Assuming this figure does in fact represent a rough upper limit to man's ability to ingest information, he might allot his capacity in various ways. For example, if a speaker were rapidly uttering random equiprobable phonemes, a listener might require all of his processing ability to receive correctly the written equivalent of the distinctive speech sounds. Little capacity might remain for perceiving other features of the speech such as stress, inflection, nasality, timing and other attributes of the particular voice. On the other hand, if the speech were idle social conversation, with far-reaching statistical constraints and high redundancy, the listener could direct more of his capacity to analyzing personal characteristics and articulatory peculiarities.

In protracted conversation the constraints of the language and the reasonably efficient human memory usually enable a listener to switch between a decoding of phonemic content and an observation of personal traits. Prosodic information can be assimilated along with phonemic features which relate directly to the written equivalent of the spoken information. The latter are customarily identified with speech intelligibility, while the former is loosely associated with speech quality. Intelligibility is conventionally quantified in terms of articulation scores and rates of receiving the written-equivalent information. Speech quality, as yet, has little basis for quantification. Until both intelligibility and quality can be suitably defined, the fidelity criteria for estimating speech information rates will not be firmly established.

1.4. Analysis-Synthesis Telephony: An Approach to Improved Efficiency

Despite the equivocal aspects surrounding estimates of human channel capacity and speech information rates, it is clear that a mismatch exists between the capacity of the conventional voice channel and the information rate of the source feeding it. One approach toward improving the match is to incorporate into the transmission system as many as possible of the constraints characterizing the production and perception of speech. This information, built into the communication link, is information that need not be transmitted. Another way of viewing the situation is that the channel, so constrained, confines the message ensemble to the sounds of speech. In general, no other sounds will be transmitted with acceptable fidelity.

Having built into the system constraints appropriate to production and perception, communication is effected by signalling certain parameters of the constraints. The nature of the incorporated constraints therefore influences the form of coding for the speech information. Assume, for example, that the limitations on the mechanical movements of the vocal tract are to be taken into account in the transmission system. One receiving device for realizing such restrictions might be a mechanical or electrical analog of the vocal mechanism. Speech information might then be coded and transmitted in terms of tract dimensions, deformations, and properties of vocal excitation.

Voice communication systems in which a conscious effort is made to improve efficiency by constraining the facility according to the characteristics of speech and hearing are customarily referred to as *analysis-synthesis systems*. The term is often taken as synonymous with speech-compression or band-saving systems. A main purpose of this monograph is to set forth the fundamental properties of speech and hearing which relate to communication systems of the analysis-synthesis type. A further purpose is to outline techniques for utilizing the properties of speech and hearing in practical transmission systems.

In attending to these objectives, the physiological and acoustical properties of the human vocal apparatus will be considered first. Next, the fundamental principles of the hearing mechanism will be examined. These basic expositions will then be followed by topics in speech analysis, speech synthesis, and speech perception. The final discussion will center on application of the preceding results to realizable speech-coding systems.

II. The Mechanism of Speech Production

2.1. Physiology of the Vocal Apparatus

Speech is the acoustic end product of voluntary, formalized motions of the respiratory and masticatory apparatus. It is a motor behavior which must be learned. It is developed, controlled and maintained by the acoustic feedback of the hearing mechanism and by the kinesthetic feedback of the speech musculature. Information from these senses is organized and coordinated by the central nervous system and used to direct the speech function. Impairment of either control mechanism usually degrades the performance of the vocal apparatus[1].

The speech apparatus also subserves the more fundamental processes of breathing and eating. It has been conjectured that speech evolved when ancient man discovered he could supplement his communicative hand signals with related "gestures" of his vocal tract. Sir Richard Paget sums up this speculation quite neatly. "What drove man to the

[1] Most of us are aware of the difficulties that partially or totally deaf persons have in producing adequate speech. Even more familiar, perhaps, are the temporary difficulties in articulation experienced after the dentist desensitizes a large mouth area by an injection of anesthetic.

