a summary review. The student without such background should be able to acquire at least the basics of articulatory phonetics. The topics to be discussed are these:

- The Speech Mechanism
- Vowels
  - Monophthongs (single vowels)
  - Diphthongs
- Consonants
  - Stops
  - Nasals
  - Fricatives
  - Affricates
  - Liquids
  - Glides
- Suprasegmentals
- Coarticulation
- Aerodynamics
- Acoustics
- Sensory Information
- Implications for Acquisition

FUNDAMENTALS OF ARTICULATORY PHONETICS

The Speech Mechanism

The anatomy of the speech production system is not within the scope of this chapter, but some general anatomical descriptions are needed to discuss the fundamentals of articulatory phonetics. The basic aspects of speech production can be understood by an examination of six principal organs or subsystems, illustrated in Figure 2.2. The *respiratory system*, consisting of the lungs, airway, rib cage, diaphragm, and associated structures, provides the basic air supply for generating sound. The *larynx,* composed of various cartilages and muscles, generates the voiced sounds of speech by vibration of the vocal folds, or it allows air to pass from lungs to the vocal tract (the oral and nasal cavities) for voiceless sounds. The *velopharynx*—the soft palate (or velum) and associated structures of the velopharyngeal port—joins or separates the oral and nasal cavities so that air passes through the oral cavity, the nasal cavity, or both. The *tongue*, primarily a complex of muscles, is the principal articulator of the oral cavity; it is capable of assuming a variety of shapes and positions in vowel and consonant articulation. For articulatory purposes, the tongue is divided into five major parts: the tip or apex, the blade, the back or dorsum, the root, and the body. These divisions are illustrated in Figure 2.3. The *lips*, along with the jaw, are the most visible of the articulators; they are involved in the production of vowels and consonants. The *jaw*, the massive bony structure and its associated muscles, supports the soft tissues of both tongue and lower lip. It participates in speech production by aiding tongue and lip movements and by providing skeletal support for these organs. Other anatomical features shown in Figure 2.2 provide general orientation or are relevant in a significant way to the processes of speech and hearing.

The respiratory system and larynx work together to provide the upper airway with two major types of air flow: a series of pulses of air created by the action of the vibrating vocal folds (for voiced sounds like the sounds in the word *buzz*) and a continuous flow of air that can be used to generate noise energy in the vocal tract (for voiceless sounds like the *s* in *see*). The basic function of the respiratory system in speech is to push air into the airway composed of the larynx and the oral and nasal cavities. The basic function of the larynx is to regulate the airflow from the lungs to create both voiced and voiceless segments. The upper airway, often called the *vocal tract*, runs from the larynx to the mouth or nose and is the site of what is commonly called

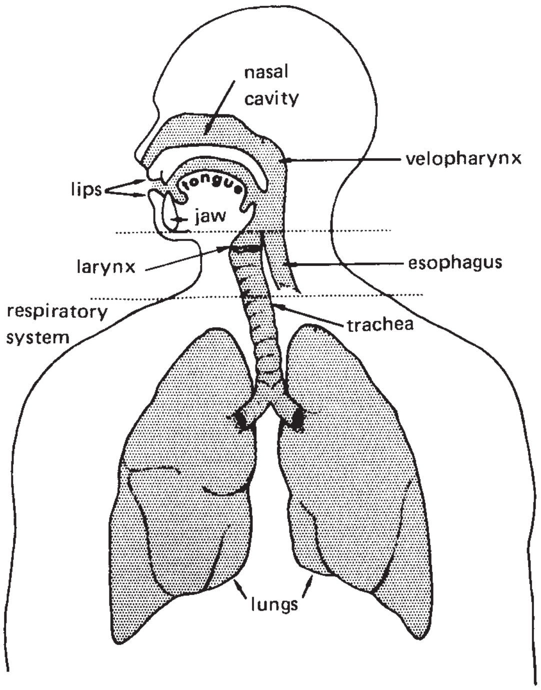

FIGURE 2.2 Organs of speech production.

