

**CHAPTER TWO**  
**SPEECH AND LANGUAGE PROCESSING: MODELS, THEORIES AND THE**  
**INFLUENCE OF CONTEXTUAL FACTORS**

---

2.1	INTRODUCTION .....	23
2.2	THEORIES AND MODELS OF SPEECH PRODUCTION .....	26
2.2.1	Motor program-based theory .....	27
2.2.2	Dynamic Systems Models.....	29
2.2.3	An information-processing model: Levelt’s theory .....	32
2.2.4	Van der Merwe's four-level framework of speech sensorimotor control .....	35
2.2.4.1	Intention to communicate verbally .....	35
2.2.4.2	Linguistic-symbolic planning .....	36
2.2.4.3	Motor planning.....	36
2.2.4.4	Motor Programming.....	38
2.2.4.5	Execution .....	39
2.2.5	Conclusion regarding models and theories of speech production .....	39
2.3	FACTORS INFLUENCING INFORMATION PROCESSING .....	40
2.3.1	Factors influencing information processing related to the task and context.....	41
2.3.1.1	Stages of learning.....	41
2.3.1.2	Automaticity .....	41
2.3.1.3	Movement time .....	43
2.3.1.4	Movement complexity .....	44
2.3.1.5	Environment.....	44
2.3.2	Factors influencing information processing inherent to the individual .....	45
2.3.2.1	Attention .....	45
2.3.2.1.1	Central resource capacity theories	46
2.3.2.1.2	Multiple resource theories	47
2.3.2.1.3	Attention and context	48
2.3.2.2	Memory.....	48
2.3.2.2.1	Working memory or short-term memory	49
2.3.2.2.2	Long-term memory	50
2.3.3	Conclusion regarding factors influencing information processing.....	51

2.4	CONTEXTS FOR SPEECH PRODUCTION .....	51
2.4.1	Speaking rate.....	52
2.4.2	Level of voluntary initiation of actions.....	53
2.4.3	Familiarity: Novel versus automatic production.....	54
2.4.4	Sound structure .....	55
2.4.5	Motor complexity.....	55
2.4.6	Length of the utterance .....	56
2.4.7	Linguistic complexity .....	57
2.4.8	First versus second language in bilingual speakers as a context for speech production .....	59
2.5	BILINGUAL SPEECH AND LANGUAGE PROCESSING.....	61
2.5.1	Defining bilingualism .....	61
2.5.2	Types of bilingualism .....	62
2.5.3	Crosslinguistic influence.....	63
2.5.4	The study of specific aspects of bilingual speech production.....	64
2.5.5	Language processing and automaticity in bilinguals.....	67
2.5.6	Neuroanatomical organization in bilinguals .....	68
2.5.7	Conclusion regarding bilingual speech and language processing.....	69
2.6	SPEECH PRODUCTION AS THE MANIFESTATION OF LANGUAGE ....	70
2.6.1	Speech as a motor skill .....	70
2.6.1.1	Degrees of freedom.....	72
2.6.1.2	Motor equivalence .....	72
2.6.2	Temporal parameters of speech production.....	73
2.6.2.1	Segmental duration .....	74
2.6.2.2	Interarticulatory synchronization.....	75
2.7	CONCLUSION.....	77
2.8	SUMMARY OF CHAPTER TWO .....	78

## CHAPTER TWO

### SPEECH AND LANGUAGE PROCESSING: MODELS, THEORIES AND THE INFLUENCE OF CONTEXTUAL FACTORS

#### 2.1 INTRODUCTION

“The sound generated by a speaker is the product of coordinated multilevel motor processes...” (Smith, 1992a:233). Considering that three to five syllables are produced per second (Ramig, 1983), “it is evident that the nervous system has the complex task of simultaneously controlling and coordinating the movements of the articulators to produce rapidly alternating vocal tract configurations” (Smith, 1992a:233). The search for principles of neural organization which underlie this uniquely human behavior has been vast, leading to the development of various theories and models attempting to explain the processes involved in the motor control of speech and their neural substrates.

Speech is the conversion of language into sound and therefore cannot be completely separated from language. Kent, Adams and Turner (1996:33) underscore this fact by saying that “although one could become completely occupied in the study of speech as a motor behavior or a conversion of articulation into a sound pattern, speech is, after all, of greatest interest because of its primacy as a language modality” and that “one of the most exciting facets of a speech production model is what it can tell us about language”. However, investigations attempting to bridge the gap between speech and language processes have been few (Maner, Smith & Grayson, 2000). Similarly few scientists have addressed the question of how speech production relates to a more general model of language formulation (Kent *et al.*, 1996).