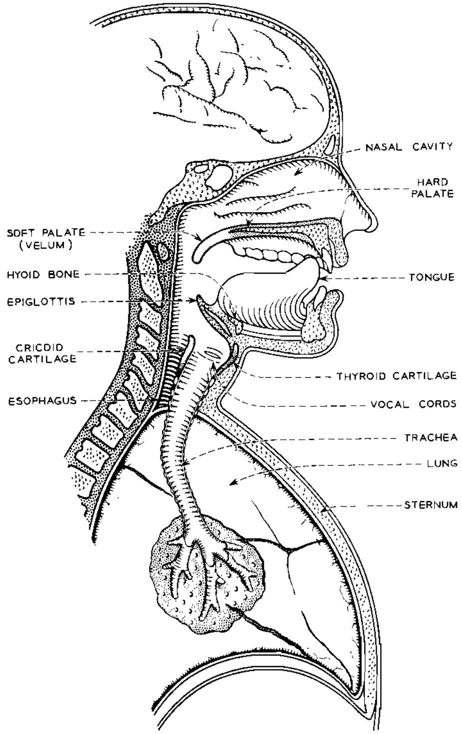

Fig. 2.1. Schematic diagram of the human vocal mechanism

invention of speech was, as I imagine, not so much the need of expressing his thoughts (for that might have been done quite satisfactorily by bodily gesture) as the difficulty of 'talking with his hands full'. It was the continual use of man's hands for craftsmanship, the chase, and the beginnings of art and agriculture, that drove him to find other methods of expressing his ideas—namely, by a specialized pantomime of the tongue and lips."

The machinery involved in speech production is shown schematically in Fig. 2.1. The diagram represents a mid-sagittal section through the vocal tract of an adult. The primary function of inhalation is accomplished by expanding the rib cage, reducing the air pressure in the lungs, and drawing air into the lungs via nostrils, nasal cavity, velum port and trachea (windpipe). Air is normally expelled by the same route. In eating, mastication takes place in the oral cavity. When food is swallowed the structures at the entrance to the trachea are drawn up under the epiglottis. The latter shields the opening at the vocal cords and prevents

food from going into the windpipe. The esophagus, which normally lies collapsed against the back wall of the throat, is at the same time drawn open to provide a passage to the stomach.

The vocal tract proper is an acoustical tube which is nonuniform in cross-sectional area. It is terminated by the lips at one end and by the vocal cord constriction at the top of the trachea at the other end. In an adult male the vocal tube is about 17 cm long and is deformed in cross-sectional area by movement of the articulators; namely, the lips, jaw, tongue and velum. The cross-sectional area in the forward portion of the tract can be varied from zero (i.e., complete closure) to upwards of 20 cm^2.

The nasal tract constitutes an ancillary path for sound transmission. It begins at the velum and terminates at the nostrils. In the adult male the cavity has a length of about 12 cm and a volume on the order of 60 cc. It is partitioned over part of its front-to-back extent by the nasal septum. Acoustic coupling between the nasal and vocal tracts is controlled by the size of the opening at the velum. In Fig. 2.1 the velum is shown widely open. In such a case, sound may be radiated from both the mouth and nostrils. In general, nasal coupling can substantially influence the character of sound radiated from the mouth. For the production of non-nasal sounds the velum is drawn tightly up and effectively seals off the entrance to the nasal cavity. In an adult male the area of the velar opening can range from zero to around 5 cm^2.

The source of energy for speech production lies in the thoracic and abdominal musculatures. Air is drawn into the lungs by enlarging the chest cavity and lowering the diaphragm. It is expelled by contracting the rib cage and increasing the lung pressure. Production of vowel sounds at the softest possible level requires a lung pressure of the order of 4 cm H_2O. For very loud, high-pitched sounds, on the other hand, pressures of about 20 cm H_2O or more are not uncommon. During speaking the lung pressure is maintained by a steady, slow contraction of the rib cage.

As air is forced from the lungs it passes through the trachea into the pharynx, or throat cavity. The top of the trachea is surmounted by a structure which is shown in additional detail in Fig. 2.2. This is the larynx. The cartilaginous frame houses two lips of ligament and muscle. These are the vocal cords and are denoted VC. The slit-like orifice between the cords is called the glottis. The knobby structures, protruding upward posterior to the cords, are the arytenoid cartilages, and are labelled AC. These cartilages support the fleshy cords and facilitate adjustment of tension. The principal outside cartilages of the larynx "box" are the anterior thyroid (labelled TC in Fig. 2.2) and the posterior cricoid. Both of these can be identified in Fig. 2.1.