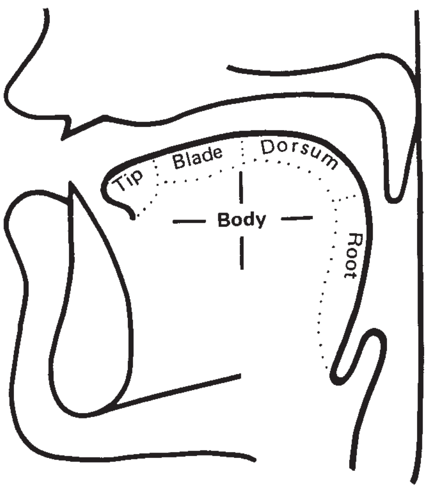

FIGURE 2.3 Divisions of tongue into five functional parts for speech articulation.

speech articulation. For the most part, this process is accomplished by movements of the *articulators*: tongue, lips, jaw, and velopharynx. The vocal tract may be viewed as a flexible tube that can be lengthened or shortened (by moving the larynx up and down in the neck or by protruding and retracting the lips) and constricted at many points along its length by actions of tongue, velopharynx, and lips. Speech articulation is thus a matter of lengthening, shortening, and constricting the tube known as the *vocal tract.*

This entire process is controlled by the nervous system, which must translate the message to be communicated into a pattern of signals that run to the various muscles of the speech mechanism. As these muscles contract, a variety of things can happen: Air may be pushed out of the lungs, the vocal folds may start to vibrate, the velopharynx may close, the jaw may lower, or the lips may protrude. The brain has the task of coordinating all the different muscles so that they contract in the proper sequence to produce the required phonetic result. The margin for error is small; sometimes an error of just a few milliseconds in the timing of a muscle contraction can result in a misarticulation.

It is appealing to suppose that speech production is controlled at some relatively high level of the brain by discrete units, such as phonemes. However, a major problem in the description of speech articulation is to relate the discrete linguistic units that operate at a high level of the brain to the muscle contractions that result in articulatory movements. For example, to say the word *stop*, a speaker's brain must send nerve instructions, in the proper sequence, to the muscles of the respiratory system, larynx, tongue, lips, and velopharynx. The full understanding of speech production therefore involves a knowledge of *phonology* (the study of how sounds are put together to form words and other linguistic units), *articulatory phonetics* (the study of how the articulators make individual sounds), *acoustic phonetics* (the study of the relationship between articulation and the acoustic signal of speech), and *speech perception* (the study of how phonetic decisions are made from the acoustic signal).

Vowel Articulation: Traditional Phonetic Description

A vowel sound is usually formed as sound energy from the vibrating vocal folds escapes through a relatively open vocal tract of a particular shape. Because a syllable must contain a vowel or vowel-like sound, vowels sometimes are called *syllable nuclei*. Each vowel has a characteristic vocal tract shape that is determined by the position of the tongue, jaw, and lips. Although other parts of the vocal tract, like the velum, pharyngeal walls, and cheeks, may vary somewhat with different vowels, the positions of the tongue, jaw, and lips are of primary consequence. Therefore, individual vowels can be described by specifying the articulatory positions of tongue, jaw, and lips. Furthermore, because the jaw and tongue usually work together to increase or reduce the mouth opening (Figure 2.4), for general phonetic purposes, vowel production can be described by specifying the positions of just two articulators, tongue and lips. Usually the vocal folds vibrate to produce voicing for vowels, but exceptions, such as whispered speech, do occur.

The two basic lip articulations can be demonstrated with the vowels in the words *he* and *who*. Press your finger against your lips as you say first *he* and then *who*. You should feel the lips push against your finger as you say *who*. The vowel in this word is a rounded vowel, meaning that the lips assume a rounded, protruded posture. Vowels in English are described as being either rounded, like the vowel in *who*, or unrounded, like the vowel in *he*. Figure 2.5 illustrates the lip configuration for these two vowels.

