessarily involve analyses of emotions and attitudes, as well as a review of the extensive intonation literature on British and American English, Dutch, Danish, Swedish, French, and other languages. This would go beyond the framework of the present book. The reader is referred to the following sources: Bolinger 1986; Brown, Currie, and Kenworthy 1980; Cruttenden 1986; Cutler and Ladd 1983; Gårding 1983; Gussenhoven 1984; Halliday 1970; 't Hart and Collier 1975; Ladd 1980, 1986; Ladd, Scherer, and Silverman 1986; Liberman and Pierrehumbert 1984; O'Connor and Arnold 1973; Pierrehumbert 1981; Scherer 1986; Thorson 1983; Vaissière 1983; Van Bezooijen 1984.

The structure of intonational phrases

The intonation contour of connected speech is organized over smaller or larger phrases, called *intonational phrases*. An intonational phrase consists of one or more phonological phrases or metrical groups, but there is no general rule dictating the size of an intonational phrase. The speaker may decide to make smaller or larger intonational phrases, depending on such factors as rate of speech, formality of the communicative situation, and so on. Still, it may help to give some examples of quite natural intonational phrases. Often the sentence as a whole is an intonational phrase, especially if it is not too long; *How are you?*, *Go and get the newspaper*, and *Henry's* *falling asleep* are normally pronounced as single intonational phrases. When the sentence is longer, or more complicated in structure, it may be broken into two or more intonational phrases. This is quite naturally done if the sentence contains a parenthetical, a nonrestrictive relative clause, or a tag question, as in the following examples:

(4) 1 2
//Connected speech // as will now become apparent //
3
consists of intonational phrases //

(5) 1 2 3
//The golden temple // which is still in use // was built by the Sikhs //

(6) 1 2
//He's your uncle // isn't he? //

The parenthetical in example 4, the nonrestrictive relative clause in example 5, and the tag question in example 6 are intonational phrases of their own. In examples 4 and 5 the inserted phrase breaks the main clause into two parts, each of which becomes an intonational phrase of its own. In example 6 the tag phrase is added to the main clause, which is an independent intonational phrase.

One intuitive test for the presence of an intonational phrase is that it can be surrounded by grammatical pauses. It is quite natural to insert such pauses at the double slash markers when reading sentences 4–6. Intonational phrases often display other metrical properties as well. The initial words are often spoken in the way of an *anacrusis*—a string of high-rate nonaccented syllables, which form sort of an "upbeat" to the phrase as a whole. This contrasts with what often happens at the end of an intonational phrase: a certain lengthening of the final syllable or of the final stressed syllable. The defining characteristic of an intonational phrase is, of course, that it displays one of a set of *tones* (meaningful pitch contours). An intonational phrase is always a sense unit of some sort: a sentence, a clause, a modifier/head combination, or a predication. Also, there is always at least one pitch accent in an intonational phrase. If no element is focused in surface structure, the pitch accent goes to the last lexical head (by the nuclear stress rule).

Examples 4–6 gave rather clear cases of intonational phrases. The clause, in particular the finite clause, is a privileged candidate for becoming an intonational phrase. A sentence modifier may then be set apart as an independent intonational phrase, as in example 7.

(7) 1 2
// Unfortunately // John had lost his purse //

A special case of intonational-phrase partitioning occurs in so-called *listing*, as in example 8.

(8) 1 2 3
//Could you bring the tent // the barbecue // the charcoal //
4
and the icebox?//

Before we turn to the melodic properties of intonational phrases, a further crucial notion has to be introduced: the *nucleus.* Each intonational phrase has one and only one nucleus. The nucleus is the most prominent pitch accent in the intonational phrase. If there is only one pitch accent, it will be the nucleus. If there are more, the last one will usually be the most prominent. This is, for most intonational phrases, a natural consequence of the nuclear stress rule. The nuclei in example 7 are the syllables *for* and *pur*; the nuclei in example 8 are *tent*, *bar*, *char*, and *ice*. Still, the nucleus may also occur in an early position. Take the second intonational phrase in example 7, but now with John carrying contrastive accent: *jOhn had lost his purse* (as opposed to Peter). In that case, *John* has the nuclear accent in the phrase.