Another challenge facing models of speech production is the explanation of the context-sensitivity of speech. The latter refers to the fact that the production of speech sounds varies with the context in which they are produced. The phonemes that make up a word are not merely realized acoustically by the assignment of a set of preset muscle commands for each phoneme and execution of these commands in serial order (Kent *et al.*, 1996). On the contrary, the temporal and spatial parameters

of speech production need to be adjusted according to the phonetic context in which the sounds are produced (Van der Merwe, 1997). Apart from the phonetic context which exerts an influence on the parameters of speech production, the other identified contexts (Van der Merwe, 1986, 1997) also influence speech production and contribute to its complexity, as mentioned in chapter one. Van der Merwe (1997) emphasizes the importance of the study of the influence of these various contexts when studying the speech of persons with neurogenic speech and language disorders, since the context of speech production might influence the process of speech sensorimotor control (Van der Merwe, 1997). Van der Merwe (1997) states that variation of certain factors, such as sound structure, have already been found to cause variation in the symptoms of AOS (Kent & Rosenbek, 1983; Van der Merwe & Grimbeek, 1990; Van der Merwe *et al.*, 1987, 1988, 1989). The variation caused in apraxic symptoms by different contextual factors, will need to be taken into account when compiling assessment and treatment procedures for persons with neurogenic speech and language disorders.

Persons with neurogenic communication disorders can exhibit deficits related to language and/or motor processes of speech production. However, linguistic and motoric aspects cannot be completely separated in the study of speech production (Robin, Solomon, Moon & Folkins, 1997). Considering that speech is a modality for language, the languages of a bilingual speaker (L1 versus L2) could serve as two different contexts for speech production. Studying the effect of speech production in L1 versus L2 in bilingual speakers, on specific parameters of speech production, has the potential to shed light on the control and interaction of speech and language processing in the brain. Studying the influence of speech production in L1 versus L2 on temporal parameters in bilingual speakers with neurogenic speech and language disorders who exhibit deficits at distinct levels of the speech production process, will provide insight into the nature of the processing mechanisms involved in bilingual speech and language processing in the presence of a neurologic lesion. Insight into the nature of the processing mechanisms involved in bilingual speech and language processing, in turn, will inform about the nature of these disorders and how different disorders react to different contextual influences.

When studying speech and language processes in the bilingual speaker with neurogenic speech and language disorders, it is necessary to employ information and concepts from several fields of research to form a backdrop against which experimental questions can be investigated and results interpreted. These include concepts related to motor control and motor learning, since speech production is in essence a motor skill. Furthermore, speech is the acoustic realization of language processing in the brain and consequently, theories and models of speech and language production need to be perused. These models and theories provide the platform from where aspects such as bilingualism in neurogenic speech and language disorders can be investigated and results interpreted. Models of speech and language production also serve to explain the underlying mechanisms involved in normal and pathological speech and language processing. Neurophysiological, as well as behavioral accounts of speech and language processing need to be incorporated into these models in order to provide a comprehensive explanation for results obtained. Finally, concepts related to bilingualism need to be reviewed in order to understand the processes involved in bilingual language processing.

In this chapter, models of speech production which have been proposed to explain normal and pathological speech and language processing, will be reviewed. Specific factors or “contexts” (Van der Merwe, 1997) which exert an influence on speech and language processing will then be discussed with reference to an information processing perspective on speech production. These factors can be either inherent to the individual or related to the task or context. From this discussion, the importance of the study of speech and language processing in bilingual speakers will become evident and concepts related to bilingualism will be reviewed. Against this backdrop, the study of the influence of speech production in L1 versus L2 in bilingual speakers, as a context for speech production, will be proposed. Since speech is a motor skill, it will become evident that the control of the temporal and spatial parameters of speech is important for obtaining perceptually accurate speech. The parameters of speech production which are studied in an attempt to gain insight into the motor control strategies of the brain in normal persons and persons with neurogenic speech and language disorders will be reviewed. Specific temporal parameters of speech production which are studied to inform about speech and language processing in

general, will finally be discussed. This discussion will provide a rationale for the investigation of specific temporal parameters of speech and the influence of L1 versus L2 speech production on these parameters in speakers with neurogenic speech and/or language disorders.

## 2.2 THEORIES AND MODELS OF SPEECH PRODUCTION

The fact that speech is an extremely complex phenomenon has led to speech production being modeled in various ways. Neural models, articulatory models, vocal tract models, functional models and models of motor control have been proposed to account for various and diverse aspects of the speech production process (Kent *et al.*, 1996). In the next section, some of the most prominent models which have been related to normal and pathological speech production will be discussed.