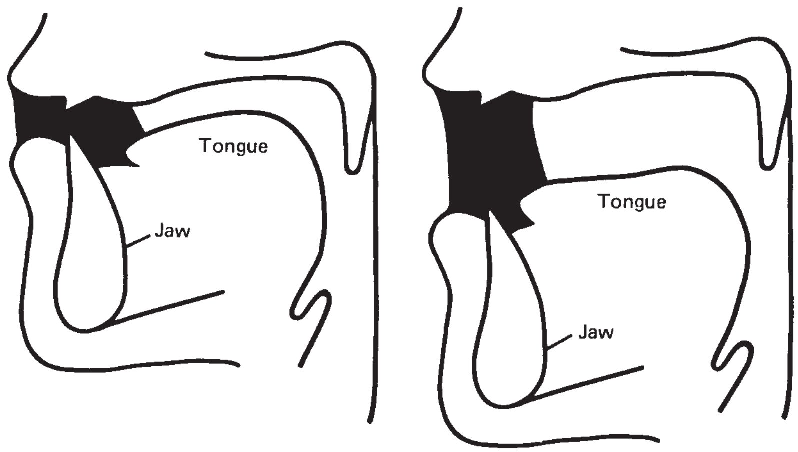

FIGURE 2.4 Variations in mouth opening (darkened area) related to lowering of jaw and tongue.

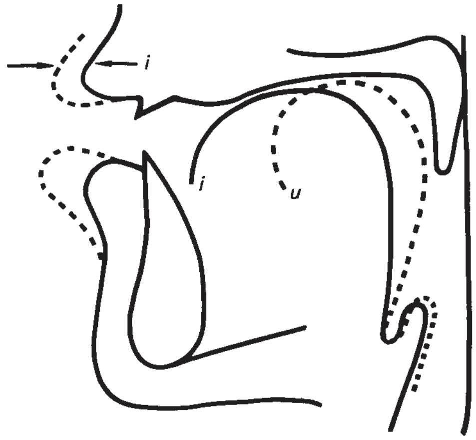

FIGURE 2.5 Vocal tract configurations for /i/ and /u/. Note lip rounding for /u/.

The tongue moves in essentially two dimensions within the oral cavity, as shown in Figure 2.6. One dimension, front-back, is represented by the motion the tongue makes as you alternately say *he*, *who* or *map*, *mop*. The other dimension, high-low, is represented by the motion the tongue makes as you say *heave-have* or *who-ha*. With these two dimensions of tongue movement, we can define four extreme positions of the tongue within the oral cavity, as shown in Figure 2.7. The phonetic symbols for these four vowels also are shown in the illustration. With the tongue high and forward in the mouth, the high-front vowel /i/ as in *he* is produced. When the tongue is low and forward in the mouth, the low-front vowel /æ/ as in *have* is produced.

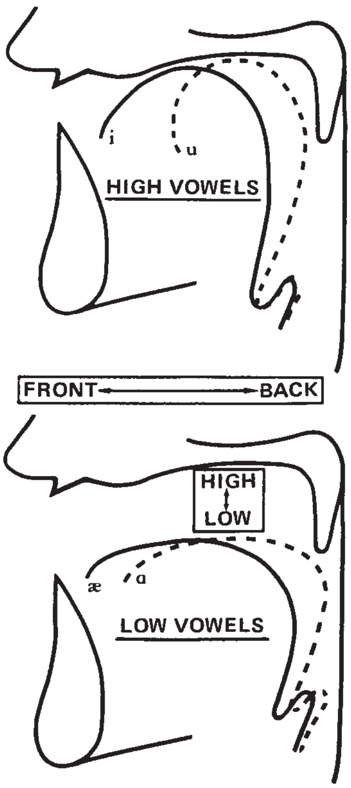

FIGURE 2.6 The two major dimensions of tongue position, front-back and high-low.

TABLE 2.4 Distinctive Feature Classifications for Selected Consonants

Feature	p	b	m	t	d	n	s	l	θ	k
Consonantal	+	+	+	+	+	+	+	+	+	+
Vocalic	–	–	–	–	–	–	–	–	–	–
Sonorant	–	–	+	–	–	+	–	+	–	–
Interrupted	+	+	–	+	+	–	–	–	–	+
Strident	–	–	–	–	–	–	+	–	–	–
High	(–)	(–)	(–)*	–	–	–	–	–	–	+
Low	(–)	(–)	(–)	–	–	–	–	–	–	–
Back	(–)	(–)	(–)	–	–	–	–	–	–	+
Anterior	+	+	+	+	+	+	+	+	+	–
Coronal	–	–	–	+	+	+	+	+	+	–
Rounded	–	–	–	–	–	–	–	–	–	–
Distributed	+	+	+	–	–	–	–	–	+	–
Lateral	–	–	–	–	–	–	–	+	–	–
Nasal	–	–	+	–	–	+	–	–	–	–
Voiced	–	+	+	–	+	+	–	+	–	–