The nuclear syllable of an intonational phrase receives *primary accent.* All other syllables receiving an intonational accent are said to have *secondary accent.* All intonational accents (for short: accents) are made by some sort of pitch movement—a rise, a fall, or some combination of a rise and a fall. But not all stressed syllables receive an intonational accent. They can become rhythmically more prominent by vowel lengthening or extra loudness. In *Mary finished her Russian novel*, for instance, there were metrical peaks on the syllables *Ma*, *fi*, *Ru*, and *no.* Not all these will undergo pitch movement. The nucleus *no* will, of course, but the other three could be pronounced at a constant pitch level. In that case they would be stressed but not accented. We will call this *tertiary stress* (as opposed to primary and secondary accent).

Though the notion of intonational phrase is widely used and accepted, it is not uncontroversial. Ladd (1986) has reviewed its theoretical and empirical foundations.

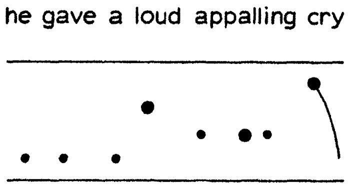
"Interlinear-tonetic" notation

There is no standard notation system for pitch contours. Such a system should be neither too concrete nor too abstract. It would be too concrete if it were to represent the actual course of an utterance's fundamental frequency (its F_0), the vibration frequency of the speaker's glottis. The phonetic plan does not provide that frequency course in all detail. The same pitch contour will be uttered at a higher frequency level by a female voice than by a male voice, and it is this "same" that we want to capture. Pitch contours are, moreover, equivalence classes. The same contour can be uttered with all sorts of accidental variations. What we want to capture are the *relevant* movements—those that carry the intonational meaning. These relevant movements are probably a rather limited set, as has been experimentally demonstrated by 't Hart and Collier (1975) and other phoneticians.

The system should also not be too abstract. It is probably insufficient to recognize just two classes of pitch, high and low, every pitch movement being just a switch of level. The most important meaning-bearing pitch phenomena are precisely in the shape and size of pitch movement. Those properties should be captured in the representation of the speaker's intonational plan.

In the following we will opt for so-called *interlinear-tonetic* notation (see O'Connor and Arnold 1973; see also Cruttenden 1986, whose analysis is largely followed in this section), which steers a convenient middle course between too concrete and too abstract a representation. This notation is exemplified in the following diagram:

(9)

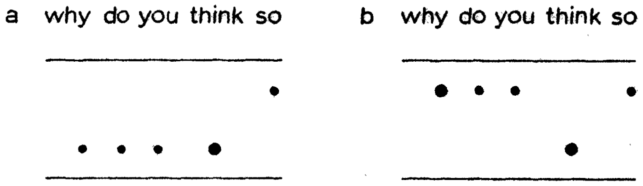


The two horizontal lines represent the upper and lower boundaries of the speaker's pitch range. The dots represent syllables. The fat dots are stressed or accented syllables. The first stressed syllable in the example is *loud.* It has pitch accent, a jump up from *a*. This is a secondary accent, since loud is not the nucleus of the intonational phrase. The syllable *pa* is also stressed, but not through pitch movement; it is in the middle of a level pitch contour. This is a case of tertiary stress. The final syllable, *cry*, is the nucleus of the intonational phrase. It has primary pitch accent, through the step up from the previous syllable. There is, moreover, a downward movement within that syllable, all the way back to the baseline. It is called the phrase's *boundary tone*, the closing movement on the last syllable. This is, as we will shortly see, an essential part of the *tone* of an intonational phrase.

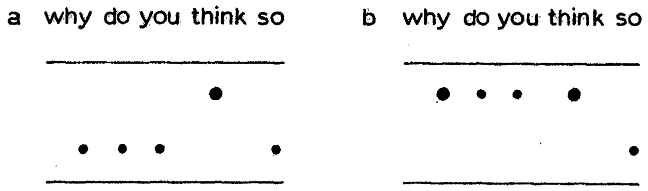
Prenuclear tune and nuclear tone

The melody of an intonational phrase can now be divided into two parts: the stretch preceding the nucleus and the part from the nucleus to the end of the phrase. These parts will be called the *prenuclear tune* and the *nuclear tone* (or just *tone*) of the intonational phrase. The intonational meaning of the phrase is essentially carried by the nuclear tone. The prenuclear tune can modify that meaning—can soften it or sharpen it—but cannot essentially change it. This means that a speaker expresses intonational meaning quite locally, mostly at the very end of the intonational phrase. This shouldn't be too surprising. The nucleus is the most prominent element in the utterance. It is an element of high interest in the interaction, one that should capture the listener's attention. It is the natural apex for the expression of intonational meaning (emotional, attitudinal, or rhetorical). The prenuclear part of the intonational phrase, on the other hand, is less prominent, often containing mostly given information to which the focused nuclear information is to be added. This distinction between the functions of tune and tone may be clarified by the following four examples.