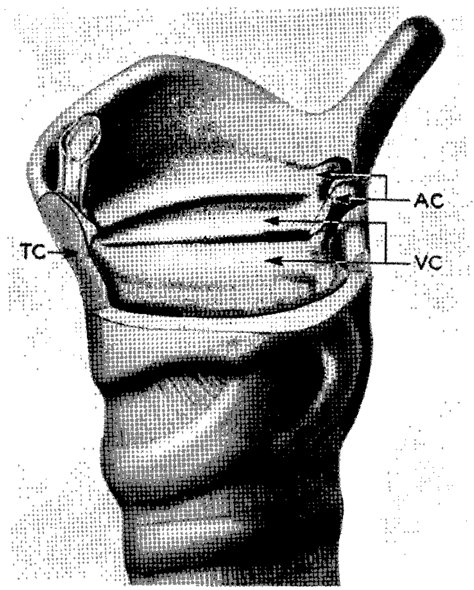

Fig. 2.2. Cut-away view of the human larynx. (After FARNSWORTH.) VC-vocal cords; AC-arytenoid cartilages; TC-thyroid cartilage

The voiced sounds of speech are produced by vibratory action of the vocal cords. Production of sound in this manner is called phonation. Qualitatively, the action proceeds in the following manner. Imagine the relatively massive, tensed vocal cords to be initially together. The subglottal pressure is increased sufficiently to force them apart with a lateral acceleration. As the air flow builds up in the orifice, the local pressure is reduced according to the Bernoulli relation, and a force acts to return the cords to a proximate position. As the cords are again drawn together the flow is diminished, and the local pressure approaches the subglottal value. The relaxation cycle is then repeated[1]. The mass and compliance of the cords, and the subglottal pressure, essentially determine the period of the oscillation. This period is generally shorter than the natural period of the cords; that is, the cords are driven in a forced oscillation.

The variable area orifice produced by the vibrating cords permits quasi-periodic pulses of air to excite the acoustic system above the vocal cords. The mechanism is somewhat similar to blowing a tone on

[1] The vibratory cycle may be started with the cords initially apart. In this case, the Bernoulli pressure first causes the cords to be drawn together. The so-called "breathy attack" is apparently produced in this manner.

a brass instrument, where the vibrating lips permit quasiperiodic pulses of air to excite the resonances of the flared horn. Over the past years the vibratory action of the vocal cords has been studied in considerable detail. Direct observations can be made by positioning a 45-degree mirror toward the back of the mouth, near the naso-pharynx. Stroboscopic illumination at the proper frequency slows or "stops" the vibratory pattern and permits detailed scrutiny.

Still more revealing and more informative is the technique of high-speed photography, pioneered by FARNSWORTH, in which moving pictures are taken at a rate of 4000 frames/sec, or higher. The technique

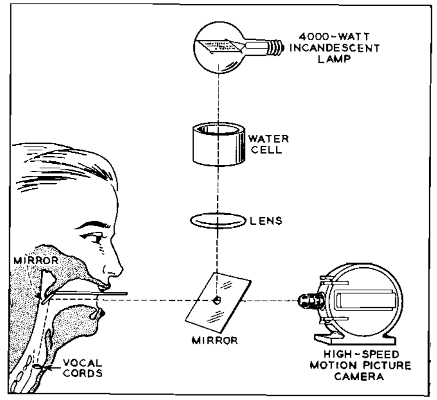

Fig. 2.3. Technique for high-speed motion picture photography of the vocal cords. (After FARNSWORTH)

is illustrated in Fig. 2.3. The cords are illuminated by an intense light source via the arrangement of lenses and mirrors shown in the diagram. Photographs are taken through an aperture in the large front mirror to avoid obstructing the illumination. The result of such photography is illustrated in Fig. 2.4. The figure shows six selected frames in one cycle of vibration of the cords of an adult male. In this case the fundamental frequency of vibration, or voice "pitch", is 125 cps.

The volume flow of air through the glottis as a function of time is similar to (though not exactly proportional to) the area of the glottal opening. For a normal voice effort and pitch, the waveform can be roughly triangular in shape and exhibit duty factors (i.e., ratios of open time to total period) commonly of the order of 0.3 to 0.7. The glottal volume current therefore has a frequency spectrum relatively rich in overtones or harmonics. Because of the approximately triangular waveform, the higher frequency components diminish in amplitude at about 12 db/octave.

The waveform of the glottal volume flow for a given individual can vary widely. In particular, it depends upon sound pitch and intensity. For low-intensity, low-pitched sounds, the subglottal pressure is low, the vocal cord duty factor high, and the amplitude of volume flow low. For high-intensity, high-pitched sounds, the subglottal pressure is large, the duty factor small and the amplitude of volume flow great. The amplitude of lateral displacement of the vocal cords, and hence the maximum glottal area, is correlated with voice intensity to a surprisingly small extent (FLETCHER). For an adult male, common peak values of glottal area are of the order of 15 mm^2.