*Feature values enclosed in parentheses indicate that the feature in question may not be specified for this sound. For example, tongue position for /p/, /b/, and /m/ is not really specified because it is free to assume the position required for the following vowel.

High sounds are made with the tongue elevated above its neutral (resting) position (see Figure 2.11a).

Low sounds are made with the tongue lowered below its neutral position (see Figure 2.11b).

Back sounds are made with the tongue retracted from its neutral position (see Figure 2.11c).

Anterior sounds have an obstruction that is farther forward than that for the palatal /ʃ/. Anterior sounds include the bilabials, labiodentals, linguadentals, and linguaalveolars.

Coronal sounds have a tongue blade position above the neutral state. In general, consonants made with an elevated tongue tip or blade are +coronal.

Rounded sounds have narrowed or protruded lip configuration.

Distributed sounds have a constriction extending over a relatively long portion of the vocal tract (from back to front). For English, this feature is particularly important to distinguish the dental fricatives /θ/ and /ð/ from the alveolars /s/ and /z/.

Lateral sounds are coronal consonants made with midline closure and lateral opening.

Nasal sounds have an open velopharynx allowing air to pass through the nose.

Voiced sounds are produced with vibrating vocal folds.

The feature assignments in Table 2.4 are for general illustration of the use of features. The features should be viewed with some skepticism because several different feature systems have been proposed, and any one system is subject to modification. It should be understood that distinctive features are one type of classification system. It should also be realized that distinctive features have an intended linguistic function that may not always be compatible with their application to the study of articulation disorders. The issue is beyond the scope of this chapter, but the interested reader is referred to Walsh (1974) and Parker (1976).

The relationship between the traditional place terms of phonetic description and the distinctive features is summarized here. For each traditional place term, the associated features

The information to the left of the arrow indicates the segments that conform to the rule. The arrow means "is realized as." Only the relevant rules are included to the right of the arrow. Other features are assumed to remain as they were. The diagonal slash means "in the context of." The dash and information that follow provide the context of the segment described by the rule. Thus, this generative phonology rule reads: Vowels are realized as nasal in the context of (in this case, specifically just before) nasal consonants.

Application to Typically Developing Children. Generative phonology has been applied to the understanding of children's speech acquisition (cf. Grunwell, 1987) as it enabled description of the relationship of children's productions to adult pronunciation in terms of phonological rules. Grunwell indicated that generative phonology has been readily applied to children's speech because generative phonological rules can explain substitutions, distortions, omissions, additions, metathesis, and coalescence (e.g., see Grunwell, pp. 176–197).

Some of the premises of generative phonology have received criticism in subsequent research. For example, there has been criticism of the premise that the child's underlying representation of the sound is adultlike (this viewpoint will be discussed later when we consider psycholinguistic theories). Additionally, there has been criticism of the premise that the rules that were applied had a corresponding reality to the processing and production systems of the child (i.e., it is not clear that we actually apply such rules in our heads when we comprehend and produce speech).

Application to Speech-Language Pathology Practice. As a theory, generative phonology has not seen broad application in the field of speech-language pathology. Hodson (2010b) describes generative phonology as the "first steps into phonologically based clinical analysis" (p. 55); however, additional knowledge gained from the theory of natural phonology (below) led to the identification of patterns in phonological analysis procedures.

Natural Phonology

The theory of natural phonology (Stampe, 1969, 1979) formed the basis of the phonological process approach to assessment and treatment of speech sound disorders and is regarded as the phonological model that has had the greatest impact on the field of SLP (Edwards, 2007). *Natural processes* (or *patterns*) are those that are preferred or frequently used in phonological systems and are identified in two ways: those that are universal across languages and those that are frequently used by young children. According to Stampe, a phonological process is a "mental operation that applies in speech to substitute for a class of sounds or sound sequences presenting a common difficulty to the speech capacity of the individual, an alternative class identical but lacking the difficult property" (1979, p. 1), and phonological processes merge "a potential opposition into that member of the opposition which least tries the restrictions of the human speech capacity" (1969, p. 443).

