Representing linguistic phonetic structure

Peter Ladefoged

1. What do we want to represent?

1.1 Limiting the scope of phonology.
- Kinds of information conveyed by speech
- Sociolinguistic differences are not the same as linguistic differences
- What counts as a linguistic contrast

1.2 Non-linguistic aspects of speech\
- Speaker attitudes, style, peronal identity
- Spoken language and written language
- The role of intonation

1.3 Universal phonetics and phonology
- Equating sounds in different languages
- IPA categories and phonological feaures
- Innate features
- The number of possible contrasting sounds

1.4 Language as a social institution
- The mental nature of language
- Language as a self-organizing social institution
- Filling holes in patterns
- Phonemes as linguistic constructs

2. What is the purpose of a linguistic phonetic representation

2.1 The nature of phonological features
- Describing sound patterns vs. specifying the lexicon
- How sound patterns arise

2.2 The need for auditory/acoustic features
- Auditory characteristics of vowels
- Other auditory features

2.3 Defining characteristics of features
- Features are defined in terms of acoustic or articulatory scales
- Articulatory features do not have invariant acoustic correlates

2.4 Multi-valued features.
- Features with an ordered set of values

1. What do we want to represent?

1.1 Limiting the scope of phonology.

Before we discuss the phonetic representation of a language, we must consider what it is we want to represent, and why we want to represent it. Phonetics is the study of the sounds of speech, so it is these sounds that we want to represent. But what is a speech sound? This in turn leads to another question, namely, what is speech?

An obvious answer is that speech is the soundwave produced by a speaker and interpreted by a hearer. But that is only part of the story. That soundwave conveys lots of information and not all of it is part of language. The aim in this book is to study how we can best represent just the linguistic aspects of speech, not other things such as the age and gender of the speaker. We are concerned with the relation between phonetics, the actual sounds, and phonology, the ways in which these sounds form significant linguistic contrasts. In studying this relation our first task is to define what we mean by significant linguistic contrasts.

Linguistic contrasts are the building blocks of words, the units that enable us to express what we know about the world. The words of a language provide us with a way of organizing our experiences, and thus give us ways of grouping these experiences into categories, so that we can qualify and relate one experience to another. As a result we are able to form concepts and manipulate ideas. Some kinds of thinking may be possible without words, but we could never develop scientific theories without them. As Wittgenstein (1958) puts it, “The limits of my words are the limits of my world.”

Speech produces words that convey information (or misinformation) about the world. In order to convey this information, words must be distinguished from one another by their phonetic characteristics. It is these and only these characteristics — the ones that enable words to convey information about the world — that we shall regard as the linguistic phonetic properties that we want to describe. Other information conveyed by spoken words, such as whether the speaker is male or female, old or young, or whether they belong to a particular sociolinguistic group, all these are not part of the language being described. They may be interesting aspects of speech, but they are not part of the linguistic phonetic structures.

In this book we will not be concerned with determining the set of phonetic features requires for conveying sociolinguistic information. Sociolinguistic information is clearly codified, although often not in terms of discrete oppositions of the kind that convey different meanings. Different degrees of diphthongization in the vowel in *plate* marks a person as belonging to a particular social class in London (or Australia, or anywhere else that uses this distin guishing characteristic). Anyone familiar with the regional accents in question can easily place a speaker by these sociolinguistic marks. Listeners are aware that these differences of degree of diphthongization are not linguistic contrasts.

Linguistic information and sociolinguistic differences between languages are sometimes conveyed by the same devices. In the South of England the words *cat* and *castle* are pronounced with different vowels, whereas in most forms of

American English the vowels in these two words belong to the same phonological unit. This kind of difference, involving the different distribution of items that might be contrastive, is a frequently used sociolinguistic device, as has been noted by many authors. Thus Gershwin and Gershwin (1933) describe two speakers who have the same set of possible phonemic contrasts but use them in different words, as in *You say* ['iðəˑ] *and I say* ['aiðə]. *You say* ['niðəˑ] *and I say* ['naiðə].' (It should be noted, however, that Gershwin and Gershwin's phonetic observations are not always reliable. They correctly observe that some people pronounce the word *tomato* as [tə'meitoʊ], whereas others say [tə'mɑtəʊ]; but they further claim to have observed the word *potato* pronounced as [pə'tɑtəʊ]. This seems very unlikely.)

In addition to these gross differences, there are many other differences between dialects and between the sounds that characterize the accents of different languages, which depend on much more subtle phonetic effects. For example, in a survey of speakers of different types of English, I found that 27 out of 30 speakers of Californian English used an inter-dental [θ] in which the tip of the tongue was protruded between the teeth in words such as *think, thin*, whereas 27 out of 30 speakers of different forms of Southern British English used a dental [θ] without tongue protrusion in these words. The difference between a dental and inter-dental [θ] is unlikely to be phonemically contrastive in any language, although it is an indicater of a difference in regional accent.

A great deal of evidence shows that languages and dialects are often differentiated from one another in ways that are *not* used to distinguish oppositions within any single language. Ladefoged and Maddieson (1996) have described many small differences between languages that are not used within languages. Fricatives can differ in small aspects of their spectral shape that are not used to distinguish one fricative from another in a language (Nartey 1982, Ladefoged & Wu 1984, Gordon et al. 2002). The degree of breathiness that occurs in one language may be different from that in another (Gordon and Ladefoged 2001) without that difference in breathiness being known to distinguish two sounds within any known language. Disner (1980) has described small cross-linguistic differences among vowels that are not contrastive within a language. Keating (1988) has shown that there are cross-linguistic differences in the extent to which different languages coarticulate adjacent stops and vowels, but similar differences have never been shown to contrast words within a language. Cho and Ladefoged (1999) have described small consistent differences in VOT between languages that have never been observed to contrast sounds within a language. These and many other papers suggest that the sociolinguistic functions of speech *cannot* be described entirely in terms of the same features as those that are used for describing phonological oppositions.

Despite the fact that phonologies are concerned with distinctions in only one lg at a time, there is a tendency for feature theories to try to account for small phonetic differences between languages. Thus Jakobson and Halle (1956) can express more phonetic detail than Jakobson, Fant and Halle (1951); Chomsky and Halle (1968) add still more features; and Halle and Stevens (1971) add further complexities so as to be able to describe yet more phonetic differences between languages, replacing the feature Voice by a set of four features, Stiff, Slack, Spread, and Constricted. This enabled them to characterize the phonetic differences between, for example, English [p] as in *spy* and the Korean so-called lax [p]; but the cost is that they no longer have the more phonologically useful opposition voiced-voiceless, and their feature system

has been considered unacceptable for phonological classification (Anderson 1978).

Much of the elaboration of feature systems serves only to describe how the speech of one group of people differs systematically from that of another group of people. But, as we have seen, there is no theoretical or empirical reason to expect speech systems to use the same devices for phonological and sociolinguistic purposes. Differences between one accent and another are often very subtle and not sufficient to distinguish the lexical meanings of words. Sociolinguistics is an interesting and worthwhile discipline; but a sociolinguistic description goes beyond the study of a particular language. For the moment, we will consider language to be a system of contrastive units that are codified in terms of discrete oppositions, used for communication by a particular group of people.

There are still many problems in this definition of language. What do we mean by a particular group of people? We could, perhaps, in the case of English, find a group of people who had been educated in the same local schools and had lived all their lives in the same district. But even these people will have been influenced by their neighbors. The odds are that they will also have listened to radio and television stations from other areas, and seen films in which people had other accents. I know from personal experience that it is often hard to find a group of a dozen people who speak in the same way.

We should also remember that in most parts of the world speakers are not like monoglot English speakers. Around half the 7,000 languages in the world are spoken by less than 10,000 people, and over a quarter are spoken by less than 1,000 people (Ethnologue 2005). The majority of people in the world are at least partially bilingual. When you grow up speaking a language spoken by only a small number of people you soon realize that you have to know how your neighbors speak. You have to trade wih them and keep the peace. It is only in major countries with a large population speaking the dominant language that there is no need to know a second language. In these large monoglot countries, however, there are still variations in the way people speak. There are different dialects and various styles of speech used in different circumstances. Speakers often do not separate these different forms of the language, using bits of different dialects and styles in a single sentence. The usual situation facing linguists trying to describe a language is that there is what Ferguson (1959) called diglossia, the existence of two languages or linguistic varieties having different status. Speakers have coexistent phonemic systems

Linguistics is often a game of let's pretend, and we have to acknowledge that we may be pretending that a particular form of a language exists with no variations due to outside influences. Ultimately we must describe a language with all its variations. But before we can describe a mixture we have to describe the components that make up the mixture.

There are also problems with a view of language considered as consisting entirely of discrete contrastive units. Sometimes these units are not as discrete as one would wish in an ideal linguistic world. Many years ago Trager and Smith (1951) suggested that there was a contrast between the words *just*, as an adjective, in phrases like *He's a just man*, and in the adverbial form as in phrases such as *just a minute.* They observed that the latter form was consistently produced with a high central vowel that could be represented by [ɪ] (barred i), in contrast with the mid central vowel [ə], which they used for the adjectival form. Should the high central

vowel be regarded as an additional phoneme in English?

In normal spoken language the degree of precision with which a sound is produced depends not only on non-linguistic factors such as the amount of whisky the speaker has drunk or the nearness of a man-eating tiger, but also on factors within the language itself. Pierrehumbert (2001) cites a number of studies showing that well-known high frequency words are pronounced with less precision than unfamiliar words. Since this is plainly a property of the language, and not something due to external causes, it should be part of the description of the phonetic structure of the language. Pierrehumbert (2003) describes good ways of tackling this problem, using an exemplar model of each sound and a probabilistic phonology. Here, however, we are concerned with basic notions of how to represent sounds and we will neglect these variations in articulatory precision. We will take it that in a basic description of a language each of the sounds making up the words in the lexicon has only one specification, irrespective of whether it is in a familiar word or an exotic technical term.

The notion of phonemic contrast is also not as simple as it might seem at first glance. There are many near neutralization phenomena that provide problems. For example, German voiced and voiceless stops contrast initially, but the contrast is at least partially neutralized in word final position (Port & O'Dell 1985). It therefore seems as if speakers of German do not make a distinction between words such as *rad* 'wheel' and *rat* 'advice'. But measurements of their recorded utterances show that there is a small, consistent difference in their production of these sounds. As Port (1996) puts it: "The phenomenon of 'incomplete neutralization' and the subtlety of this incompleteness reveal vividly that speech sounds do not fall into discretely distinct phonetic types."

Similar phenomena found by Labov (1991) are what he calls 'near mergers', words that some speakers hear as being the same, but which are nevertheless found to be different when analyzed acoustically. In the incomplete neuteralization phenomena exemplified above for German, two sounds that are distinct in some contexts are barely distinct in another context. In near mergers it is a matter of two sounds being very nearly the same wherever they occur. Labov found near mergers for the vowels in *sauce* and *source* in New York City, *fool* and *full* in Albuquerque and Salt Lake City, *line* and *loin* in Essex, and *ferry* and *furry* in Philadelphia. Faber and Di Paolo (1995) found that listeners typically heard two sounds as merged before speakers had in fact merged them in production in many of these cases. For speakers with a near merger, should the two vowels in these words be considered linguistic contrasts? They are, after all, different.

Ladefoged & Everett (1996) discuss other borderline cases in which a sound occurs in so few words in a language that it is not clear whether it should be considered part of the phonology. For example, in Margi there is a labiodental flap (Ladefoged 1964) that occurs in only five words in an extensive dictionary of the language (Hoffmann 1963). In Wari' there is a special kind of bilabial trill that occurs in only 24 words (MacEachern, Kern & Ladefoged 1997). Is five words too few, but 24 enough to count as a phonological segment?