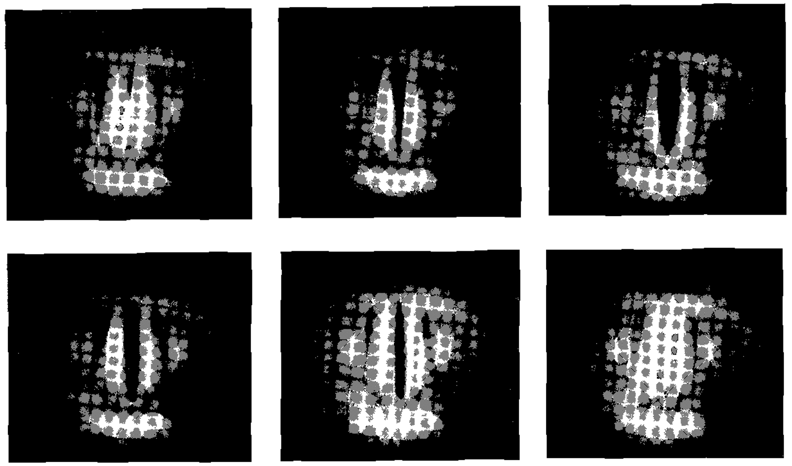

Fig. 2.4. Successive phases in one cycle of vocal cord vibration. The total elapsed time is approximately 8 msec

Because of its relatively small opening, the acoustic impedance of the glottal source is generally large compared to the acoustic impedance looking into the vocal tract, at least when the tract is not tightly constricted. Under these conditions changes in tract configuration have relatively small (but not negligible) influence upon the glottal volume flow. For tight constriction of the tract, the acoustic interaction between the tract and the vocal-cord oscillator can be pronounced.

Another source of vocal excitation is produced by a turbulent flow of air created at some point of stricture in the tract. An acoustic noise is thereby generated and provides an incoherent excitation for the vocal system. The unvoiced continuant sounds are formed from this source. Indirect measurements and theory suggest that the spectrum of the noise, at its point or region of generation, is relatively broad and uniform. The vocal cavities forward of the constriction usually are the most influential in spectrally shaping the sound.

A third source of excitation is created by a pressure buildup at some point of closure. An abrupt release of the pressure provides a transient excitation of the vocal tract. To a crude approximation the aperiodic excitation is a step function of pressure, and might be considered to have a spectrum which falls inversely with frequency. The closure can be effected at various positions toward the front of the tract; for example, at labial, dental, and palatal positions. The transient excitation can be used with or without vocal cord vibration to produce voiced or unvoiced plosive sounds.

Whispered speech is produced by substituting a noise source for the normally vibrating vocal cords. The source may by produced by turbulent flow at the partially closed glottis, or at some other constricted place in the tract.

2.2. The Sounds of Speech

To be a practicable medium for the transmission of information, a language must consist of a finite number of distinguishable, mutually-exclusive sounds. That is, the language must be constructed of basic linguistic units which have the property that if one replaces another in an utterance, the meaning is changed. The acoustic manifestations of a basic unit may vary widely. All such variations, however—when heard by a listener skilled in the language—signify the same linguistic element. This basic linguistic element is called a *phoneme* (BLOCH and TRAGER). Its manifold acoustic variations are called *allophones*.

The phonemes might therefore be looked upon as a code uniquely related to the articulatory gestures of a given language. The allophones of a given phoneme might be considered representative of the acoustic freedom permissible in specifying a code symbol. This freedom is not only dependent upon the phoneme, but also upon its position in an utterance.

The set of code symbols used in speech, and their statistical properties, depend upon the language and dialect of the communicators. When a linguist initially studies an unknown language, his first step is to make a phonetic transcription in which every perceptually-distinct sound is given a symbol. He then attempts to relate this transcription to behavior, and to determine which acoustically-distinguishable sounds belong to the same phoneme. That is, he groups together those sounds which are not distinct from each other in meaning. The sounds of each group differ in pronounciation, but this difference is not important to meaning. Their difference is merely a convention of the spoken language.