In Stampe's view, the child's underlying representations are akin to adult forms. Natural (or innate) phonological processes apply to these underlying representations, resulting in the child's productions (or surface forms). For example, it is assumed that children have the adult form of a word, such as *tree* /tri/, in their underlying representation. However, natural processes such as cluster reduction are applied because the child (at least temporarily) has some limitation to produce a particular sound or group of sounds. In this case, the surface form (child's production) would most likely be [ti]. Later, in the discussion of psycholinguistic models, we will critique the notion that children's underlying representations are akin to the adult form. A shortcoming of this theory is that some errors may fit into more than one category. For example, if a child attempted to say *dance* /dæns/ and said [dæn] instead, it is not clear if this is an example of final consonant deletion, cluster reduction, stridency deletion, or some combination of these.

Application to Typically Developing Children. Natural phonology has provided insight to the understanding of typical speech acquisition. Natural processes are described as innate rules that are systematically applied to speech production until children learn to suppress them. Because these rules are universal, they are meant to apply to all children speaking all languages. Thus, speech acquisition is a progression from these innate speech patterns to the pronunciation system of the language(s) learned by the child. By applying natural phonology to English speech acquisition, Grunwell (1987) presented a table of the ages of suppression of phonological processes by typically developing children, such as cluster reduction, fronting, and stopping. Other researchers have also provided lists of natural phonological processes (e.g., Ingram, 1976; Shriberg and Kwiatkowski, 1980). Shriberg and Kwiatkowski advocated the clinical use of eight "natural processes": (1) final consonant deletion, (2) velar fronting, (3) stopping, (4) palatal fronting, (5) liquid simplification, (6) cluster reduction, (7) assimilation, and (8) unstressed-syllable deletion.

Application to Speech-Language Pathology Practice. The phonological pattern/process approach to assessment and intervention based on natural phonology transformed the way that SLPs viewed children's speech sound errors. Since Ingram's (1976, 1989a) seminal work on phonological impairments in children, SLPs increasingly have applied descriptive linguistic-based models to their clinical activities. Ingram's application of natural phonology was widely accepted by SLPs in the 1970s and 1980s and remains popular for directing the assessment, analysis, and intervention of children with speech sound disorders (Bankson and Bernthal, 1990a; Khan, 1982; Shriberg and Kwiatkowski, 1980; Weiner, 1979). Assessment approaches were developed to specifically assess subgroups of sounds within a given phonological pattern (e.g., *Bankson-Bernthal Test of Phonology* [*BBTOP*] [Bankson and Bernthal, 1990a]). Phonological processes were also described as part of a broader analysis procedure for several speech-sampling tools (e.g., *Phonological Assessment of Child Speech* [*PACS*] [Grunwell, 1985]) and as stand-alone analyses to be applied to conversational speech (e.g., *Natural Process Analysis* [Shriberg and Kwiatkowski, 1980]). One of the goals of intervention based on natural phonology is "to teach children to suppress innate simplification processes" (Hodson, 2010b, p. 55).

Limitations of the application of natural phonology to SLP practice have been identified. First, although most SLPs can readily describe children's nonadult productions using phonological process terms such as *cluster reduction* and *fronting*, SLPs' use of phonological processes are descriptive rather than an application of the theoretical tenets of natural phonology. Shriberg (1991, p. 270) described this as an "atheoretical use of process terminology." Second, natural phonology does not account for "nonnatural" simplifications in children's speech (Hodson, 2010a). Many children with highly unintelligible speech produce speech sounds in a way that cannot be classified using natural phonology. Terms such as *backing* and *initial consonant deletion* are in the literature to describe phonological processes that are not seen in children with typical speech acquisition (Dodd, 1995b).