(10)



(11)



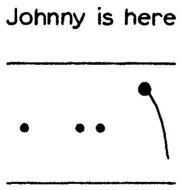
In all four examples, *think* is the nucleus; the nuclear tone stretches over *think so*, the prenuclear tune over *why do you.* Examples 10a and 10b carry the same nuclear tone, a full rise (from *think* to *so*). When the question is put in this way, it conveys the speaker's genuine interest in the answer. The nuclear tones of examples 11a and 11b are also the same; they are both full falls. Here the speaker conveys a touch of disagreement, a certain distancing from the interlocutor's position. These open versus distanced attitudes expressed by the tones in examples 10 and 11 are then further modified by the prenuclear tunes. The a tunes are low-level tunes; the b tunes are high-level. The high-level tunes tend to strengthen the character of the following tones. There is more empathy in the speaker's openness in example 10b than in 10a, and there is slightly more emphasis in the speaker's distancing in 11b than in 11a.

These are, of course, rather subtle interactions. It is not surprising to find quite diverse accounts of prenuclear tunes (alternatively called "pretonic accents"). Not only do intuitions differ substantially between authors, but it is likely that real differences in the character of these tunes exist between various dialects of English. The following discussion will be limited to a short review of the nuclear tones.

Some common tones

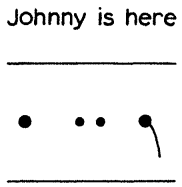
How is the nuclear accent made? Let us review seven of the more common tones. They have three distinguishing features, namely (i) whether they fall or rise from the nucleus, (ii) whether the nucleus itself is high or low with respect to the previous tune, and (iii) whether another pitch change occurs after the nuclear move.

Tone I: High-fall The nucleus peaks up from the preceding level, then falls all the way down to the base level. Example:



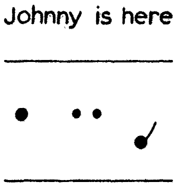
This tone expresses seriousness, in a matter-of-fact way. It is probably the most common tone for declaratives.

Tone II: Low-fall There is a fall, but without a step up toward the nucleus. Keeping the prenuclear tune at the same mid-level pitch, the tone looks like this:



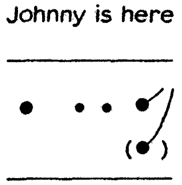
This tone also expresses seriousness, but it is somewhat less involved and more businesslike.

Tone III: Low-rise The nucleus starts at a lower level than that at which the previous tune ended. The nuclear move is a small rise, like this:



Whereas tones I and II do the "natural" thing in reaching the to-be-accented nucleus—namely, jump up—here the nuclear accent is made by stepping down. There is something contradictory in this tone: The speaker accents and plays down at the same time. Many authors interpret this as a way of reassuring. The rise at the end invites a reaction, but the addressee is given to understand that the matter is not a crucial one.

Tone IV: High-rise The nucleus starts slightly up from the previous tune, and then rises:

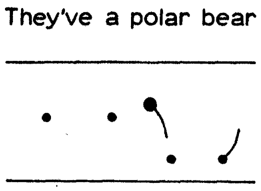


This tone is used when seeking confirmation. It is mild, and it expresses genuine interest, maybe with a touch of incredulity. It also has an echoic function, seeking confirmation for what was just said, as in:

A: I saw Johnny coming in. B: Johnny is here?

A variant of tone IV is the *full-rise*, which is indicated in parentheses. As in tone III, the nucleus starts low; but there is the high-pitched ending of tone IV. Semantically, the full-rise is more similar to the high-rise than to the low-rise.