Features of speech which may be phonemically distinct in one language may not be phonemic in another. For example, in certain Chinese dialects pitch inflections are crucial in signifying distinctive speech sounds. In Western languages this generally is not the case. Another striking example is the agglutinative language of the South African Hottentots in which vocal clicks, completely foreign to Western languages, are phonemic.

The preceding implications are that speech is, in some sense, discrete. Yet an oscillographic representation of the sound pressure wave emanating from a speaker producing connected speech shows surprisingly few gaps or pause intervals. Connected speech is coupled with a near continuous motion of the vocal apparatus from sound to sound. This motion involves changes in the configuration of the vocal tract as well as in its modes of excitation. In continuous articulation the vocal tract dwells only momentarily in a state appropriate to a given phoneme.

The statistical constraints of the language greatly influence the precision with which a phoneme needs to be articulated. In some cases it is merely sufficient to make a vocal gesture in the direction of the normal configuration to signal the phoneme. Too, the relations between speech sounds and vocal motions are far from unique, although normal speakers operate with gross similarity. Notable examples of the "many-valuedness" of speech production are the compensatory articulation of ventriloquists and the mimicry of parrots and myna birds.

Despite the mutability of the vocal apparatus in connected speech, and the continuous nature of the speech wave, humans can subjectively segment speech into phonemes. Phoneticians are able to make written transcriptions of connected speech events, and phonetic alphabets have been devised for the purpose. (One of the earliest dates from the Hindus around 300 BC.) The often-accepted standard in modern times is the alphabet of the International Phonetic Association (IPA). This alphabet provides symbols for representing the speech sounds of most of the major languages of the world.

A phonetic symbol used for a phonetic transcription is conventionally enclosed in brackets [ ]. When used to indicate a phoneme, it is usually enclosed in virgules // (FAIRBANKS). In the remainder of this book the former would often be appropriate, particularly with reference to the characteristics of specific utterances. Generally, however, the broad phonetic properties of an utterance and the phoneme group to which it belongs will be of more importance. The latter notation will therefore be used exclusively to enclose all phonetic symbols.

Classification of speech sounds is customarily accomplished according to their manner and place of production. Phoneticans have found this method convenient to indicate the gross characteristics of sounds.

For example, the articulation of vowel sounds is generally described by the position of the tongue hump along the vocal tract (which is often, but not always, the place of greatest constriction) and the degree of the constriction. This classification method will be employed in the following discussion of speech sounds. The examples extend to the sounds of English speech of General American (GA) dialect.

2.21. Vowels

The vowel sounds of GA speech are normally produced exclusively by vocal cord (or voiced) excitation of the tract. In normal articulation, the tract is maintained in a relatively stable configuration during most of the sound. The vowels are further characterized by negligible (if any) nasal coupling, and by radiation only from the mouth (excepting that which passes through the cavity walls).

If the nasal tract is effectively coupled to the vocal tract during the production of a vowel, the vowel becomes nasalized. When the 12 vowels of GA speech are classified according to the tongue-hump-position/degree-of-constriction scheme, they may be arranged as shown in Table 2.1. Along with each vowel is shown a key word containing the vowel.

Table 2.1. *Vowels*

Degree of constriction	Tongue hump position		
	front	central	back
High	/i/ eve /I/ it	/ɜ˞/ bird /ə˞/ over (unstressed)	/u/ boot /ʊ/ foot
Medium	/e/ hate* /ɛ/ met	/ʌ/ up /ə/ ado (unstressed)	/o/ obey* /ɔ/ all
Low	/æ/ at		/ɑ/ father

* These two sounds usually exist as diphthongs in GA dialect. They are included in the vowel table because they form the nuclei of related diphthongs. See Section 2.27 for further discussion. (See also PETERSON and LEHISTE.)

The approximate articulatory configurations for the production of these sounds (exclusive of the two unstressed vowels) are shown qualitatively by the vocal tract profiles in Fig. 2.5 (POTTER, KOPP and GREEN). The physiological basis for the front-back/high-low classification is particularly well illustrated if the profiles for the vowels /i, æ, a, u/ are compared[1].

[1] These profiles, and the ones shown subsequently in this chapter, mainly illustrate the oral cavity. The important pharynx cavity and the lower vocal tract are not drawn. Their shapes may be deduced from x-rays (see Figs. 5.29 through 5.31, for example).