One question that remains unresolved with natural phonology is whether the process labels being applied actually represent mental operations going on inside the head of the child. However, because such labels do capture "patterns" of errors being observed, the term *phonological patterns* is frequently used in place of *phonological processes*. For example, the title of a popular assessment tool in this area is the *Hodson Assessment of Phonological Patterns* (Hodson, 2004).

Nonlinear Phonology

Nonlinear phonology refers to a collection of theories that focus on the hierarchical nature of the relationships between phonological units. Goldsmith introduced nonlinear phonology in his doctoral dissertation (1979) and later expanded upon it (Goldsmith, 1990). These theories include autosegmental theory, metrical theory, moraic theory, feature geometry theory, and underspecification theory. Nonlinear phonology attempts to account for the idea that production of speech involves more than just production of a sequence of phonemes; it takes into account many elements (features, segments, syllables, feet, words, and phrases) both independently and in relation to one another; hence, the term *nonlinear*. There are two main tiers in nonlinear phonology:

1. The *prosodic tier* focuses on words and the structure of words and includes a number of levels: word tier, foot tier, syllable tier, onset-rime tier, skeletal tier, and segmental tier (see Figure 3.1).
2. The *segmental tier* focuses on the segments or speech sounds and the features that make up those sounds (see Figure 3.2).

In the prosodic tier, the *word tier* simply denotes words. Immediately below the word tier is the *foot tier*, which refers to grouping of syllables, and syllables may be either strong (S) or weak (w). A foot can contain only one strong syllable (but can also contain other weak syllables). A foot that includes a weak syllable can be either Sw (left prominent, or trochaic), or wS (right prominent, or iambic). Below the foot is the *syllable tier*. A syllable consists of one prominent phoneme (the peak), which is usually a vowel and less prominent phonemes (generally consonants) that can appear before or after the peak. Consonants that appear before the vowel are known as *onsets,* and consonants that appear after the vowel are *codas.* The peak and the coda together make up the *rime.* All languages allow syllables without a coda, which are sometimes called *open syllables* (e.g., CV). Some languages do not allow for *closed syllables* that have a coda (e.g., CVC). Across the world's languages, open syllables occur more often than closed. Below the syllable tier is the *skeletal tier*, which includes slots for the individual speech sounds.

In the *segmental tier*, features are described according to three nodes: the root node, the laryngeal node, and the place node (see Figure 3.2). The root node [sonorant] and [consonantal]