Marginal sounds of this kind occur not only in little known languages but also in English and other major languages. I have a few words that I pronounce with a nasalized vowel and no final consonant, such as *restaurant, salon, crouton*

(contrasting with *aunt, talon. coupon*). Is this nasalized vowel a phoneme of my English?

Speakers are often aware of some distinctive phones in their language. Many speakers of Russian with no knowledge of linguistics think that there is a sixth vowel in their language, [ɪ], not realizing that it is the allophone of [i] that occurs in hard consonants. Many speakers of British English recognize that they have a glottal stop in *button* (and some will recognize one in *butter*), and consider this sound to be different from that usually represented by *t.* Should this allophonic awareness be considered part of the language?

There are no simple answers to these questions, and accounting for them requires a more complex theory of phonetic representation than I will present here. (One approach that I would favor is that of Pierrehunbert (2005).) For the moment we will take it that our major task is to represent clear linguistic differences that occur among citation forms of words in the lexicon. Once we can describe this part of language we can expand it to cover the problems raised above. But we cannot do that until we have a basic description from which we can start stating the probabilities of other events.

1.2 Non-linguistic aspects of speech

There are many functions of speech that cannot, and should not, be expressed in terms of any of the traditional sets of phonological features. These functions include to convey the attitude of the speaker to the topic under discussion, to the person addressed, and, indeed, to the world in general. These attitudinal differences may not be codified in the same way as linguistic information. They are conveyed largely by intonation, which is, of course, also used for conveying linguistic information. However, many aspects of intonation, such as sarcastic or simpering intonations are not part of language, and therefore should not be described within phonology. We should not build phonological devices that have special units for contrasting these possibilities. Again, the study of the techniques for conveying these aspects of speech is a worthwhile pursuit; but it is not part of the study of the formal system of a language, which is what we are considering here.

Similarly, we should not consider emotional effects. There seems to be something in common to expressions of anger, astonishment, sorrow, doubt, and love in many different languages, making them general human traits, rather than linguistic communicative devices. To the extent that they are not the same in all languages, they are learned, cultural, aspects of behavior. Many Englishmen consider it normal to speak in a phlegmatic way with a narrow intonation range that Americans consider as indicative of boredom. The Navaho tend (by American English standards) to speak very softly. As Nihalani (1983) has pointed out, Indian English typically sounds rude or aggressive to speakers of British or American English. All these are learned aspects of particular cultures; but they should not be considered as part of language.

Yet another aspect of speech that should be mentioned here is that conveyed by the style. Again neglecting entirely the lexical and syntactic content, I can say the sentence *We hope for fine weather* in a way that sounds as if I am conveying factual information, or saying a prayer in public, or reading a poem on the radio. These

differences in style are equivalent to speaking in a different dialect. They don't depend on the same phonetic devices as are used for distinguishing one word from another.

The final kind of information conveyed by speech that nobody would call a function of language, is that it signals the identity of the speaker. When I walk into a house and call out *Hi, it's me* this is all the information I am conveying. I am not really making a declarative statement. Provided it was normal for me to do so I could have established my identity by walking in and saying any other phrase, maybe *Tra la la* or *Hi de hi*. Personal information can sometimes be difficult to distinguish from sociolinguistic information. We each speak in the way that we do partly because our particular vocal organs have certain characteristics, but also because we choose to use within limits, our own personal style of speech. Often what might seem to be a personal characteristic of a particular speaker is in fact something that he or she has chosen to copy, which is shared by a small sociolinguistic group. But irrespective of whether this learned vocal behavior conveys personal or sociolinguistic information, it is still not part of the formal system of a language.

To be more precise about what should be included within language (and hence within phonology), I suggest that we should consider language to be just the system of sounds that we need for modeling our known world. Considered in this way, spoken language is the direct counterpart of written language. Instead of speaking of a language being *reduced* to writing, implying that some part of spoken language is not present in the written form, it would be better to say that (virtually) all that is *language* can be expressed in speech or in writing — and all the sociolinguistic, attitudinal, emotional, stylistic and personal information, that is present in speech and less evident in writing is not part of what we want to define as language. Language is represented by all the encoded aspects of speech except those that convey information about the speaker's identify, attitude, emotions, sociolinguistic background, or style of speech, in so far as these are not conveyed by syntactic or lexical devices.

Writing may convey some sociolinguistic, personal, or stylistic information. But, neglecting small differences in spelling, such as *color* vs. *colour*, it is almost impossible to tell from printed pages whether they have been written by an Englishman or an American, or indeed, by a speaker of one of many other forms of English. You could gain some sociolinguistic information if I were to use certain marked phrases or lexical items such as talking about a *lift* as opposed to an *elevator*, or a *full stop* as opposed to a *period*. In addition, other more informal lexical items might convey stylistic or personal information. But in all these cases, when the written language does convey sociolinguistic or other information, it does so by means of precisely those phonological devices that are used to convey information about the topic under discussion. We do not need a special form of phonetic representation to be able to handle information of this sort. All we need is to be able to identify the linguistic oppositions of exactly the same kinds as those that are used to distinguish words.

Writing also conveys certain aspects of intonation. From the syntax, morphology, word order and punctuation we can determine something (but far from everything) about the intonation that a given sentence could have. As Bolinger (1977) has pointed out, the semantics further circumscribe the possible intonations, but again only to a limited extent. The intonational aspect of regional accents are also not

represented on the printed page. When speakers of Irish, Welsh, American, or Scottish English read a page such as this one aloud, there will be differences in their intonation patterns that cannot be ascribed to anything written down. From the point of view I have been developing so far, these differences are not part of the language; they convey only sociolinguistic information about the speaker.

The role of intonation is undoubtedly the most problematic part of this attempt to distinguish what is formally part of the language and what is not. Much work on intonation takes note of aspects of speech that are not represented in the written language, and might, or might not, be considered to be truly part of the language. A generation ago, linguists used to argue whether four pitch levels and a number of junctures were sufficient to capture all the meaningful contrasts in English (see, for example, Trager & Smith 1951, Stockwell 1960). But it was often not clear whether the contrasts being discussed were really linguistic contrasts. Similar discussions are still in progress using a different set of phonetic devices. Thus Pierrehumbert (1980) has an elaborate discussion of the many different intonations that can occur on the phrase *Manny came with Anna.* In these discussions the notion of a meaningful contrast is not the same as it is in discussions of other aspects of phonology. Intonation includes aspects of speech that are conveying the speaker's attitude rather than objective information about the world. Outside of discussions of intonation, these aspects of speech are not usually considered to be part of the language. For example, lengthening segments to indicate superlatives, as in [ɪt **wəz** bɪːːg], is plainly a gradient matter, rather than a simple contrast, and would not lead to discussions of how many degrees of vowel length there are in English. (A possible exception is discussed by Prince (1980), who has a brief mention of the realization of emphasis in Finnish.)

If we limit phonology to the study of simply those aspects of speech that convey information that can be expressed in terms of a formal system of contrasts, then the relation between phonological and phonetic units becomes much more straightforward. The phonologist is no longer under an obligation to describe the phonetic details that characterize one speaker, one dialect or one language as opposed to another. This is a slightly different view from that which I have often expressed before (Ladefoged 1971, 1980), and still different from that propounded by Chomsky and Halle (1968), and later by Halle (1983, 1988). I used to maintain that "a linguistic theory should be able to characterize both the oppositions within a language (the differences between the members of the set of all possible sentences) and the contrasts between languages (all and only the features which mark the sounds of the language as being different from the sounds of other languages)." (Ladefoged 1971:275.) Chomsky and Halle do not explicitly discuss this point, but their work can be read in a similar way; and a similar point of view has certainly been expressed by Schane (1973) who says "Linguistically significant differences are those which characterize native control of a language." Now, having no reason to believe the distinguishing phonetic characteristics of languages are always expressable in terms of phonological features, and also seeing no formal way in which gradient sociolinguistic features can be considered along with categorical phonological features, I prefer to consider these aspects of speech separately. This view makes the difference between one dialect and another, or one language and another, a part of sociology, describable in the same way as any other indexical behavior, such as the dress, appearance, or patterns of belief that characterize a

particular group. Such things are part of culture; and there is authority (although perhaps not recognized as literally *good* authority) for saying: “Linguistics is not to be confused with culture.” (Stalin, 1950.) This view of language would also regard phonetic differences conveying emotion, or those aspects of the speaker's attitude that cannot be expressed by syntactic or lexical devices, as part of the subject matter of psychology.

1.3 Universal phonetics and phonology

We will now consider an elaboration of this theory of the proper domain of phonology, which leads to a theory more similar to that of Chomsky and Halle (1968). The major weakness of the theory we have been discussing so far is that it does not allow us to say much about the nature of human language as a whole. To do this properly we must be concerned with the general properties of all languages, not just the set of contrasts within a particular language, which is all that is needed for conveying linguistic information. If we are to say something about the nature of human language we must have an overall phonetic framework that allows us to sum up what is different among languages and what is common to all of them. We want to be able to describe speech sounds in terms of some absolute phonetic standards, so that we can equate phonological patterns, distinctions, and rules among languages. This kind of phonetic representation will undoubtedly enable us to say more about the phonetic differences between languages. But the phonetic framework will not have been set up for this purpose, and there is no reason to expect it to be fully sufficient for discussing differences between languages.. As we noted above, sociolinguistic differences between dialects and languages are only sometimes the same as differences that convey linguistic information such as distinguishing words.

We want a theory of phonetics that will permit discussion of phonological universals, allowing us to make statements such as 20% of all languages have a particular set of vowels, all languages have at least two of the three stop consonants [p, t, k], and no language has more than 7 voiceless fricatives. In order to do this we have to know the set of possible sounds, and to decide when a sound in one language is the same as a sound in another. This is a historic problem. Before the rise of distinctive features (Jakobson, Fant & Halle 1951) and phonlogical features (Chomsky & Halle 1968), linguists such as Joos (1950) advocated ad hoc descriptions for each language. Hockett (1955) thought that “it is impossible to supply any general classificatory frame of reference from which terms can be drawn in a completely consistent way for the discussion of every individual language.” But all linguists (including those just cited) in fact use phonetic descriptions that imply some absolute frame of reference. As was pointed out many years ago by Fischer-Jorgensen (1952), unless we are able to use categories such as alveolar and bilabial, which refer to observable phenomena, we cannot describe the first and last sounds in *bib* as being in some way the same, and also know that the first sound in *bib* is not to be identified with the last sound in *did*.

This problem was first formulated by the International Phonetic Association when establishing a set of symbols to form an International Phonetic Alphabet. In 1900 the IPA published an *Exposé des principes* containing a table showing the recommended alphabet. This table was set up so that it contained “les sons distinctif de toutes les langues étudiées jusqu’ici”. Similarly, the 1912 English version, in a

section headed “Principles of transcription of languages hitherto not transcribed”, notes, long before the phoneme became a popular notion, “It is necessary to ascertain what are the *distinctive* sounds in a language, i.e. those which if confused might conceivably alter the meanings of words.” (Emphasis in the original.) The 1949 edition of the *Principles* makes the same point: “There should be a separate letter for each distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word.”

The second principle in the IPA *Principles* is the one that is directly concerned with equating sounds in different languages. It says: “When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound.” This principle is sometimes difficult to put into practice, as there is no definition of what is meant by very similar shades of sound.