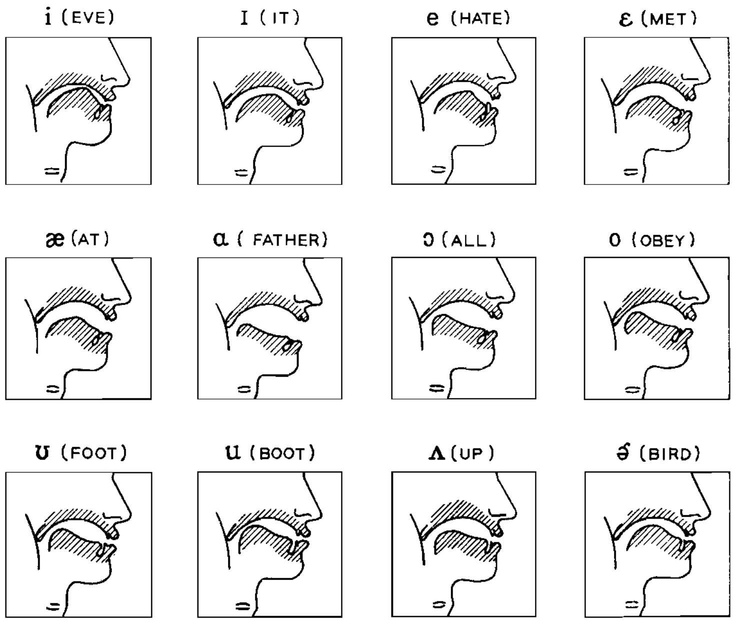

Fig. 2.5. Schematic vocal tract profiles for the production of English vowels. (Adapted from Potter, Kopp and Green)

2.22. Consonants

The consonants constitute those sounds which are not exclusively voiced and mouth-radiated from a relatively stable vocal configuration. They generally are characterized by greater tract constrictions than the vowels. They may be excited or radiated differently, or both. The short-time dynamic motions of the vocal apparatus are crucial to the production of an important class of consonants. Those consonants for which vocal motion is not requisite may be uttered as sustained sounds (as vowels may be) and hence are termed *continuants*.

2.221. Fricative Consonants. Fricatives are produced from an incoherent noise excitation of the vocal tract. The noise is generated by turbulent air flow at some point of constriction. Common constrictions for producing fricative consonants are those formed by the tongue behind the teeth (dental), the upper teeth on the lower lip (labio-dental), the tongue to the gum ridge (alveolar), the tongue against the hard or soft palate (palatal or velar, respectively), and the vocal cords constricted and fixed (glottal). Radiation of fricatives normally occurs from the mouth. If the vocal cord source operates in conjunction with the noise

source, the fricative is a voiced fricative. If only the noise source is used, the fricative is unvoiced.

Both voiced and unvoiced fricatives are continuant sounds. Because a given fricative articulatory configuration can be excited either with or without voicing, the voiced and voiceless fricatives form complementary pairs called *cognates*. The fricative consonants of the GA dialect are listed in Table 2.2, along with typical "places" of articulation and key words for pronunciation.

Vocal tract profiles for these sounds are shown in Fig. 2.6. Those diagrams in which the vocal cords are indicated by two small lines are the voiced fricatives. The vocal cords are shown dashed for the glottal fricative (h).

Table 2.2. *Fricative consonants*

Place of articulation	Voiced	Voiceless
Labio-dental	/v/ vote	/f/ for
Dental	/ð/ then	/θ/ thin
Alveolar	/z/ zoo	/s/ see
Palatal	/ʒ/ azure	/ʃ/ she
Glottal		/h/ he

2.222. Stop Consonants. Among those consonants which depend upon vocal tract dynamics for their creation are the stop consonants. To produce these sounds a complete closure is formed at some point in the vocal tract. The lungs build up pressure behind this occlusion, and the pressure is suddenly released by an abrupt motion of the articulators. The explosion and aspiration of air help to characterize the stops. The closure can be labial, alveolar, palatal or velar. The stop can be produced with or without simultaneous voicing. In fact, a voiced consonant may employ voiced excitation to build up the requisite pressure, in which case voicing starts before the pressure release. The cognate pairs of stops, with typical places of articulation, are shown in Table 2.3.