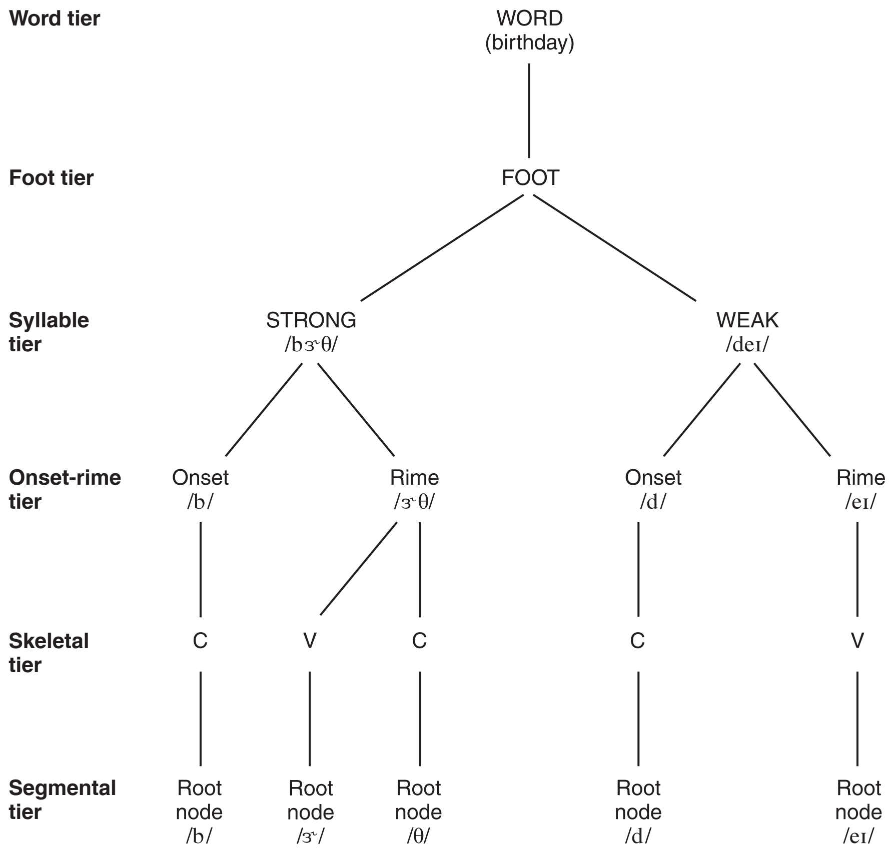

FIGURE 3.1 An example of the prosodic tier representation for the word *birthday.*

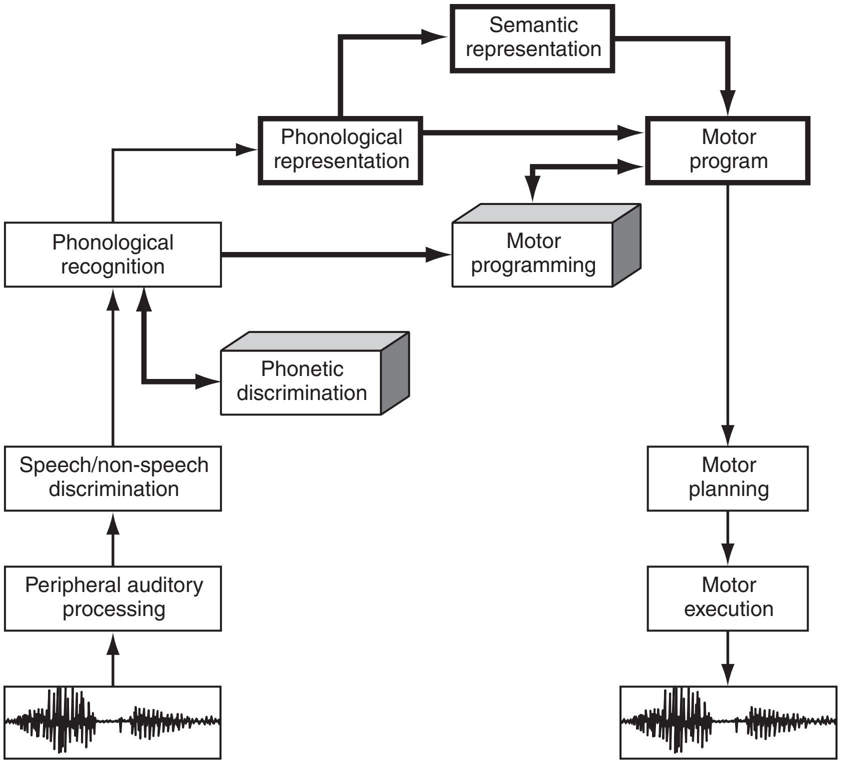

FIGURE 5.1 Speech-processing model by Stackhouse and Wells (1997).

Note: The broad arrows and shaded boxes represent processes hypothesized to occur off-line.

Source: From *Children's Speech and Literacy Difficulties: A Psycholinguistic Framework* (p. 350) by J. Stackhouse and B. Wells, 1997. Copyright John Wiley & Sons Limited. Reproduced with permission.

both real-word picture naming and nonword repetition, suggesting that she was having problems with her already stored *motor programs* as well as difficulty creating new motor programs (i.e., *motor programming*).

Stackhouse and Wells (1997) did not initially develop their model with the specific intention of identifying subgroups. However, Vance, Stackhouse, and Wells (2005) suggested that using profiles based on a particular pattern of problems might provide additional insights into the nature of a child's problem. Thus, using the model to identify subgroups seems at least plausible. One might, for example, evaluate large groups of children with SSDs of unknown origin to determine whether certain profiles are more common than others. Stackhouse (2000) used this approach in a longitudinal study to suggest how different profiles might help differentiate children with SSDs who have later literacy problems from those who learn to read normally.