Phonetic theory in the early days of the IPA was greatly influenced by the works of Sweet and Bell, both of whom had developed systems for classifying all the sounds that were known to distinguish words in the world’s languages. Bell’s *Visible Speech* (1867) and Sweet’s *Handbook of Phonetics* (1877) provided iconic symbols for showing the articulatory elements present in a sound. These same elements (or at least a subset of them) were used to define the symbols of the International Phonetic Alphabet. Throughout its history this phonetic alphabet has consisted of symbols defined in terms of intersections of phonetic categories that are similar to phonetic features. Most of the symbols are defined by the terms naming the rows and columns of the charts, together with the convention concerning voicing: when there are two symbols in a cell the first one is voiceless, and if there is only one symbol it is voiced. The symbol definitions also include the heading of the part of the chart in which they appear, such as “Consonants (Pulmonic)”. In addition a few symbols and all the diacritics are defined by supplementary notes. The whole work, principles, charts, symbols and notes, constitutes the IPA’s theory of phonetic representation.

Given this background we may now compare an IPA description with a feature specification of the kind that is commonly used. The location of [m] on an IPA chart explicitly indicates that this sound is: [+ voiced, + bilabial, + nasal]. By means of the labels above this section of the chart, the IPA system also indicates that this is a consonant made with the pulmonic airstream mechanism. In much the same way, Chomsky and Halle (1968:5) note that they will use symbols as "informal abbreviations for certain feature complexes." For them this symbol would be a shorthand way of designating the feature values: [+ voiced, + nasal, + anterior, – coronal, + sonorant] etc.

It is important at this stage to make a clear distinction between the name of a feature and the possible values that it can have. A feature is a physiological (or perhaps acoustic) property. The value of the feature denotes the extent to which the sound has this property. I will distinguish between features and their values by capitalizing the name of a features and placing the value of a feature in square brackets. Thus the IPA feature Vowel Height may be said to have values such as [close], [close-mid], [open-mid], [open] (or, as I prefer, [high], [mid-high], [mid-low] and [low]), and in standard phonological feature theory the feature High has values [+high] and [–high]. These conventions will be extended later when the hierarchial structure of features is discussed.

Both the IPA and standard phonological feature theory assume that there is a limited

set of phonetic categories; but neither of them states how this set can be delimited. Within general phonetics there is no theory that defines what counts as a speech sound for linguistic purposes. Obviously it is not just any sound that can be made with the vocal organs. We can make all sorts of grunts and snorts that are unlikely ever to be part of human speech. Accordingly, before we can have a valid theory of phonetics, we should be able to delimit the class of sounds that we wish to represent. There are obviously a number of borderline cases. For example in English, are a pair of dental clicks [| |], the clicking noise written *tut tut* or *tsk tsk*, part of the linguistic system, or is this just an exclamation? Presumably the latter, as this sound is not used in forming any words (one cannot say *he was* [||ɪŋ]).

Phoneticians of all traditions behave as if there were a well-defined framework that allows them to describe linguistic sounds in terms of categories such as voicing or nasality. (The precise feature set used is irrelevant to the argument at the moment.) But it should be fairly clear that although we have a well-defined phonetic framework, we cannot always describe sounds in terms of categories. Whenever we find a sound that has not been described previously in the phonetic literature, we simply invent new categories, or permit combinations of categories that had previously been declared impossible. Thus when I first heard a labiodental flap in Margi I had no qualms in extending the meaning of the term flap (Ladefoged 1968), despite the fact that the gesture involved is nothing like that in any other flap that had been previously reported. And when Tony Traill heard strident vowels in !Xõo, in which the main source of acoustic energy is the movements of the aryepiglottic folds (Traill 1985, Ladefoged and Traill 1980), we simply added ventricular to the list of possible phonation types. Similarly neither I nor any other linguist seems to worry about saying at one moment (Ladefoged 1971) that true velar laterals cannot be made, and shortly afterwards (Ladefoged, Cochran and Disner, 1977) describing these sounds in Melpa and other languages spoken in Papua New Guinea. Nor is it considered odd that in the first edition of my textbook (Ladefoged, 1975) there was no mention of epiglottal sounds, but in the second edition (Ladefoged 1982) they are described along with those made at other places of articulation. Phonologists working in standard feature theory also feel at liberty to add features as convenient.

This kind of behavior should be worrisome, because it makes it very plain that there is no theoretical basis for the existing phonetic framework. If we can simply add to it or modify it in some way whenever we find a new contrast in a previously uninvestigated language, then we can hardly say that the framework specifies “the phonetic capabilities of man” (Chomsky and Halle, 1968). All it specifies is the set of contrasts that have been observed in languages to date. In other words, it has a functional linguistic basis rather than an *a priori* basis dependent on our articulatory, auditory and cognitive capabilities.

An alternative proposal that has been made by Catford (1977a), Lindblom (1984) and others is that the set of speech sounds to be used in our descriptions is, in fact the *complete* set of sounds that can be made with the human vocal apparatus. From a linguistic point of view this is not a very fruitful suggestion. As Pike (1943) pointed out, there are numerous marginal speech sounds — clapping the teeth together, talking with a pulmonic ingressive airstream — so many that it seems profitless to try to classify them. Listing every sound that that we can possibly make will not help us find out what languages are like. What we want to know is the set of sounds that a language could use — those that are not too difficult to produce in running

speech and that are sufficiently different from one another to be clearly distinguishable.

Another notion is to argue that there *is* a well determined set of phonetic possibilities and that we are innately endowed with these phonetic capabilities. The problem then becomes one of trying to find out what these innate capabilities are. But this does not seem to be a legitimate position; there is no way of making an independent determination of these capabilities. It is not like saying that there are such things as chemical elements, even though, at a particular moment in history, we might not have discovered the complete set of elements. In the chemical case, if we find something lying around, there are tests that we can do to decide whether it is an element or not. We can predict (if we are as bright as Madame Curie) when we find an apparent gap in the Periodic Table, that there ought to be a substance (which we might call radium) that could be discovered. In the case of speech sounds there is nothing comparable that can be done. There are lots of things available for our inspection — whistling, burping, ventricular phonation — but there is no way of telling whether they are part of human phonetic capabilities in a linguistic sense without waiting to observe them in a language. Nor is even this a foolproof test. They might have been phonological features in a language spoken until yesterday in Outer Yucca. Hockett (1955) is correct in saying that there is no principled way in which we can determine the limits of the phonetic framework. As a result, it seems uninteresting to talk about our being innately endowed with a set of phonetic capabilities. This notion is largely untestable. The best that we can do is to test whether the observed phonetic possibilities are in fact things that people use when speaking particular languages.

It therefore seems that, whether we like it or not, our phonetic framework must be set up simply on the basis of observations of phonological contrasts. But these observations will enable us to say only when two sounds in the same language are different; they will not help us to decide whether two sounds are the same or not when they occur in different languages. It's all very well for the IPA to say "When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound." What is meant by "very similar shades of sound."

There is no easy way of saying whether two sounds should be considered to be the same or not. No two sounds in different languages are ever absolutely the same, if only because they are usually spoken by different speakers. If we imagine that they have been spoken by a perfect bilingual speaker of the languages in question, our problem is to decide whether the two sounds are the same or whether they are sufficiently different so that, if they had occurred in one language they could have distinguished words. For example, the high front vowel [i] that occurs in French *si* is different from that in English *see*, but is the difference sufficient to distinguish words? Almost certainly not. Speech sounds must be more than just noticeably different to be used contrastively in a language. Listeners have to be able to distinguish words reliably in comparatively noisy surroundings. Languages lose contrasts that are not easily heard.

At the moment we have no way of giving a quanititative account of how far apart two sounds must be for them to distinguish words in a language. There is work on the just noticeable differences between speech sounds, but little informed work on the magnitude of the difference necessary to be viable in the rough and tumble of

everyday language use. In calculating how similar but different speech sounds might be, the best we can do is to consider languages that appear to outsiders to have sounds that are hard to distinguish. Thus we can make a conservative estimate of the number of possible vowel distinctions by reference to phonological observations in a number of languages. There are three major parameters of vowel quality: height, backness and rounding. Thinking just of how many vowel heights there are, we can note German and Swedish vowels with four vowel heights. But there are also some Germanic dialects, such as the Dutch dialect of Weert (Heijmans and Gussenhoven,1998) that have 5 vowel heights, at least among front vowels. Now consider how many degrees of backness there are. Some languages, such as Norwegian contrast front, central and back vowels, as we will see in chapter 3. The third major aspect of vowel quality to be considered is rounding. Rounding is not always a two-way opposition, with vowels being either rounded or unrounded, Ladefoged and Maddieson (1996) conclude that there are two kinds of rounding, one in which the corners of the mouth are drawn together and the lips are protruded, and the other in which the lips are simply compressed vertically, without involving any protrusion or drawing of the corners together. These two kinds of rounding are said to contrast in Swedish high front vowels (Malmberg 1951, Linker 1982).

Superficially it might look as if there might be 5 heights, 3 degrees of backness and 3 rounding possibilities, unrounded, protruded rounding and compressed rounding. This would allow for 5 x 3 x 3 = 45 contrasts. But there are many constraints on the possible combinations. There are 5 contrasting front vowels differing in height, but probably only 4 contrasting back vowels differing in this way, and only 3 height possibilities for central vowels, giving a total 5 + 4 + 3 = 12 rather than 15 possible contrasting vowels without considering rounding. Some of these vowels have not been observed as having rounding contrasts in any language. There are probably at most 3 front rounded vowels, even considering different types of rounding, one contrast in rounding among central vowels, and 3 contrasts among back vowels, adding a total of 7 more to the 12 already noted. This gives a conservative estimate of 19 vowels that could be distinguished in terms of the major features height, backness and rounding. In one sense this is far too few, as there are 28 vowel symbols on the IPA chart. But it seems unlikely that there is any language that could use all the vowels on the chart contrastively.

Now consider some secondary features of vowel quality such as nasalization. I doubt that all the 19 vowels that could contrast with each other could also have nasalized counterparts, but it is at least possible (in my opinion) that a language could have 10 contrasting nasalized vowels, although none that I know of has more than 8 (reference to come). Adding 10 to the 19 takes the total set of contrasting possibilities up to 29. Some forms of American English have 5 vowels with r-coloring, as in the words *here, there, part, port, poor* bringing us up to 34 possilities for contrasts. There are West African languages such as Degema that have 5 vowels contrasting in ATR. This make the total 39. Vowels can also vary in phonation type. In Jalapa Mazatec, there are 5 contrasting breathy voiced vowels, as well as 5 creaky voiced vowels, all contrasting with modal voiced vowels, making the total 49. !Xóõ has 3 vowels that involve another voice quality, epiglottal or strident voicing, making the total 52.

We could consider many other variations in phonation type. Some South East Asian languages such as Bruu have contrasting voice qualities that involve a degree

of tension of the vocal folds that is similar to that found in creaky voice but different in that the vocal fold are not so tense, resulting in what has been called stiff voice. Javanese has an opposite deviation from modal voice, a form of slack voice similar to breathy voice, but with the vocal folds being closer together. There are a large number of variations in voice quality used in different languages, but we will consider only those in which there is a clear contrast in glottal states. No language contrasts stiff and creaky vowels, or slack and breathy voiced vowel, so we will not add these possibilities.

So far we have been considering each of the secondary features of vowel quality simply in contrast to modal vowels. But ATR, stridency, rhoticism, creaky voice and breathy voice no doubt could interact with each other, although clear cases of such interactions are not easy to find. The only obvious examples are ATR vowels, which are contrastively nasalized in Akan (adding another 5 possibilities). I do not know of contrastively nasalized r-colored vowels but they could occur (5 more possibilities), Similarly creaky voice and nasalization is not problematic (5 more), and breathy voice and nasalization could occur (5 more). Epiglottalized nasalized vowels are also possible (adding another 3). Adding these 23 possibilities would bring the total up to 75. As there is no reason why all these vowels should not have contrasting long and short variants, this number should be doubled, to around 150. None of this takes into account how many contrasting diphthongs we should add. When we consider that the diphthongs could also be nasalized, creaky, breathy and so on, it seems that a very conservative estimate might suggest that there are at least 200 contrasting vowel possibilities.