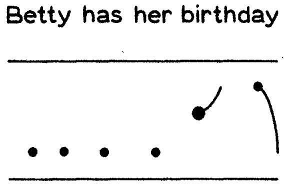
Tone V: Fall-rise This is a cup-like movement beginning at the nuclear syllable, and extending over any remaining syllables. The movement usually begins slightly up from the preceding tune. An example is this: Somebody asks *Isn't that a very small zoo?*, and the answer is



This tone characteristically expresses some reservation. What the speaker says is meant to be in contrast or even contradiction to what was apparently assumed or expected. The presence of the polar bear shows that the zoo is not to be belittled. The rise at the end suggests some (implicit) follow-up, the right conclusion to be drawn (*so the zoo cannot be that insignificant*). The same tone on *holidays* in *she didn't take holidays* could go well with a continuation like *she quit her job!* The tone can even affect the scope of negation, as Cruttenden (1986) showed. If somebody says *I am not going to perform anywhere*, the scope depends on the tone. With tone I on *anywhere* it means that the speaker is not going to perform at all; however, with tone V it means that the speaker *is* going to perform but not in just any place (say, only in Carnegie Hall).

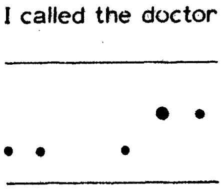
Sometimes, there is another kind of meaning expressed by this tone: a self-justificatory "I told you so" or "I am telling you so". This would hold for the following sentences (nuclear vowel capitalized): *I knEw he was mean* and *You will nOtice he drinks.*

Tone VI: Rise-fall Here the nucleus has a rising pitch obtrusion, but the phrase ends in a full fall. The tone stands out best when the nucleus is a step up from the preceding tune, like this:



This, like the other tones with final fall (I and II), expresses completion; no continuation is invited or suggested. But in addition it expresses a great deal of enthusiasm or being impressed. The "cap" of this tone contrasts with the "cup" of tone V, and there is a related contrast in meaning. There is no reservation; rather, there is full endorsement, unrestricted commitment, overriding potential doubt or opposition.

Tone VII: Level Here an even pitch is maintained, or at most slightly raised from the nucleus to the end of the intonational phrase. The only accent is in the step up or the step down to the nucleus. The following example is one with a step up:



The most characteristic trait of this tone is its nonfinality. By not turning to baseline, as in tone I or tone II, it conveys that there is more to come. But the speaker is not inviting a reaction, either. This is, therefore, typically a neutral tone for a nonfinal phrase. It is, for instance, a good tone for listing, as in example 8 above: *Could you bring the tent, the barbecue, the charcoal,....*

The meanings of these major nuclear tones are, in part, conventional, and specific to English (or even to certain dialects of English). Still, there are aspects to these meanings that are probably more universal. Tones with a final fall express completeness, finality. Tones with a final rise express nonfinality, openness. One might say that falling tones assert the position of the speaker, whereas rising tones reach out to the addressee (to invite a reaction, to challenge, or whatever else).

Key and register

There are at least two other potentially universal aspects to intonation. The first one concerns *key*. Key is the range of movement in an intonational phrase. A speaker can make more extended falls or rises by raising his high pitch level, the peaks of his intonation (not by lowering the baseline). Brazil, Coulthard, and Johns (1980) distinguish three levels for the peak intonation: high, mid, and low. The speaker must select one of these for each and every intonational phrase. Changing key may, among other things, have a function in backgrounding and foregrounding. The three nuclei in example 5 above (repeated here as example 12) are *tem*, *use*, and *Sikhs*.

(12) The golden temple, which is still in use, was built by the Sikhs.

A natural intonation for the three intonational phrases of this sentence would involve mid, low, and high key, respectively, i.e., moderate pitch accent on *temple* (which is given information), minor pitch accent on *use* (which constitutes background information), and a major step up on *Sikhs* (which constitutes the newly focused information in the utterance). This relation between pitch range and intended attentional effect might well be universal in the world's languages. Key is probably also related to rate. Deese (1980) found that high-rate stretches of speech tended to have subdued intonation.

A second aspect is *register*. Register is the pitch level of the baseline. The scream is universally high register, and high-register speech may have a similar root: It expresses emotion and tension. It also expresses "smallness"—the register of the child, and metaphorically the register of helplessness and deference.

The key notions for the speaker's intonational plan are the intonational phrase, its nucleus, its tune, its tone, its key, and its register. They are the codeterminants of the intonational contour. The speaker surely has no intonational lexicon with whole ready-made contour templates. The closest thing to that may be something like a tone lexicon, a relatively small set (but larger than seven) of canonical tonal contours. These contours are expressive and meaningful. In the process of phonological encoding, they have to be projected onto the stretch of speech that extends from the nucleus to the end of the intonational phrase.