Table 2.3. *Stop consonants*

Place of articulation	Voiced	Voiceless
Labial	/b/ be	/p/ pay
Alveolar	/d/ day	/t/ to
Palatal/velar	/g/ go	/k/ key

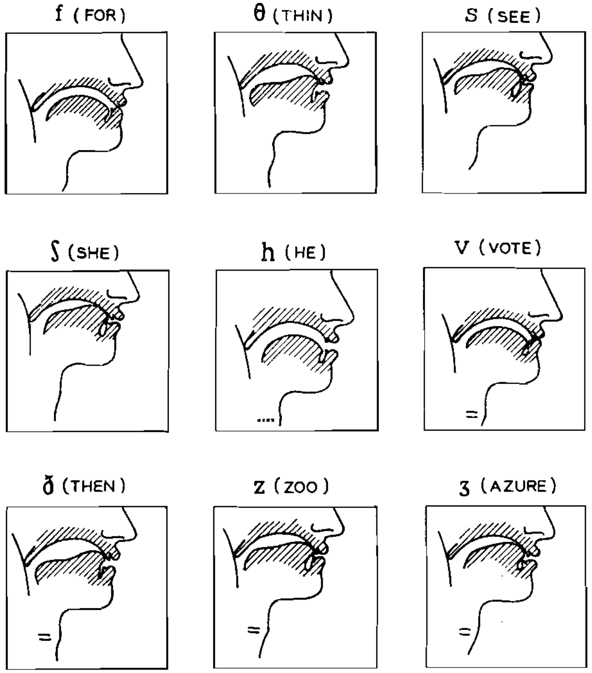

Fig. 2.6. Vocal tract profiles for the fricative consonants of English. The short pairs of lines drawn on the throat represent vocal cord operation. (Adapted from POTTER, KOPP and GREEN)

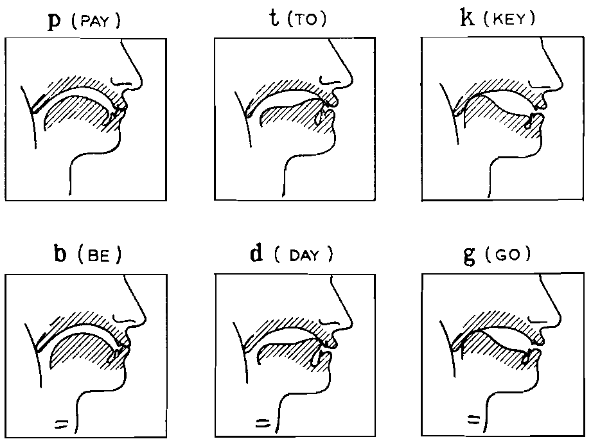

Fig. 2.7. Articulatory profiles for the English stop consonants. (After POTTER, KOPP and GREEN)

Articulatory profiles for these sounds are shown in Fig. 2.7. Each position is that just prior to the pressure release.

2.223. Nasal Consonants. The nasal consonants, or nasals, are normally excited by the vocal cords and hence are voiced. A complete closure is made toward the front of the vocal tract, either by the lips, by the tongue at the gum ridge, or by the tongue at the hard or soft palate. The velum is opened wide and the nasal tract provides the main sound transmission channel. Most of the sound radiation takes place at the nostrils. The closed oral cavity functions as a side branch resonator coupled to the main path, and it can substantially influence the sound radiated. Because the nasals can be sustained, they are classed as continuants. The GA nasal consonants are listed in Table 2.4, and their vocal profiles are illustrated in Fig. 2.8.

Table 2.4. *Nasals*

Place	
Labial	/m/ me
Alveolar	/n/ no
Palatal/velar	/ŋ/ sing (no initial form)

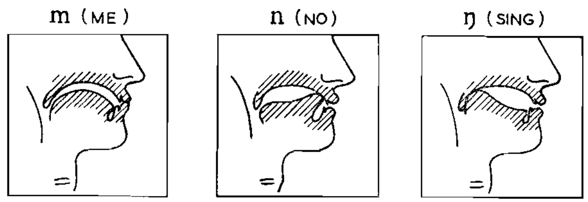

Fig. 2.8. Vocal profiles for the nasal consonants. (After POTTER, KOPP and GREEN)

2.224. Glides and Semivowels. Two small groups of consonants contain sounds that greatly resemble vowels. These are the glides /w, j/ and the semivowels /r, l/ (FAIRBANKS). Both are characterized by voiced excitation of the tract, no effective nasal coupling, and sound radiation from the mouth. The glides are dynamic sounds, invariably precede a vowel, and exhibit movement toward the vowel. The semivowels are continuants in which the oral channel is more constricted than in most vowels, and the tongue tip is not down. These sounds for the GA dialect are listed, according to place of articulation, in Table 2.5. Their profiles, for the beginning positions, are given in Fig. 2.9.