The largest vowel inventory reported by Maddieson (1984) is in !Xu, with a total of 46 items consisting of 13 short monophthongs, 11 long monophthongs, and 22 diphthongs. This language does not have a number of the distinctions we have been discussing, such as front rounded vowels, or back unrounded vowels, or rhoticized vowels. In view of all the possibilities we have been considering, it is apparent that the limitation in the !Xu inventory size (to a mere 45 vowels) is not because the language has exhausted the phonetic possiblities available to it. Furthermore, when we consider all the possibilities it seems that Shaw (1916) may have underestimated the number of distinguishable vowels. In the play *Pygmalion* (and in the *My Fair Lady* version, Lerner and Loewe, 1956) Colonel Pickering expresses admiration for the expert phonetician, Henry Higgins, who is able to distinguish 130 different vowels, as opposed to Pickering's 24. I think a reasonable phonetician (maybe me, before I became somewhat deaf) could identify 200 different vowels.

The number of distinct consonants is also considerable, even if we limit ourselves fairly strictly to what must be called single segments (i.e. disregarding all affricates, prenasalized stops, etc, although many of them function as single phonological segments). The IPA chart has 82 symbols for consonants, without taking account of oppositions such as that between dental and alveolar stops (which contrast in many Australian languages), voiceless nasals (as in Burmese), or differences between aspirated and unaspirated obstruents (Sindhi has 25 stop consonants, only 10 of which appear as distinct symbols on the IPA chart). We must also consider all the secondary articulations, such as labialization, palatalization, velarization and pharyngealization, which would far more than double the number of possibilities. And we have to note differences in phonation type, as well as airstream mechanisms of the kind that form clicks, ejectives and implosives. A very conservative estimate

would place the total number of consonantal segments as being up in the hundreds, perhaps as high as 600, giving a total of 800 possible segments in all. A comparable number occurs in Maddieson's (1984) survey of the phonological segments that occur in 317 languages, selected so as to exemplify the range of the world's languages. He found that when he considered just the contrasting sounds, without taking into account variations in length, he had to recognize about 650 phonetically distinct segments. Probably no language uses more than about 150, (Maddieson reports !Xu as having a total of 141, consisting of the 46 vowels mentioned above plus 95 consonants, some of which may better be regarded as sequences.) A theory of phonetic representation has to specify far more possibilities, perhaps, as indicated above, around 800.

There are many reasons why languages use only a small part of the total phonetic space. Perhaps the most important is that it isn't necessary to use a large number of segments; languages can have a sufficiently large stock of words while using only a small number of segmental oppositions. Hawaian has only 13 segments, /p, k, ʔ, m, n, w, l, r, i, e, a, o, u/, but it is a perfectly viable language. (It is an endangered language, but that is because of the dominance of another culture, not because of any linguistic failings of its own.) The phonological devices used by languages do not require more than a small part of the wealth of phonetic possibilities. Nevertheless if we want a theory of phonetic representation that will account for all the sounds of the world's languages we have to be able to describe around 800 speech sounds.

1.4 Language as a social institution

We have discussed what a language is for, and the limits of what we want to consider as language, but have said very little about what a language is. One popular answer, deriving largely from the many works of Noam Chomsky, is that a language is a formal system in a speaker's mind. This is a valid view, accounting neatly for a speaker's ability to generate an infinite number of well-formed sentences from a small set of primitives. But it is not entirely satisfactory in that it takes two or more to communicate; there must be a speaker and at least one listener. And the speaker and the listener do not share a mind.

It seems to me that the mental nature of language has been somewhat misleadingly presented by Chomsky (1975). His notion that language is an organ of the mind is not very helpful. It is somewhat like saying that digestion is an organ of the body. Digestion is an ability that involves many components, including some things that are normally called organs such as the liver and the pancreas, as well as a number of other things such as saliva, mastication, and bowel movements. Digestion is like language in that it is a system. But neither of them is an organ in the usual sense.

Another way of describing a language is to consider it as an observable social institution, without having to consider what goes on in people's minds. When we consider any social institution we find that it is governed by different principles from those that govern the behavior of individuals. Principles such as communicative efficiency and identification with a group apply to a language considered as a system. A language is the result not only of individual acts but also of internal forces that make it to some extent self-organizing.

In order to make this point clear it is worth considering two other examples of self-organizing social institutions. We may begin by comparing a language with a moral code — a system of value judgments applicable in a given community. Any moral code is clearly a product of a society, and is strongly influenced by the surrounding culture. Moral judgements that originally had some utilitarian function rapidly become ritualized. Like pronunciations of words they are as they are because that is the way things are done in a certain society. But morality is also a property of an individual, at least to the extent that the individual can choose to perform moral acts. Morality, or at least the capacity for performing moral acts, may even be innate, making it similar to the capacity for learning language. Certainly one way to think of morality is as an over-developed herd instinct — a self-organized, innate drive for the preservation of the herd rather than the individual.

The moral code that we observe (or feel guilty about) is only one example of a social institution. As another, very different, example consider the economic system. There are obvious market forces affecting the price of goods and the cost of labor (the far from inexorable 'laws' of supply and demand). There are also Galbraithian forces such as the conflict between the company management (whose aim is usually growth, which leads to bigger managerial responsibility and salaries) and the company ownership (the shareholders) who want bigger profits, which may well be achieved without growth and with less management. All these forces, and many more (government, international affairs, and perhaps morality, although that seem unlikely in the present business economy) add up to form a social institution, the economic system, which nobody understands and which is certainly not part of anybody's competence. Without people there would be no economic system. It is like language in that it takes at least two to trade. Furthermore, just as people 'know' the rules of their language, in the same sense everyone 'knows' their economic system. We all have some understanding of how to buy and sell things, and what our labor is worth. But it is obviously ridiculous to take a mentalist approach. Nobody would call economics an organ of the mind.

Many sound patterns are the result of language being (like morality and economics) a self-organizing social institution. As an example, consider how patterns result from the filling of a gap in a phonological system. It has often been observed that languages tend to fill holes in the patterns of their segmental inventories. Thus Antilla (1972) shows that Proto-Baltic Finnic had a system "in which the short vowels had one degree of height more than the long ones, and contained the only front rounded vowel in the whole system....(Modern) Finnish has filled every single gap and ended up with perfect symmetry." To take another example, it is not at all surprising for a language such as English, which at one stage had four voiceless fricatives /f, θ̟, **s,** ʃ/ and three voiced ones /v, ð, z/, to acquire the missing voiced fricative [ʒ], as we have done recently. But it should be noted that this does not occur because of a biological drive acting on individual speakers and listeners. Filling holes in an abstract phonological system does nothing for the communicative efficiency of an individual. There is no increase in ease of articulation or auditory distinctiveness for any currently existing possible utterance.

Adding a new sound is like the emergence of a new species in biological evolution. If there is an ecological niche to be filled, events such as the random mutation of genes will conspire to fill it. In the same way a new phoneme is more likely to occur (to be borrowed or to be phonologized from an existing allophone), if

it fits nicely into an existing pattern. This cannot be explained in terms of the behavior of individual speakers and listeners, just as the development of a new species is not due to the action of individual members of existing species. The communicative efficiency principle does not apply to how languages organize their sounds. When we discuss phonological systems we have moved to considering language not as part of an individual's behavior, but as a self organizing institution.

Many of the patterns that we can observe in languages are patterns that occur simply in language considered as a social institution. In addition to the hole in the pattern phenomena, there are what Kisseberth (1970) called phonological conspiracies. Dauer (1983) has assembled an excellent case for regarding stress timing in English in this way. Many people have observed an apparent tendency in English for stresses to recur at regular intervals of time. But it seems that this may be due to a fortuitous combination of circumstances. Several properties conspire to make this happen. Firstly English words have a somewhat regular stress pattern; secondly there are alternative stress patterns available for some words; thirdly it is often possible to drop stresses in some sequences of words; fourthly weak syllables can have fully reduced vowels. These and other possibilities lead to sentences being formed with stresses at fairly regular intervals.

Perhaps the most startling conspiracy — one that seems to have deceived by far the majority of linguists — is the appearance of phonemes. Accounts of human behavior in terms of phonemes are nearly always examples of what has been called the psychologist's fallacy — the notion that because an act can be described in a given way that it is necessarily structured in that way. As will be shown later, phoneme size units play only a minor role in human behavioral acts such as normal speaking and listening. They are, nevertheless, great imaginary objects for use in describing linguistic aspects of speech.

Lindblom (1984) has suggested a nice analogy that can be extended to make this point clear (perhaps in ways that he might not approve of). He has pointed out that termite nests appear to the outside observer to have a most intricate structure. There are great pillars and arches that rival those of medieval cathedrals. But it does not follow from this that individual termites know about arches. In fact they are simply following a very straightforward pattern of behavior, governed (in nest building) by a single rule: deposit grains of earth near other grains of earth that are scented with a termite secretion. At first this leads to random depositing of earth. But very soon the deposits are on top of other recent deposits and the pillars grow. As two pillars grow taller the scent is strongest on the sides closest to each other; and so those two sides grow together and form an arch. All from a single, simple, rule. Phonemes may be like arches in termite nests, visible to outside observers, but having no role in the activity of the individuals producing them.

Speech appears to be composed of sequences of segments because of the interactions of the different systems of which it is composed. The complex gestures involved in producing syllables have diverse parts that look as if they are categorically distinct. We call these diverse parts vowels and consonants, but we must always remember that these are just names for readily distinguishable aspects of the stream of speech. Those of us who have been exposed to an alphabetic tradition may be influenced so that we are very conscious of the possibility of describing speech in terms of vowels and consonants. But illiterates may have little or no concept of speech segments (Morais et al, 1979). Similarly those involved in

adult literacy campaigns report that the concept of the segment is far from self evident. Intelligent adults who have been taught to write only a few words cannot readily perform tasks such as naming other words that begin with the same segment. They have to be taught to do this. It is an ability one acquires when learning to read and write proficiently. A language consultant who has been working extensively with a linguist will be able to learn the phonemic principle (Sapir 1949) just as a child can learn to read and write. But this is hardly evidence for phonemic units in the normal process of speaking and listening.

We can carry the termite analogy a step further still. The individual termite cannot be considered responsible for the design of the arches and pillars in a termite nest, but it does not follow that it is not interesting for outsiders to describe these pillars and arches. They are a necessary part of the termite nest; when a pillar or an arch is needed to support the edifice which the community requires, if the nest is to be one which survives, then the requisite structure will evolve. In that sense, a termite nest is self-organized. In a similar way a language gets the sounds that it needs to function as a viable communicative device. The segments and phonemes are present in the structure of that abstract entity, the language, considered as a social institution. They are units that are necessary for describing patterns that occur in this observable institution.

At this point it is worth considering the nature of speech errors, as they might be taken to show that segments are real aspects of speech, stored in a speaker's brain. It is quite clear that the mistakes that people make when talking are not random segmental errors. We never say *act* when we mean to say *cat*, nor do we say *dog* when we mean to say *God*, as we would be likely to do if the segments were stored separately. Instead we make errors that show that the so-called separate segments are actually stored as parts of a larger whole. One of the most common types of error is the spoonerism, the interchange of syllable initial sounds as when, for example, *Our dear old queen* becomes *Our queer old dean*. Such errors are typically confusions involving consonants at the beginning of one syllable with consonants at the beginning of another. Mistakes involving consonants in different parts of the syllable do occur, but they are no more frequent than chance substitutions of one syllable for another.