The discussion will include, *inter alia*, two theories of motor control which have been related to speech production. Theories of motor control can be divided into two main groups depending on the emphasis placed on either the movement commands issued by the central components of the motor control system, or on the environment (Magill, 2001; Kent *et al.*, 1996). The two models of motor control which will be discussed are schema theory proposed by Schmidt (1975, 1988) and a dynamic system model proposed by Kelso and Tuller (1981). *Motor program theory* gives prominence to commands issued by the central nervous system, while *dynamic pattern theory* “gives more influence to movement commands specified by the environment and to the dynamic interaction of this information with the body, limbs and nervous system” (Magill, 2001:47).

A third model of speech production proposed by Levelt (1989) will then be discussed as a functional model of speech production, incorporating the concept of information processing which attempts to explain the ways various types of information regulate speaking. Finally, a model of speech sensorimotor control proposed by Van der Merwe (1997) will be reviewed, since this model is specifically aimed at describing the processes involved in normal speech production and speech production of persons with neurogenic speech and language disorders due to breakdown at different levels of the speech production process. Van der Merwe’s (1997) model incorporates the

concept of motor plans and programs, as well as the influence of contextual factors on speech production.

### 2.2.1 Motor program-based theory

The *motor program* forms the central part of theories that support central control of movement (Magill, 2001). Kent *et al.* (1996:14) define a motor program as “a plan or prescription of movement”. A motor program for movement implies that speech movements are available for execution in a pre-assembled form which directs the muscular and articulatory regulation during speech production. The concept of motor programming for speech has been criticized because of its “excessive rigidity, failure to account for corrections in the ongoing movement, and inability to assess the status of the periphery and take proper advantage of these initial conditions” (Kent *et al.*, 1996:15). According to these researchers, motor program theory would not be able to account for the adaptation of movement to the context.

Kent *et al.* (1996) propose that an approach that may avoid some of the objections to the concept of motor programs is a generalized motor program (GMP) or schema theory as proposed by Schmidt (1975, 1982, 1988). Schmidt (1988) proposes two control components which are involved in the learning and control of skills, namely, the *GMP* and the *motor response schema*. The *GMP* is responsible for controlling the general characteristics of classes of actions, such as, kicking, walking and running, whereas the *motor response schema* provides the specific rules governing an action in a given situation, thus providing parameters to the GMP. According to Schmidt’s theory (1975, 1982, 1988), motor programs are composed of *schemata*, which refer to “learned relationships among movement outcomes, control signals and boundary conditions” (Kent *et al.*, 1996:16). The schemata are based on four kinds of information which is stored, namely, the initial conditions, the outcome, the sensory consequences and the parameters used during execution of the movement. After each response, this information is stored and the relationships among them abstracted (Schmidt, 1982). These schemata are strengthened by experience with the consequence that practice in situations with different motoric requirements will enhance a particular schema (Kent *et al.*, 1996; Schmidt, 1982, 1988).

According to Schmidt (1988) the GMP controls a class of actions rather than a specific movement or sequence. A class of actions refers to a range of different actions that have a common but unique set of features. These invariant features form the basis of what is stored in memory. When a person wants to perform a specific action in a specific situation, the appropriate motor program (GMP) needs to be retrieved and movement-specific parameters then need to be added in order to meet the specific demands of the situation or context. Invariant features include the *relative time* of the components of the skill, the *relative force* used in performing the skill and the *order or sequence* of the components.

Although the GMP has invariant features which are invariable from one performance of a skill to another, the *parameters* which are applied can be varied. The parameters include overall force, overall duration and the muscles that must be used in execution of the skill. These aspects are adapted according to the requirements of the *situation or context* (Schmidt, 1988; Shea *et al.*, 1993). Schmidt's theory accounts for the performance of novel actions. A person can thus use rules from the motor response schema and add appropriate parameter characteristics to the GMP to perform a novel action (Magill, 2001). In Schmidt's theory, the importance of adapting the movements to the context is evident, although the context is not the primary determinant of the movements which occur in this view of motor control.

- **Schmidt's schema theory related to speech motor control**

Schmidt's schema theory can also be applied to speech production. In this regard, Kent *et al.* (1996:17) state that "The *motor response schema* generates a motor program based on the desired outcome and the initial conditions", where the desired outcome refers to the desired goal of movement and the initial conditions to the initial position of the articulators in space. The desired outcome, for example, a spatial target for a phonetic unit and the initial conditions (the current state of the speech system, including the positions of the articulators) serve as the inputs to the speech production mechanism and are needed for response recognition as a motor response schema. Feedback signals in the form of efference copy, proprioceptive feedback and auditory feedback are generated. These are thus the expected sensory consequences based on previous experience and outcomes. Actual feedback signals are compared

with expected feedback once the movement is in progress and discrepancies between these two are used to adjust the motor response schema. The motor response schema can thus be adapted to deal with novel responses (Kent *et al.*, 1996).