Classification by Symptomatology

Barbara Dodd and colleagues classify SSDs of unknown origin in a different manner. Dodd, Holm, Crosbie, and McCormack (2005) note that "there is as yet no theoretically adequate or clinically relevant explanation of disordered speech. . .current models of the speech-processing chain. . .fail to disentangle causal, comorbid, and consequent difficulties" (p. 44). Dodd and colleagues argue that "surface error patterns" (i.e., symptoms) provide the best perspective on classification and that such patterns do, in fact, explain the nature of the disorder. Dodd (2014) also offers specific treatment recommendations for most of the five subgroups she proposes.

The *Diagnostic Evaluation of Articulation and Phonology (DEAP)* (Dodd et al., 2006) was specifically developed to identify subgroups of children with SSDs.

The first subgroup proposed by Dodd is *articulation disorder,* in which the child produces consistent substitution or distortion errors on a limited number of phonemes (often /s/ or /r/). Errors do not change whether the production is spontaneous or imitated. In a cohort study of 320 children with SSDs, Broomfield and Dodd (2004) classified 40 (12.5%) children into this subgroup. According to Dodd (2014), traditional motor-based therapy (see Chapter 9) is considered the most appropriate treatment approach for this subgroup.

The second and largest of Dodd's subgroups, *phonological delay,* includes children whose errors can be described using phonological process (pattern) labels that are also seen in much younger typically developing children (i.e., developmental patterns). Broomfield and Dodd (2004) classified 184/320 (57.5%) children into this subgroup. Dodd (2014) suggested that this subgroup might respond best to what we have here termed linguistically based approaches (see Chapter 10).

Dodd's third subgroup, *consistent atypical phonological disorder*, includes children who produce one or more nondevelopmental patterns (i.e., those not usually produced by typically developing children). In this case, the errors are produced consistently. Broomfield and Dodd (2004) classified 66/320 (20.6%) children into this subgroup. A specific type of linguistically based approach called *contrast therapy* may be best suited for these children, according to Dodd (2014).

Dodd's fourth subgroup, *inconsistent phonological disorder*, includes children who produce nondevelopmental error patterns but do so inconsistently. In this case, *inconsistency* refers to variations in output of repeated productions of the same words (the same description of inconsistency mentioned previously in our discussion of CAS). In this case, however, none of the other signs of CAS would be present (see Table 5.3). Broomfield and Dodd (2004) classified 30/320 (9.4%) children into this subgroup. Dodd (2014) recommended a core vocabulary approach (to be discussed in Chapter 9) for this subgroup.

Dodd's final subgroup is CAS. In her 2014 paper, Dodd's definition is very reminiscent of (although not identical to) our earlier discussion of CAS. She says that this subgroup demonstrates

> Speech characterised by inconsistency, oromotor signs (e.g., groping, difficulty sequencing articulatory movements), slow speech rate, disturbed prosody, short utterance length, poorer performance in imitation than spontaneous production. CAS is rare, and reliable identification is clinically challenging. It may involve multiple deficits affecting phonological and phonetic planning as well as motor program implementation. (p.193)

Dodd and colleagues have attempted to empirically validate the categories. For example, Dodd (2011) compared 23 children classified as delayed against an age-matched group of 23 children classified as disordered (producing five or more instances of at least one atypical error pattern). The delayed group made significantly fewer consonant errors overall as well as fewer types of errors. Although the groups did not differ on general measures of language ability, the disordered group performed significantly less well on a nonlinguistic rule-learning task and showed less cognitive flexibility on two tests of executive function. Dodd interpreted the findings to mean that the disordered group was less able to sort out the sound system of the language. Validation of Dodd's categories is also being sought by demonstrating that children in different groups respond differentially to different treatment approaches, as mentioned above. For example, Dodd and Bradford (2000) presented case study data on three children: One child with a consistent phonological disorder received the most benefit from a phonological contrast (contrast therapy) approach, whereas two children with inconsistent phonological disorders initially responded most quickly to a core vocabulary approach. Once these latter children's errors became more consistent, one of them responded more quickly to phonological contrast therapy. These same associations between subgroup membership and response to therapy were also observed in a follow-up group study that included a total of 18 children (Crosbie, Holm, and Dodd, 2005).