Words must be stored in our brains in some way, but it is unlikely that a word such as *cat* is stored as instructions involving the segments **k**, **æ** and **t**. The evidence from speech errors is that larger units are involved. When talking we may make the mistake of pulling out one syllable that has a great similarity to another syllable, but we seldom simply misarrange the sounds within a syllable. I've never heard anyone mispronounce *cat* as *tack*, a mistake they could quite conceivably make if the sounds existed as separate items in the brain. There is no way of describing speech errors as just a matter of producing the segments in the wrong order. All descriptions of observed speech errors must make reference to syllables.

Several years ago Halle commented, "Almost every insight gained by modern linguistics from Grimm's law to Jakobson's distinctive features depends crucially on the assumption that speech [or, in my terms, language] is a sequence of discrete entities." (Halle, 1964). But this is an assumption made for the purpose of describing phenomena. The descriptive value of segments is not evidence of their presence in the human mind.

We will consider later arguments from speech error data and from language acquisition that might be considered to be indicative of mental activity involving segment size units. The notion proposed here, that language is a social institution, should not be taken as a claim that this is the only way to view language. Studying the mental aspects of language is a fascinating, worthwhile, interesting topic. But it is a more limited approach than studying language in all its institutional glory.

Despite the value of segments as descriptive units, it seems almost certain that the phonemic principle is not part of our genetic endownment (as it surely must be for those who view it as an innate ability). The manipulation of phonemes is an acquired ability. Evolutionists teach us that such things are properties of a culture, and not of an individual's physiology. The invention (not, for me, the discovery) of the alphabet occurred far too recently for it to have become part of our DNA. Indeed, as Gould (1981) put it: *"Homo sapiens* arose at least 50,000 years ago and we have not a shred of evidence for any genetic improvement since then ... All that we have accomplished, for better or worse is a result of cultural evolution."

Our endeavors include building (like termites) social institutions such as language, morality, and economic systems. Each has, to a great extent, become its own thing, so that it is no longer entirely explicable in terms of outside forces. The evolution of language has involved its feeding upon itself, so that it must be described partly in terms of unique principles. Descriptions of language have to be partly accounts of social institutions and not just explanations of mental activities. Like termites who do not know how to build an arch, ordinary speakers and listeners do not know all the sound pattern of English. They just follow the trail provided by other people.

2. What is the purpose of a linguistic phonetic representation

2.1 The nature of phonological features.

The next question that we must consider more fully is what the phonetic representation is for. The short answer for the purposes of this book is that it is required for the phonology of a language. There are other reasons for making phonetic representations of speech, such as noting speech errors or recording the prelinguistic babblings of infants, but we will not be concerned with them here. They are part of what Abercrombie (1967) called impressionistic transcriptions rather than accounts of the formal aspects of language.

A phonology must provide a way of representing all the distinct lexical items within each language. It must also account for the patterns of sounds that occur within languages and those that occur across languages. In order to describe patterns of sounds it is necessary to arrange them into groups, by categorizing them in terms of phonological features. Phonological features therefore serve two different purposes: they group sounds together for the purposes of the formal statements that describe and elucidate phonological patterns, and they distinguish items in the lexicon.

There is no intrinsic reason why the features required for the one purpose should be the same as those required for the other. Some languages (e.g. !Xóõ, according to Trail 1985) have little or no phonology — i.e. no alternations explicable in terms of formal statements — although they obviously have lexical items that have to be given distinct representations. Conversely, there maybe phonological alternations in a language that depend on some feature that is not used to distinguish words in that language. For example, in the Papuan language Yeletnye (Maddieson, personal communication) the feature Voice is not needed to distinguish lexical items. (Hawaiian and Maori are other more wellknown languages that are similar in this respect.) But although the feature Voice is not needed lexically in Yeletnye there have to be statements that refer to it. For example, oral stops are [+voice] when surrounded by [+voice] elements, except stem initially. (Statements that could be made using other featuresm but with far less explanatory power.)

As well as making it possible to distinguish words in the lexicon and to describe the patterns of sounds in languages, there are other tasks that a theory of phonology must perform. It must explain why languages are more likely to have certain contrasting sounds and not others. Maddieson (1984) lists the contrasting segments in a wide variety of the the world’s languages, making it evident that some kinds of sounds occur more frequently. A proper theory of phonology must account for the inventories that Maddieson observes.

In order to construct a theory of phonology that will account for all possible lexical items, all observable patterns of sounds within languages, and also take into account the probabilities of different segments occurring in the languages of the world, we must consider how languages get to be the way they are. Languages are the products of speakers and listeners. The acts of speaking and listeners leave their mark on languages in different ways. One of a speaker's goals is to communicate without undue articulatory effort. Opposing this goal is the listener’s requirement that the sounds of a language are sufficiently different from each other, and that the individual sounds within sequences of sounds are easy to recognize. As Boersma

(1989) puts it: "(T)he development of any sound system aims at three goals:(1) Maximizing the ease of articulation. (2) Minimizing confusion in the vocabulary by maximizing the perceptual distinctions between words. (3) Maximizing the perceptual salience within words, i.e. maximizing the perceptual contrast between adjacent sounds in the speech chain." With regard to the last point, he reminds us of Kawasaki's (1982) notion that "sequences of acoustically similar sounds such as [wu] or [ji] are avoided in the world's languages in favour of sequences with a greater acoustical dynamic variation like [wi] or [ju]".

The balance between the conflicting forces is clearly different in different circumstances. A speaker addressing an unfamiliar audience on a complex topic may talk slowly with a careful articulation. Two close friends exchanging information will be able to talk more quickly and with a considerable reduction in the degree of precision of the articulatory movements. On some occasions, when the listener almost certainly knows the words to expect, virtually no distinctive articulations may be needed. Lovers in bed need little articulatory precision; and soldiers on the parade ground can interpret the wordless bellowing of the drill sergeant with great rapidity.

As listeners become more familiar with particular words or phrases, speakers will be able to use more articulatory assimilations. This topic has been well enough covered in the vast literature on historical phonology to need little further documentation here. We might, however, note that as well as obvious assimilations that occur in pronouncing items such as *handkerchief* and *in between* as [ˈhæŋkətʃif] and [ɪm bəˈtwin], many cases of apparent dissimilation are actually examples of economy of effort. This is the case for Grassman's law, which states that the first of two aspirated stops in a word will become deaspirated (so that, for example Indo-European *[thrikhos] becomes classical Greek [trikhos] 'hair'). This can be interpreted as dissimilation, the consonants in a word become more unlike one another. But it is also an example of economy of effort. Aspirated consonants are perceived as being very distinct from all other sounds (Singh and Black, 1966); but they are also costly in that they use considerable respiratory energy. A word with two such sounds is very costly, and an obvious candidate for pruning in any attempt to reduce the overall effort required for an utterance. Ohala (1981) has given a convincing account of dissimilation in terms of the listener as the source of sound change. I would only add to his account that in Grassman's law the conditions were ripe for it to occur because it satisfies the biological drive for economy of communication.

The notions of auditory distinctiveness (both of the sounds in the lexicon and those in sequences) and economy of articulatory effort are important in the formation of phonological patterns. Because sounds that were produced in similar ways usually have a similar acoustic structure, features that group sounds in articulatory terms will usually also group them in acoustic terms. But the reverse is not always true; some segments can sound very similar although they were produced in quite different ways. There are thus two different kinds of phonological features: those that have an auditory basis, and those that group sounds in articulatory terms.

We must also note that some phonological patterns are the result of historical processes that may, or may not, be in any sense known to speakers of the language. For example, consider the pairs of words *bomb vs. bombardier, iamb vs. iambic, paradigm vs. paradigmatic.* It is clear that when adjacent underlying oral or nasal

non-coronal stops occur at the end of a word, one of them is deleted. The same is true at the beginning of a word, in this case applying to all stops, coronal or not: *mnemonic vs. amnesia, gnostic vs. agnostic, pterygoid vs. helicopter.* This is a pattern in the language, considered as a social institution.

There are other examples of patterns arising from historical processes that were, when they occurred, due to articulatory or auditory aspects of the sounds in question, but which now cannot be explained in terms of how they are currently heard or produced. These patterns have to be described in terms of *ad hoc* features that have neither an auditory/acoustic nor a physiological basis. An example of this kind of phonological description occurs whenever the vowels of English are described in terms of the feature Tense. This is a notoriously difficult feature to define. It does not have a unique defining articulatory or acoustic property; but it nevertheless specifies a very real grouping that has to be considered part of the sound pattern of English.

Another kind of phonology might prohibit the use of *ad hoc* features that had no physical basis on the grounds that a phonology of this kind is too unconstrained, and too difficult to learn. From the point of view outlined in the first chapter, this is irrelevant in that a phonology has to describe all the patterns in the language considered as a social institution. The feature Tense functions like *i,* the square root of –1, in mathematical formulae. It is a useful concept that has no physical reality but has great value in explaining the way things work.

It should be noted that any non-physical arbitrary feature that simply groups segments as a result of diverse historical events in a particular language or group of languages is not part of a universal phonetic feature set. When we are concerned with the general nature of human language, then we must consider only phonological descriptions made in terms of phonetic features that have a physical basis that could apply to all languages. Comparative phonologies that are restricted in this way are, of necessity, different from those of a single language in which *ad hoc* features might be permitted.

2.2 The need for auditory/acoustic features

As we have noted, the physical basis for phonological features may be physiological or acoustic. There are several important natural classes that are the result of sounds having an acoustic structure such that they have certain auditory properties in common. It is somewhat ironic that this great insight of the Prague school, much touted by Jakobson, Fant and Halle (1951), should now be overlooked by the phonologists who are their successors. The present situation arises partly because of the view of phonology presented by Chomsky and Halle (1968) in *The Sound Pattern of English* (henceforth SPE), in which features are considered to be mental entities. From this point of view it is just a matter of exposition as to whether features are defined in articulatory or acoustic terms.

It is, of course, true that all features have both articulatory and acoustic properties. As soon as a feature has been defined in terms of either acoustic or articulatory properties it can be regarded as a linguistic unit that characterizes the lexical items of a language. These lexical items have to be capable of being both spoken and heard; and therefore the features that characterize them must have both kinds of properties. But, as noted by Lieberman (1970), it does not follow from this that we should consider the linguistic function of a feature as being *required* in both

domains or that we can define it equally well in either way. The salient characteristics of the feature Coronal is in the articulatory domain, in that [+coronal] implies an articulation in the coronal region. The salient characteristics of the feature Sibilant is that [+sibilant] sounds have considerable aperiodic acoustic energy above 3,000 Hz, with no particular correlates in the articulatory domain. The feature Voice is a good example of a feature that has both kinds of characteristics. It can be defined both in physiological terms (having regular vibrations of the vocal folds) and in the acoustic domain (having an acoustic structure with well defined harmonics).

The main point in this chapter is to show that there are these two kinds of features. As nobody seems to doubt that some, if not all, features can have an articulatory basis, the discussion of these features will be left to a later chapter. Table 2.1 lists a number of features that have specifically auditory rather than articulatory correlates.

Table 2.1 A set of features that determine auditorily based natural classes without having necessary articulatory correlates.