### 2.2.2 Dynamic Systems Models

Dynamic systems theory, also known as task dynamics and action theory (Kent *et al.*, 1996) proposes that the motor control system can constrain the degrees of freedom by acting as a single unit (Bernstein, 1967; Tuller, Fitch & Turvey, 1982). In this theory “motor behavior is viewed in terms of the interactions between biomechanical and environmental variables” (Kent *et al.*, 1996:19). Dynamic systems theory describes and explains the control of coordinated movement by emphasizing the role of information in the environment and the dynamic properties of the body and limbs (Magill, 2001; Tuller *et al.*, 1982).

In essence, this theory proposes that the degrees of freedom of a motor system can be constrained by groups of muscles functioning as functional units, termed coordinative structures (Tuller *et al.*, 1982). Muscles are thus not controlled individually, but independent muscles are constrained to work as a functional unit or coordinative structure (Shea *et al.*, 1993). The main idea that proponents of this view wish to convey is that actions cannot be considered independent of their context (Kelso & Tuller, 1981). These researchers state that "While a coalitional style of control embodies the advantageous characteristics of heterarchies, namely free dominance, reciprocity and distributed function, it possesses the additional control advantage of effectively reducing the degrees of freedom of the system" (Kelso & Tuller, 1981:232). In their view, control of the degrees of freedom is "accomplished by the contextual framework that operates as a constraint on possible movement".

Proponents of dynamic systems theory, postulate that the contextual constraints specify the parameters of the motor system. The movement itself thus represents the context in which the action takes place and movement originates from the functional coalition of various structures. Depending on the movement to be executed, a coalition between the neural and anatomic structures, such as the muscles, is established. Movement is the result of the selection of synergies (coordinative

structures) for a particular action. The coalitional style of organization is in direct opposition with the hierarchical models of control (Kelso & Tuller, 1981; Kelso *et al.*, 1983).

Magill (2001) states that both the GMP and dynamic pattern theory include the concept of relative time invariance as a characteristic. The source of the invariance differs, however, in that in the GMP relative time is an invariant feature of the GMP and is included in the movement commands that are sent to the muscles. Relative time invariance in this case is indicative of a class of movements that are controlled by the same GMP. The dynamic pattern view on the other hand prefers the term “temporal pattern”. In this view the invariable temporal pattern is the result of the interaction between the person and the characteristics of the task and/or environment or due to the mechanical dynamics involved in the body and limb movements themselves (Magill, 2001).

- **Dynamic systems theory related to speech motor control**

Kent *et al.* (1996) apply the dynamic systems theory to speech production in the following way. Dynamic systems theory proposes that a complex system, such as the speech production system, can be simplified when the individual components of the system are functionally linked. The degrees of freedom are reduced by these functional groupings or synergies among the muscles which comprise the system. The effective control of the system is accomplished by appropriate combinations of the synergies. These synergies are “task specific, context sensitive, and adaptive” (Kent *et al.*, 1996:19). They possess both essential and nonessential parameters. Essential parameters are qualitative aspects of a movement’s structure, for example, lip closure for the bilabial stop consonant /b/, whereas nonessential parameters are quantitative, for example, differing displacements of the lower lip in the bilabial closure movement when variations are introduced in phonetic context, stress or speaking rate (Kent *et al.*, 1996).

According to Kent *et al.*, (1996:19) when this is applied to the speech production process, essential parameters can account for phonetically distinctive characteristics of movements, while nonessential parameters can account for the effects of stress, rate

and other “scalar variables that operate within the phonetic requirements of the movements”. When producing a word, a general form of the intended action is contained in the phonological prescription. The coordinate grouping, among the many possible elements, determines the details of the motor events. Equations of constraint specify how the group members/muscles interact within the limits of the particular action and its context. Coordination in this view is described as “a blending of the dynamics of the participating synergies” (Kent *et al.*, 1996:20). In motor program theory, coordination is preassigned by a motor program, whereas in dynamic systems, coordination results from the system dynamics. Dynamic system theory interprets invariant phase relations between task-coupled articulators during speech production as indicative of the operation of these functional synergies (Kent *et al.*, 1996).