Summary

Phonological encoding is the speaker's construction of a phonetic plan. This chapter has outlined the structure of phonetic plans, both at the word level and at the level of connected speech. The plans for words are built on their morphology, and they involve a phonological organization at various levels or tiers. The derivational and inflectional morphology of words were reviewed, and it was suggested that, for almost all words used by a native speaker of English, this internal structure is stored in his mental lexicon.

The phonetic form of words involves qualitative and prosodic aspects. The qualitative aspects concern the phonetic material that is to fill successive slots of the skeletal tier. This material can be categorized as more or less sonorous (V or C). It is further characterized by various other phonetic features—particularly, laryngeal and supralaryngeal ones, such as manner and place features. One could, in fact, represent the quality information of segments on a set of independent "feature tiers." Successive phones are given some syllabic organization: Each syllable contains a sonorous peak, which may be flanked by less sonorous phones. A syllable's rime characterizes it as either strong or weak, and this, in turn, affects the rhythmic structure of the word in which it figures.

The prosodic aspect of a word's phonetic plan involves its metrical and intonational organization. Each word has an internal organization of more or less stressed syllables, and this is part of the stored code for the word. This metrical pattern can be conveniently represented by way of metrical grids. The syllable that carries main stress is the one that will be given pitch accent if that is specified in surface structure. A word's intonation is not lexically stored in languages like English, although it is (to some degree) in so-called tone languages. In English, the intonation of a word depends on whether it receives pitch accent and on the melody of the intonational phrase in which it partakes.

A speaker's plan for connected speech is not a simple concatenation of word plans. First, a speaker may select clitic allomorphs instead of free-standing words when producing connected speech, such as *I've* instead of *I have*. These and other forms of cliticization create new word-like units—phonological words—and usually involve some degree of resyllabification. There are also various other segmental and syllabic changes that can occur at word boundaries in connected speech.

Connected speech also involves specific metrical planning. We considered an organization into what some call "phonological phrases"—stretches of speech leading up to lexical heads of NPs, VPs, or APs. Some such metrical packaging undoubtedly exists in speech; it highlights a sentence's surface-structure organization. There are, in addition, intonational phrases, consisting of one or more of these coherent metrical units. They are domains for the assignment of intonation.

Intonation of connected speech was the final subject of this chapter. It was stressed that the function of intonation is expressive rather than representational, and that this makes it very special in a speaker's phonetic plan. We distinguished between two parts of an intonational phrase: the prenuclear part, whose tune has a qualifying or modifying function, and the remaining part from nucleus to end of phrase, whose tone carries the intonational meaning. We concentrated on different nuclear tones and their approximate expressive functions. Two further properties of a phrase's intonation were discussed: its key and its register.

Chapter 9

Generating Phonetic Plans for Words

How does a speaker generate the phonetic form of a word, given the developing surface structure? This chapter will characterize the process of phonetic planning as the spelling out of stored form representations and their projection on pronounceable syllables. The stored representations involve, in particular, the morphological, metrical, and segmental composition of words. A major task of phonological encoding is to generate a string of syllables that the Articulator can accept and pronounce. Syllables are basic units of articulatory execution. As was outlined in the preceding chapter, they consist of phones which, when executed, are complex and temporally overlapping articulatory gestures. The adult speaker, we conjecture, has an inventory of syllables. They need not be generated from scratch over and again. Rather, these stored articulatory patterns are addressed during phonological encoding on the basis of the spelled-out word representations. Also, certain free parameters are set, such as a syllable's duration, stress, and pitch. The eventual phonetic plan is a string of such specified syllables.

This chapter deals with the phonological encoding of single words. There is a certain drawback to this: It may seem as if phonological encoding is a wasteful process. Spelling out a word's segmental makeup also makes available the stored syllabification of the word, i.e., the segments' abstract grouping in syllables and syllable constituents. Why then should there be a second phase where strings of segments are used to address stored syllable representations? The main reason for this seemingly roundabout way of phonetic planning is to be found in the generation of connected speech. A word's stored syllabification is not sacrosanct. In connected speech, words often form coalitions with their neighbors that lead to so-called resyllabification. A phrase like *I gave it*, for instance, is easily resyllabified as *I ga-vit*. This enhances the fluency of articulation. To make an optimally pronounceable phonetic plan, the Phonological Encoder needs the seg-