ROOT	FEATURE	named values	*Traditional term*	Approx acoustic definition
AUDITORY	HEIGHT	high	*high (close)*	Frequency of F1
		mid-high	*mid-high (half-close)*	
		mid	*mid*	
		mid-low	*mid-low (half-open)*	
		low	*low (open)*	
	BRIGHTNESS	bright	*front unrounded*	Freq. of (F2' – F1)
		neutral	*neutral*	
		dark	*back rounded*	
	GRAVE	+grave	*grave*	Mid freq. noise
		–grave	*–grave*	
	SIBILANT	+sibilant	*strident / sibilant*	High freq. noise
		–sibilant		
	SONORANT	+sonorant	*sonorant*	Formant structure
		–sonorant		
	RHOTIC	+rhotic	*r-colored*	F3 lowering
		–rhotic		

The most outstanding features of the auditory type are properties of vowels. A problem that arises in discussing these features is that it has not been generally recognized that vowels have both articulatory *and* auditory properties. What may (loosely speaking) be called the height and backness of the body of the tongue are important considerations in accounting for articulatory interactions between consonants and vowels. Descriptions of these interactions will require us to use

features that actually specify articulations.As we will see in Chapter 4, there is no problem interpeting the auditory features of vowel quality in articulatory terms so that they can be used in this way.

We are not used to thinking of vowels simply as sounds to be judged in terms of auditory qualities. Our difficulties are further compounded by the fact that the auditory features operate in terms of multivalued scales, rather than binary oppositions, as is the case with most other features. We all realize that the notes on a piano are arranged on a scale going from low to high. But most people find difficulty in thinking of any properties other than musical scales on which sounds can be ordered; loudness is the only other auditory property that is generally recognized. Nevertheless it is a fact that some pairs of sounds - [i] and [e] for example - are more alike than others - for example [i] and [a]. Accordingly there must be some property or properties of vowel sounds that enable listeners to make these judgements. These properties may, of course, be in the articulatory domain. It may be that listeners are simply referring their auditory impressions to their tacit knowledge of the articulations required to produce vowels, as suggested by the motor theory of speech perception (Liberman and Mattingley, 1985). But before we assume that this is the case, we should consider the auditory judgments that people make, and see whether they are in fact equivalent to their articulatory gestures.

A necessary first step in describing how the vowels [i e a o u] differ from one another in auditory terms is to decide how many auditory properties are involved. It turns out that from an auditory point of view, the five vowels [i e a o u] differ in terms of two major auditory properties. If we ask observers to listen to pairs of vowels from this set, and to say which pair of vowels is furthest apart, which is most alike, and so on, we can obtain measures of the auditory distances between each of them. XXXX (1977), who performed an experiment of this sort on a larger set of vowels, found that the auditory distances between [i e a o u] could not be represented by putting all these vowels at certain distances apart on a straight line (i.e. in one dimension), but could be properly represented on a two dimensional graph. In other words, there is no single auditory property distinguishing these vowels, but they can be described in terms of two distinct auditory properties.

We may call these two auditory properties Height and Brightness. Height is a well known term for describing an aspect of vowel quality, but again I would like to emphasize that here it is being regarded as simply an auditory property. As such, it cannot be defined in anything other than impressionistic terms reflecting a listener's judgments. It is like pitch, for which the only definition offered by the Acoustical Society of America is that it is that property of a sound that enables it to be ordered on a scale going from low to high. In the case of vowel Height the definition is that it is the property that enables a vowel to be ordered on an auditory scale of vowel height going from low to high. But just as pitch can be related to fundamental frequency, so can vowel Height be related to measurable acoustic parameters. The major acoustic correlate of the auditory property vowel Height is the frequency of the first formant (F1).

The other auditory property has been termed Brightness (*Helligkeit*) (Trubetzkoy 1929, 1939, Fischer-Jorgensen 1985). The acoustic correlates of Brightness are not fully established (it is another topic on which more research remains to be done), but we take it for the moment that it is a function of F2. Brightness is a combination of all three traditional vowel features, Highness, Backness, and Round. High front

unrounded vowels have the highest value of Brightness, low back neutral vowels have a mid value and high back rounded vowels have the lowest value.

The explanatory power of the two auditory features for vowels can be exemplified by reference to notions of vowel raising, as will be discussed in the next section. Here we will consider a different point, namely, the way in which the auditory aspects of vowels can explain the dominance of the five vowel system [i e a o u]. From an auditory point of view, the fact that these particular vowels are selected is not surprising. As suggested by Ladefoged (1971:76) there is probably "some kind of principle of maximum distinctiveness whereby the auditory differences between the vowels in a language tend to be kept at a maximum." Liljencrants and Lindblom (1972) showed that if the auditory properties in question depend on the first two formants, then the most likely 5 vowel set would be [i e a o u]. Further work (summarized by Lindblom 1986) confirmed this view.

Because this is an important finding it is worth considering a slightly different version of the Liljenkrants and Linblom experiment. The first step in calculating how vowels are dispersed in an auditory space is to determine the boundaries of this space. Using a computer model of the vocal tract (Ladefoged & Lindau1989), we can calculate the formant frequencies of different tongue positions combined with different degrees of lip rounding. A set of 147 tongue shapes varied throughout the range of possibilities was combined with seven degrees of lip rounding and the corresponding formants calculated. A plot of the first two formants of these 1029 vowels is shown in figure 2.1.

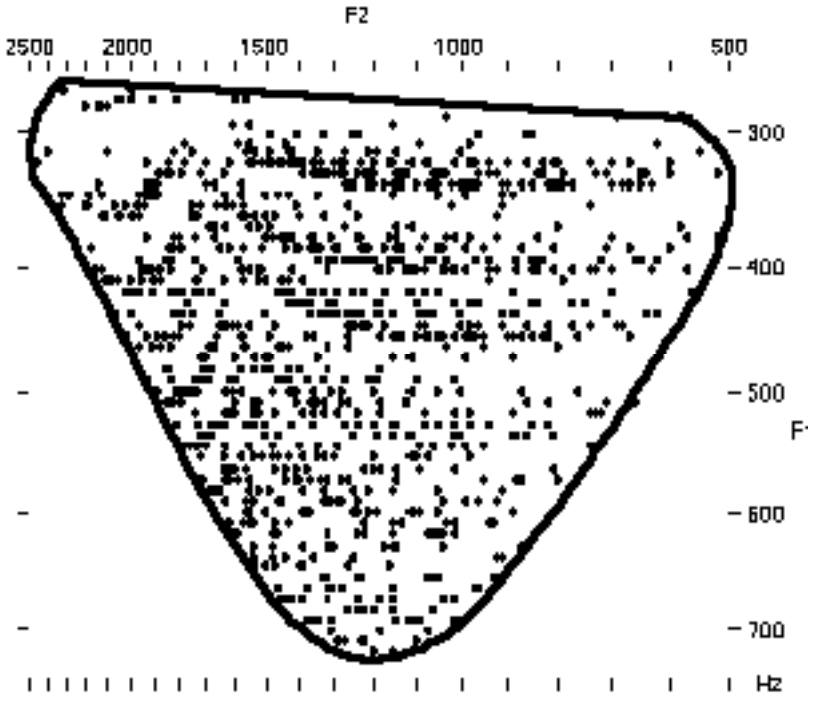

Figure 2.1. A formant plot of the first two formants of 1029 vocal tract shapes described in the text, and a curve showing the boundaries of the formant space for vowels.

A curve has been drawn round the outside of the 1029 points representing the boundaries of the formant space for vowels. The computer model is that of an average male speaker. Anyone with the same vocal tract characteristics as those

built into the model could not produce a vowel with formant frequencies outside the bounded area.

The shape of the vowel space in Figure 2.1 may not be an appropriate representation of the vowel space from a listener's point of view. The formant frequencies have been plotted in a way that may providde a good acoustic representations of vowel quality. A bark scale that reflects listener's judgments of pure tones has been used, making what appear to be equal difference in the pitch of pure tones equal distances apart. In addition the scale for F2 has been compressed in relation to that of F1. The distance between frequencies on the F2 scale is one third that of the same frequencies on the F1 scale. This has been done because F2 has a lower amplitude than F1, and therefore may be less important.

Languages as diverse as Swahili, Spanish, and Hawaiian have five vowels, with qualities something like [i e a o u]. These and only these vowels are used by approximately 20% of the world's languages (Maddieson 1984). On the basis of the material reviewed in the last chapter, we can very conservatively estimate that there are at least 50 distinct vowels (neglecting differences in phonation type, etc) that languages might have chosen as the vocalic elements of their segmental inventories. The likelihood of the same five being chosen so frequently is therefore comparable with the likelihood of playing poker and finding that one hand in five always had the Ace, King, Queen, Jack, Ten of Spades. It is therefore an absolutely astounding fact that so many languages have the vowels /i,e,a,o,u/, and any theory of phonology that does not offer an explanation for this fact must be considered to be seriously lacking.

We can now see a two part explanation for this fact. The first part suggests that the optimal number of vowels that a language needs to have is in the neighborhood of five. Three vowels (or even less) might be sufficient for conveying an appropriate number of distinctions among possible lexical items, but we may hypothesize that this is not optimum, and puts some strain on the system. More than five vowels, on this hypothesis, may be more than needed for optimal communicative efficiency. Given that five vowels is the modal number then the second part of the proposal is that the distinctiveness principle applies to the auditory features, making the most favored set [i e a o u].

A word of caution is necessary at this point. Although the most favored set of vowels can be represnted by [i e a o u], the precise values of these symbols are not the same in all five vowel languages. We can get

From an articulatory point of view, there is no reason why front unrounded and back rounded vowels should be more common than the reverse combinations. From an auditory point of view, the choice of these particular vowels is part of the tendency among languages to maximize perceptual contrasts. Phonologists who regard all features as having only articulatory definitions have no explanation for the remarkable facts of vowel distribution. But it should be emphasized that this does not mean that articulatory features do not have a role in the description of vowels. The action of the body of the tongue in the production of a vowel is specifiable in terms of physiological features that are also applicable to consonants (and thus show the relations between vowels and consonants). There should be no doubt that in order to form the correct phonological classes of vowels these sounds have to be characterized in both physiological and auditory terms.

One of the oldest established auditory features is Grave, which groups some labial and velar sounds in accordance with their spectral characteristics. Sounds such as [p, k, f, x] are produced in very different ways, but they sound similar because they have a comparatively large amount of aperiodic acoustic energy in the lower part of the spectrum. This similarity is reflected in historical changes such as English [x] to [f] in words such as *rough, tough*, or the parallel phenomena in Danish whereby *lugt* becomes *luft* (Basbol 1974). These changes are completely inexplicable in articulatory terms. Phonological alternations involving the feature Grave have been described by Hyman (1975) who shows that different vowel allophones occur before Grave consonants in Fe'fe'. Other cases of phonological alternations include spirantization in Hebrew triggered by the Grave consonants /p, b, k/.

Chomsky and Halle discarded the feature Grave because they found it did not provide a satisfactory basis for characterizing differences in place of articulation. This is undoubtedly true; from an articulatory point of view the feature Grave does not distinguish the appropriate natural classes. But this does not mean that it fails to characterize a natural class of sounds from an auditory point of view. Throwing out Grave just because it does not have a useful articulatory correlate is as bad as it would be to throw out Nasal just because it does not have a unified set of acoustic correlates that form a basis for a natural class.

Note that the feature Grave as defined in Table 2.1 is not exactly the same as the feature proposed by Jakobson, Fant and Halle 1951). Their definition was "the predominance of one side of the significant part of the spectrum over the other." It was intended to include both consonants and vowels. The feature Grave as defined here is in practice restricted to obstruents (and, perhaps, voiceless approximants) because it stipulates that the auditory characteristic of a Grave sound is that there is salient *aperiodic* energy in the lower part of the spectrum. In speech, this type of energy occurs only in stop bursts and fricatives (and, perhaps, a voiceless labial-velar approximant). There is no auditory property of this sort that links particular vowels with particular consonants. (But, as we will see later, there are links between particular vowels and consonants specified by certain articulatory features.)