Table 2.5. *Glides and semi-vowels*

Place	
Palatal	/j/ you
Labial	/w/ we
	(no final form)
Palatal	/r/ read
Alveolar	/l/ let

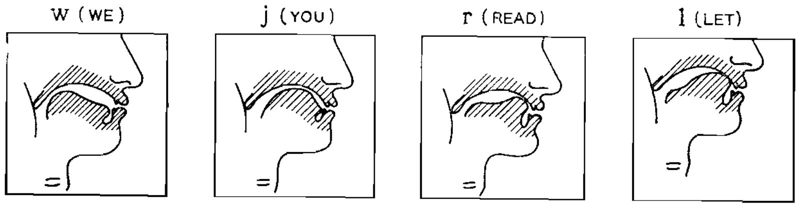

Fig. 2.9. Vocal tract configurations for the beginning positions of the glides and semivowels. (After POTTER, KOPP and GREEN)

2.225. Combination Sounds: Diphthongs and Affricates. Some of the preceding vowel or consonant elements can be combined to form basic sounds whose phonetic values depend upon vocal tract motion. An appropriate pair of vowels, so combined, form a diphthong. The diphthong is vowel-like in nature, but is characterized by change from one vowel position to another. For example, if the vocal tract is changed from the /e/ position to the /ɪ/ position, the diphthong /eɪ/ as in *say* is formed. Other GA diphthongs are /ɪu/ as in *new*, /ɔɪ/ as in *boy*; /aʊ/ as in *out*, /aɪ/ as in *I*, and /oʊ/ as in *go*.

As vowel combinations form the diphthongs, stop-fricative combinations likewise create the two GA affricates. These are the /tʃ/ as in *chew* and the /dʒ/ as in *jar*.

2.3. Quantitative Description of Speech

The preceding discussion has described the production of speech in a completely qualitative way. It has outlined the mechanism of the voice and the means for producing an audible code which, within a given language, consists of distinctive sounds. However, for any transmission system to benefit from prior knowledge of the information source, this knowledge must be cast into a tractable analytical form that can be employed in the design of signal processing operations. Detailed inquiry into the physical principles underlying the speech-producing mechanism is therefore indicated.

The following chapter will consider the characteristics of the vocal system in a quantitative fashion. It will treat the physics of the vocal and nasal tracts in some depth and will set forth certain acoustical properties of the vocal excitations. The primary objective—as stated earlier—is to describe the acoustic speech signal in terms of the physical parameters of the system that produced it. Because of physiological and linguistic constraints, such a description carries important implications for analysis-synthesis telephony.

III. Acoustical Properties of the Vocal System

The collection of olfactory, respiratory and digestive apparatus which man uses for speaking is a relatively complex sound-producing system. Its operation has been described qualitatively in the preceding chapter. In this chapter we would like to consider in more detail the acoustical principles underlying speech production. The treatment is not intended to be exhaustive. Rather it is intended to circumscribe the problems of vocal tract analysis and to set forth certain fundamental relations for speech production. In addition, it aims to outline techniques and method for acoustic analysis of the vocal mechanism and to indicate their practical applications. Specialized treatments of a number of these points can be found elsewhere[1].

3.1. The Vocal Tract as an Acoustic System

The operations described qualitatively in the previous chapter can be crudely represented as in Fig. 3.1. The lungs and associated respiratory muscles are the vocal power supply. For voiced sounds, the expelled air causes the vocal cords to vibrate as a relaxation oscillator, and the air stream is modulated into discrete puffs or pulses. Unvoiced sounds are generated either by passing the air stream through a constriction in the tract, or by making a complete closure, building up pressure behind the closure and abruptly releasing it. In the first case, turbulent flow and incoherent sound are produced. In the second, a brief transient excitation occurs. The physical configuration of the vocal tract is highly variable and is dictated by the positions of the articulators; that is, the jaw, tongue, lips and velum. The latter controls the degree of coupling to the nasal tract.

[1] For this purpose G. FANT, *Acoustic Theory of Speech Production*, is highly recommended. Besides presenting the acoustical bases for vocal analysis, this volume contains a wealth of data on vocal configurations and their calculated frequency responses. An earlier but still relevant treatise is T. CHIBA and M. KAJIYAMA, *The Vowel: Its Nature and Structure*. Another excellent and more recent analysis of vowel articulation is G. UNGEHEUER, *Elemente einer akustischen Theorie der Vokalartikulation*.