Another auditory feature that is of importance in grouping consonants I have called Sibilant, following the traditional phonetic usage. It is not exactly equivalent to the Jakobsonian feature Strident as in that system [f, v] were called [+strident] in order to distinguish them from [ɸ, β]. This produced the rather unnatural class of strident sounds [f, v, s, z, ʃ, ʒ], a fact recognized by Chomsky and Halle (1968), who called [f, v] [–strident] and devised a different way of distinguishing [ɸ, β]. So as to make it plain that I am using a traditional definition, I have retained the older term Sibilant rather than Strident. Sibilant has been used for centuries (e.g. by Holder, 1669, and many phoneticians after him) to identify the class of sounds [s, z, ʃ, (ʒ)]. (Holder did not recognize [ʒ], a distinctive sound only in more recent English).

It is interesting to consider whether it might be possible to give an articulatory definition of this feature, in that Sibilant sounds are always pronounced with the jaw raised so that there is a narrow gap between the upper and lower front teeth. The high frequency aperiodic acoustic energy that gives rise to the auditory characteristics of this feature is due to the jet of air striking this narrow gap (Catford 1977a, Shadle 1985). There are two objections to this articulatory definition of

Sibilant. Firstly, other sounds such as [i], also have considerable jaw raising. Secondly, it is unlikely that this articulatory attribute is the reason for their acting together in historical changes and morphological alternations. There is no evidence showing that jaw position is a salient characteristic of sounds causing them to be grouped together, whereas the auditory grouping of these sounds is evident in the perceptual confusion data of Miller and Nicely (1955) and its reanalysis by Shepard (1972), as well as in the perceptual similarity judgments reported by Ingram (1975).

The claim that Sibilant should be defined in auditory rather than acoustic terms is not a matter of whether there is or is not a feature of this kind. There is little doubt that sibilants form a natural class of sounds that act together in phonological rules. Nor is it a matter of formal evaluation of rules. Given that there is a feature Sibilant the system for evaluating its use within a phonology will be the same irrespective of its phonetic attributes. What is at stake is whether the auditory definition provides a better explanation for the grouping than a definition in terms of the articulatory attributes. Until there is some evidence for the shared articulatory properties being the reason for this grouping, the well attested salient auditory characteristics are clearly the basis for the natural class of Sibilants.

The next auditory feature listed in Table 2.1 is Sonorant, a feature defined here in a slightly different way from the definition in SPE. In many languages sounds such as [m, n, l, r] act together as a class. For instance, in American English these sounds are syllabic after a stop or a fricative, as at the ends of the words *prism, hidden, table, razor*, but not after after other sonorants as in *film, kiln*. The feature Sonorant is hard to define meaningfully in articulatory terms. The notion 'spontaneous voicing' (Chomsky and Halle, 1968) does not get at the essence of what it is that causes these sounds to be grouped together. What matters is that they are consonants that have a periodic, well-defined, formant structure and are comparatively steady state. They behave the same way within a language not because they are made alike, but because they sound alike.

Similar observations can be made about the feature Rhotacized, which is associated with a lowering of the frequencies of the third and fourth formants. As has been shown by Lindau (1985) many forms of *r* share this auditory characteristic, although they may have been produced by different articulatory means.

2.3 Defining characteristics of features

So far we have considered three kinds of phonological features: those that reflect some *ad hoc* grouping of sounds resulting from diverse historical processes; those that can be associated with a physiological property (which will be discussed more fully in the next chapter); and those that can be associated with an auditory property. When we are concerned with general properties of human language, then the features that reflect some *ad hoc* grouping of sounds are irrelevant. In considering phonological universals we can use only phonological features that reflect an articulatory or auditory property of some kind; and that property must be the same in all languages. There is a long and valid tradition to this effect in feature theory. Jakobson, Fant, and Halle (1951) state: “A distinctive feature cannot be identified without recourse to its specific property.” Chomsky and Halle (1968) make it quite plain that “... each feature is a physical scale defined by two points.” Later this view was extended by Stevens (1983), who noted: “Each feature is represented in the

sound wave as a unique acoustic property to which the auditory system responds in a distinctive way." This view is even stronger than Chomsky and Halle's, in that it claims that there is a unique *acoustic* correlate of each feature. But for all these authors, each segment, S, is composed of a number of features, F, each of which is related to some physical parameter, P, as sketched in Figure 2.1.

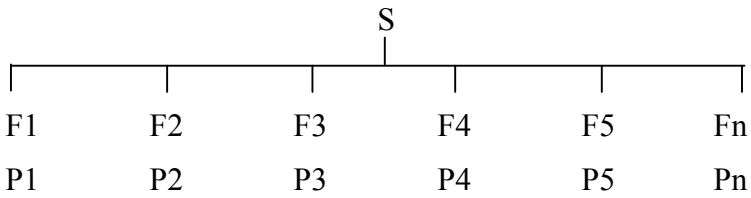

Figure 2.1. The relationship between a segment (S), features (F), and physical parameters (P) in traditional feature theories.

Stevens' view that there is unique acoustic correlate of each feature is based on the notion that the perception of speech involves specific feature detectors (Stevens and Blumstein 1975, Blumstein and Stevens 1975, Stevens 1983). There is very little evidence that this is correct. According to Stevens and Blumstein (1981): "The theory of acoustic invariance has been elaborated most completely for place of articulation in stop consonants." But, while recognizing Stevens and Blumstein's considerable achievements, their success rate in determining the correlates of place of articulation is not very high. They suggest that there is a particular relationship between the spectral burst and the formant frequencies for each place of articulation. But they find the properties proposed for detection of place of articulation are present in only about 83 % of initial stops, 75 % of final stops if exploded, and 77% of initial nasal consonants. Further work by Lahiri, Gewirth, and Blumstein (1984) led to some improvement in these percentages for particular places of articulation. By allowing the invariant acoustic properties to be both time varying and relative they were able to distinguish bilabial and dental/alveolar stops more than 91 % of the time. But this improvement in distinguishing places of articulation was applicable to only one manner of articulation (stop), and only one place in the syllable (initial). Phonological features group sounds in terms of place irrespective of whether they are stops, nasals, or fricatives, and irrespective of whether they are initial or final in the syllable. If correct detection of place of articulation could be achieved in 90% or more of both initial and final positions for stops, fricatives, and nasals, then the feature-detector hypothesis could be considered more seriously. There have been suggestions (Stevens and Blumstein 1981, Stevens 1983) of possible invariant properties for several other phonetic categories. However it is clear that there are no reports of the analysis of natural speech that show the presence of invariant acoustic properties, however complexly defined, for a wide range of phonological features.

It is always possible to argue that the failure to find an invariant acoustic property for a particular feature is due to our lack of diligence in looking for it. The claim could be that the invariance is there, but we just have not found it. This is rather like saying that part of the moon is made of green cheese. This is a very difficult claim to disprove. When we all fail to find the green cheese, it could be said that this is just because we have not looked in the right part of the moon. Of course, there are many occasions in scientific endeavours when it is appropriate to seek structures that are almost certainly present but very difficult to find. This is the situation at the moment

in genetics where people are searching for the correlates of inherited characteristics in the structure of the DNA. But the search for invariant acoustic properties of all phonological features is not like that.

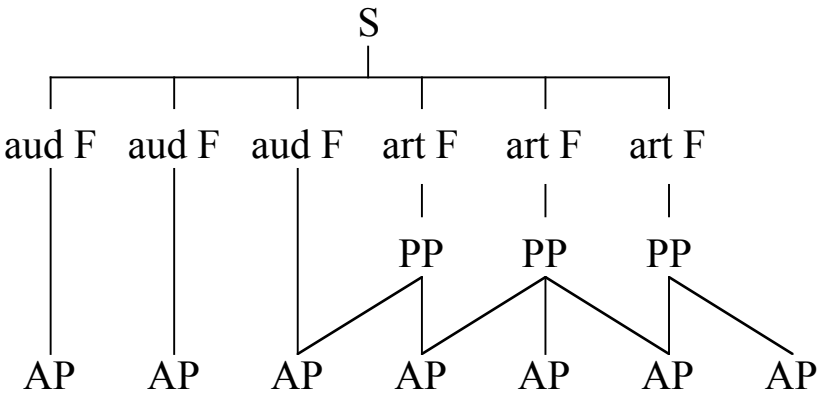

Figure 2.2. The relationship between a segment (S), auditory features (aud F), articulatory features (art F), physiological parameters (PP), and acoustic parameters (AP).

There is no reason to expect all phonological features to be in a one-to-one relation to acoustic scales. It is much more likely that the articulatorily based features will be associated with variations in many acoustic parameters. This view of phonological features is illustrated in Figure 2.2. A segment is still described in terms of features, where the features represent convenient labels for classes of sounds needed by the phonology. Some features (those that reflect auditorily based natural classes) have simple, direct acoustic correlates; but articulatory features are connected to acoustic parameters in a complex, many-to-many relationship. Thus the auditory features 1, 2 and 3 are each associated with a single acoustic property; and the articulatory features 4, 5 and 6 are each associated with a single physiological property, but the relationships between these physiological properties and acoustic poperties are complex.

There are problems in the notion that each feature should be definable in terms of a specified property ("a physical scale" as Chomsky and Halle, 1968, put it). Across languages there are often situations that are partially the same, and we would miss a significant generalization if we did not group them together. For example, there are similarities in the processes referred to as vowel harmony in African languages and register distinction in Mon Khmer languages (Gregerson 1976). These distinctions often involve several articulatory parameters that are related to one another, such as tongue-root movements, tongue height variations, vertical larynx movements, and changes in laryngeal tension (and hence in phonation type). But these factors are not all describable in terms of a single feature. We undoubtedly need some way of expressing these similarities.

One way of recognizing similarities among different features is by grouping them under a higher level feature in a feature hierarchy. As will be shown in the next chapter, we can show many parallels in the phonological processes that occur in different languages by arranging the features in a hierarchy in which the higher nodes, rather than the features themselves, may be used to refer to more general phonological processes. In this way we can avoid weakening the notion that each

terminal feature must be definable in terms of a single scale in the articulatory or the acoustic domain. Higher nodes act as cover terms that have multiple properties, but terminal features have only a single phonetic correlate.

2.4 Multi-valued features.

There is one further formal consideration affecting both articulatory and auditory features that must be clarified before we go on, in the next chapter, to complete the set of phonological features by describing the required articulatory features. This is the question of whether all features have to be binary. Implicit in the proposal for the auditory features Height and Brightness as described above is the notion that these features are not binary; each of them permits a range of values along a scale. Thus the feature Height has to have at least four possibilities, [high], [mid-high], [mid-low] and [low] to account for the vowels of many languages, such as Danish, a point that will be discussed further in the next chapter.

The process of going from [low] to [mid] on the Height scale is the same as the process of going from [mid] to [high]. Only with this notion of scalar features can we maintain the notion of a two-dimensional vowel space defined in terms of Height and Brightness. Without this notion there is little likelihood of a true explanation not only of why vowel systems are as they are, but also of many phonological processes such as the English vowel shift.

It is, of course, possible to describe vowel shift changes in terms of binary features as illustrated by Chomsky and Halle (1968). But any rule using binary features that has to account for [+high] becoming [–high] in the same circumstances as [–low] becomes [+low] inevitably misses a linguistically significant generalization. Simply by virtue of having to use two separate features it cannot show the unity of the process that is expressed by a rule of the form: [n high] -> [n+1 high]. Binary feature theories also have to have a marking convention that prohibits a vowel from being simultaneously [+high] and [+low]. All this is equivalent to arranging, in a rather cumbersome notation, that [high], [mid] and [low] form an ordered set of values on a single scale. So why not say just that?

Simple examples of vowel raising can be found in a number of languages spoken in Southern Africa. The changes are hard to state in terms of the conventional SPE features that permit the specification of only three vowel heights. In the Nguni languages there are five vowels, which, in traditional IPA terms, have the qualities [i, ɛ, a, ɔ, u]. Each of these vowels is fairly similar to the corresponding cardinal vowel, except [a], which is retracted so that it is in between cardinals [a] and [ɑ]. The unmodified qualities of these vowels in Zulu are shown by the solid points in the vowel chart on the left in figure 2.3

ˈpitə ˈlædɪfəʊgɪd rɛprɪˈsɛntɪŋ ˈlɪŋgwɪstɪk fəˈnɛtɪk ˈstrʌktʃə

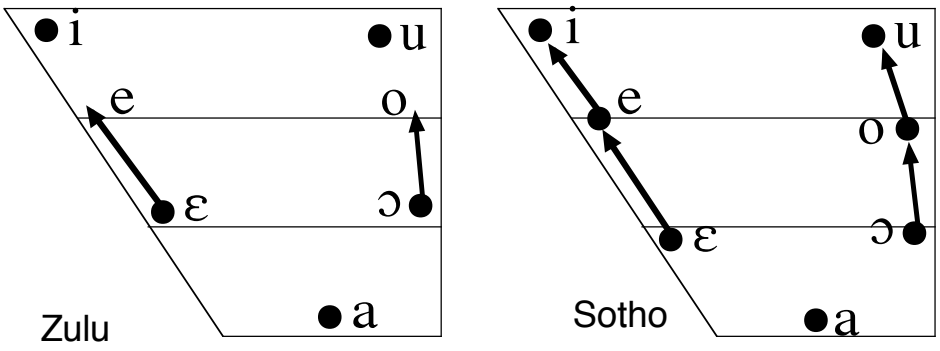

Figure 2.3. Vowel raising in Zulu and Sotho.

The two mid vowels [ɛ, ɔ] have raised variants that occur whenever the following vowel in the word is [i] or [u]. (The phonological conditions can be further elaborated, but this is sufficient for our purpose here.) These variants can be considered as distinctive phones in the terminology suggested by Ladd (2002). They have the phonetic qualities indicated by the arrow heads in the figure. Let us assume that we want to express this process as a change whereby [ɛ] goes to [e] and [ɔ] goes to [o]. We can readily express the change using a multi-valued feature of Height, with four possible values:

[mid-low] --> [mid-high] / (C) [high]

This rule is much simpler (and more explanatory) than the corresponding rule using SPE features. As the Zulu vowels and their alternants have to be described in terms of four vowel heights, [high], [mid-high], [mid-low] and [low], the SPE features High and Low would be insufficient, and it would be necessary to bring in another feature such as Tense. But this would be a purely arbitrary use of Tense; the vowels [e, o] differ from [ɛ, ɔ] in the same way, physiologically and acoustically, as [ɛ, ɔ] differ from [a].

The situation is complicated still further in languages of the Sotho group, which have seven underlying vowels [i, e, ɛ, a, ɔ, o, u], as shown in the chart on the right in figure 2.3. In these languages all four of the mid vowels [e, ɛ, ɔ, o] have raised variants in similar circumstances to the raised variants in the Nguni languages. The changes are also summarized in figure 2.3. It seems obvious that we need a multi-valued feature vowel Height to make the correct generalizations.

Further examples will be provided in the next chapter of multi-valued articulatory features in which there are at least three mutually exclusive values, arranged in an ordered series. The evidence that some phonological features are sometimes multivalued is simply overwhelming. Many are binary; but in the case of some features we need additional values both in phonological rules and in accounting for universal patterns of sounds among languages. There is no basis for the argument that all features must be binary because phonological rules have been shown to operate in terms of such features. Phonological rules have never been entirely binary; about one third of the rules in SPE — those involving Stress — use a multi-valued notation that permits scalar quantities. The fact that much of the patterns of sounds within languages can be described in binary terms does not show that it is not equally possible to use another formalism.

I hope that phonologists will soon recognize that we are due for our own little

Copernican revolution. It is possible to describe many of the observations of astronomy, and to predict future eclipses while still maintaining that the earth does not move and the sun goes round it. But as Galileo (1633) whispered after being forced to retract his Copernican heresies *Epur si muove* [still it does move]. Like Galileo, I will not go to the stake for my belief. But five vowel systems are most favored. Mid vowels are between high and low vowels. An interface between phonetics and phonology must allow some phonological features to have non-binary values.

References (for chapters 1-4, but not complete)

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Atal, Bhishnu, J. J.Chang, M. V.Mathews & J. W. Tukey. 1978. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America,* 63.1535-1555.

Bloch, Bernard & George L. Trager. 1942. *Outline of Linguistic Analysis*. Special Publication of the Linguistic Society of America. Waverly Press.

Bosch, Anna & Kenneth de Jong. 1997. The Prosody of Barra Gaelic Epenthetic Vowels. *Studies in the Linguistic Sciences* 27.1-15.

Browman, Cathe & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica*, 49, 155-180.

Browman, Cathe., Louis Goldstein, J, Kelso, Philip Rubin, & Eliot Saltzman. 1984. Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America*, 75, S22-S23 (A).

Catford, John (Ian). 1988. Phonetics of Caucasian languages. MS

Cho, Taehong & Peter Ladefoged. 1999. Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27.207-229.

Cho, Taehong, Sun-Ah Jun & Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30.193-228.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press

Clements, G. Nickerson. 1985. The geometry of phonological features. *Phonology Yearbook*, 2, 225-252.

Clements, G. Nickerson. 1986. Syllabification and epenthesis in the Barra dialect of Gaelic. *The phonological representation of suprasegmentals*, eds. K. Bogers, Harry van der Hulst, & Maarten Mous. Dordrecht, Holland: Foris. 317-336.

Duanmu, San. 2000. *The Phonology of Standard Chinese*. New York: Oxford University Press.

Esling, John. 2000. Crosslinguistic Aspects of Voice Quality" in *Voice Quality Measurement*, eds. R.D. Kent and M.J. Ball. San Diego: Singular Publishing Group, 25-35.

Esling, John & Jerold A. Edmondson. 2002. The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai. *Phonetics and its applications. Festschrift for Jens-Peter Köster on the occasion of his 60th birthday*, eds. Angelika Braun & Herbert R. Masthoff. Stuttgart: Franz Steiner Verlag. 38-51

Esling, John H., & Jimmy G. Harris .2005. States of the glottis: An articulatory phonetic model based on laryngoscopic observations. *A Figure of Speech: A Festschrift for John Laver,* eds. W. J. Hardcastle & J. Beck. Mahwah, NJ: Lawrence Erlbaum Associates. 347-383.

Fuchs, Susanne, Pascal Perrier, Christian Gengı, & Christine Mooshammer. forthcoming. What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents.*Towards a better understanding of speech production processes*, eds. Jonathan Harrington & Maria Tabain. New York: Psychology Press.

Fujimura, Osamu. 2000. The C/D Model and Prosodic Control of Articulatory Behavior. *Phonetica* 57:128-138.

Gordon, Matthew, Paul Barthmaier & Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141-174.

Gordon, Matthew, & Peter Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383-406.

Gregerson, Kenneth. J. 1976. Tongue root and register in Mon-Khmer. *Austroasiatic Studies, Part 1*, eds. P. N. Jenner, L. Thompson, & S. Starosta. Honolulu: University of Hawaii Press.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hardman, Michael J. 1966. *Jaqaru: Outline of Phonological and Morphological Structure*. (Janua Linguarum, Series Practica 22.) Mouton, The Hague.

Harshman, Richard, Peter Ladefoged, & Louis Goldstein. 1977. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America* 62.693-707.

Krane, M.H., D. J. Sinder, & James L. Flanagan. 2000. Aeroacoustic modeling of speech sound production. In *Proceedings of 5th Seminar on Speech Production: Models and Data*,eds.xxx. Kloster Seeon, Germany. 2000: 177–180

Ladefoged, Peter. 1958. Use of palatography. *Journal of Speech and Hearing Disorders* 22.764-774.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.

Ladefoged, Peter 1975. *A Course in Phonetics*. (1st edition) Orlando: Harcourt Brace.

Ladefoged, Peter 2006. *A Course in Phonetics*. (5th edition) Boston: Thomson/Wadsworth.

Ladefoged, Peter, Richard Harshman, Louis Goldstein & Lloyd Rice. 1978. Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America* 64.1027-1035.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind, & St John Skilton. 1998. Phonetic Structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28.1-.42.

Ladefoged, Peter and Mona Lindau. 1989. Modeling articulatory-acoustics relations: a comment on Stevens' 'On the quantal nature of speech.' *Journal of Phonetics* 17.99-106.

Liljencrants, Johan, & Gunnar Fant. 1975. Computer program for VT-resonance frequency calculations. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1975(4), 15-20.

Linker, Wendy. 1982. Articulatory and Acoustic Correlates of Labial Activity in Vowels: A Cross-linguistic Study (UCLA Working Papers in Phonetics 56). Ph. D. thesis. University of California, Los Angeles.

Maddieson, Ian. 1989. Linguo-labials. *VICAL 1: Oceanic Linguistics, Part 2*. eds, R. Harlow & R. Cooper. Linguistic Society of New Zealand, Auckland. 349-376.

Malmberg, Bertil. 1951. *Svensk Fonetik*. C. W. K. Gleerup, Lund.

McGowan, Richard. 1992. Tongue-tip trills and vocal-tract wall compliance. *Journal of the Acoustical Society of America* 91.2903-2910.

McGowan, Richard. 1993. The quasi-steady approximation, boundary-layer separation, and vibrations in the vocal tract. *Journal of the Acoustical Society of America* 93.2295.

Rubin, Philip, Tom Baer, & Paul Mermelstein. 1981. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America. 70,* 321-328.

Saltzman, Elliot. 1986. Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series 15*, Eds . H. Heuer & C. Fromm, New York: Springer-Verlag. 129-144.

Saltzman, Elliot (in press). Dynamics and coordinate systems in skilled sensorimotor activity. *Mind as motion*, Eds. R. Port & T. Van Gelder. Cambridge, MA: MIT Press.

Scully, Celia. 1990. Articulatory synthesis. *Speech Production and Speech Modelling*, eds. William J. Hardcastle & Alain Marchal. Dordrecht, The Netherlands: Kluwer Academic Publishers. 1990. 151–186.

Shadle, Christine H. & Robert I. Damper. 2001. Prospects for articulatory synthesis: A position paper. *Proceedings of the 4th ISCA Tutorial and Research Workshop onSpeech Synthesis (Blair Atholl, UK)*, pages 121–126.

Sinder, D. J., M. H. Krane, and James L. Flanagan. 1998. Synthesis of fricative sounds using an aeroacoustic noise generation model. *Proceedings of 16th International Congress Acoustics*, Seattle, 249–250.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30. 655–688.

Stone, Maureen, Melissa Epstein & Khalil Iskarous. 2004. Functional segments in tongue movement. *Clinical Linguistics and Phonetics 18(16-18), 507-522.*

Trager, George., & Henry Lee Smith. 1951. *An outline of English structure* (Studies in Linguistics: Occasional Papers 3). Norman, OK: Battenburg Press.

Vanvik, Arne. 1973. A phonetic-phonemic analysis of Standard Eastern Norwegian. *Norwegian Journal of Linguistics* 27: 130-39.

Wedekind, K. (1976). Kolá, Tpalá, ou 'Kpalá? Notes sur les occlusives double dans la langue Daba-Mousgoy (Kola) du Nord-Caneroun. *African Languages/Langues Africaines*, 2, 91-104.