Kent *et al.* (1996) ascribe the following disadvantages to dynamic system models. These models do not clearly link responses with phonological input to speech production. Furthermore, some predictions about phase relations have not been confirmed experimentally. Acoustic and language-specific timing factors are also neglected in this theory. According to these researchers advantages include, minimizing of the degrees of freedom and recognition of biomechanical properties of the system in relation to tasks.

Neither dynamic systems theory nor motor program models specify how speech production progresses from an intention to communicate to the achievement of perceptually accurate speech. In other words, these models do not explain how the linguistically formulated message is transformed into an acoustic signal. Apart from the motor processes involved in speech production, it is necessary to recognize the importance of the cognitive and language processes which occur in the brain. The model of speech production proposed by Levelt (1989) emphasizes the cognitive and language processes involved in speech production and will be reviewed in the following section.

### 2.2.3 An information-processing model: Levelt's theory

Since speech is one way to express language, it is evident that speech production cannot be viewed exclusively from a motor control perspective. A number of interactions and conversions take place in the process of going from thought to speech. Speech conversions involve neuromotor, myomotor and articulatory rules, whereas language conversions involve linguistic rules (Netsell, 1982). Consequently all of these aspects need to be incorporated in a model of speech production. Smith (1992a:263) posed an important question related to this issue by asking, "Can a general motor theory account for speech production, or must the linguistic elements of speech ultimately be intimately intertwined with motor processes at every level of the production process?". According to Smith (1992a:263) this debate is unlikely to be resolved soon and provides "a special vigor to the study of speech movement control".

One of the many approaches that could be taken in the study of motor learning and control and in the explanation of speech as the externalization of language processing which occurs in the brain, is an information processing approach. In this view, humans are regarded as active processors of information rather than passive recipients (Shea *et al.*, 1993). The basic assumption of this approach is that a number of cognitive processes are required for correct execution of movement by an individual. Furthermore, the fact that responses can vary in different situations, underscores the influence of contextual factors on speech and language processing. Current circumstances and past experiences that are stored in memory are considered when planning and executing a response in a specific situation (Shea *et al.*, 1993; Stelmach, 1982). Abbs (1988:168) states "the neurophysiologic mechanisms of speech control cannot be separated functionally from the less well-understood neural processes involving motivation, attention, conceptualization and sensory awareness, all of which have been implicated increasingly as influential factors in motor execution processes".

An information processing approach to the study of movement, examines the mental operations which occur after a stimulus has been received and a response has been initiated (Stelmach, 1982). This approach could be especially informative when

compiling models of speech and language processing, since language formulation for speech production involves mental processing which is largely inaccessible to man. The speaker can be viewed as “a highly complex information processor who can, in some still rather mysterious way, transform intentions, thoughts, feelings into fluently articulated speech” (Levelt, 1989:1). Levelt (1989:1) proposes that “Developing a theory of any complex cognitive skill requires a reasoned dissection of the system into subsystems, or processing components”.

According to Levelt (1989), a theory of speech production will involve various processing components. There are thus various processing systems that underlie speech, in other words, “which translate the speaker’s intentions into overt speech” (Levelt, 1989:1). These processing components are specialized and work in a rather autonomous manner. Levelt (1989) underscores this issue by saying that the processing components which underlie speech work in a “highly automatic, reflex-like way” which allows these components to work in parallel. Parallel processing is, according to Levelt (1989), a prerequisite for uninterrupted fluent speech.

Levelt’s (1989) theory of speech production incorporates aspects of information processing and is one of the few theories of speech production which have attempted to explain how language is converted to speech. Levelt (1989) proposes the following processing components to be involved in speech and language processing, namely, the Conceptualizer, Formulator, Articulator, Audition processing component and Speech-Comprehension System. The mental activities involved in speaking require the person’s conscious attention and include conceiving of an intention, selecting the relevant information to be expressed, ordering this information, monitoring of one’s own productions and keeping track of what was said during a conversation. Levelt (1989) refers to these mental activities as conceptualizing and to the subserving processing system as the Conceptualizer. The product of this processing component is termed the preverbal message. This message needs to be encoded by the speaker by making use of both procedural and declarative memory. Paradis (1995a, 1998) proposes that procedural memory underlies implicit linguistic competence, which entails the incidental acquisition and the automatic use of language, while declarative memory subserves metalinguistic knowledge which is acquired consciously and

stored explicitly in the brain. The output of the Conceptualizer is the input to the following processing component, namely, the Formulator.

The Formulator translates a conceptual structure into a linguistic structure through grammatical and phonological encoding. The grammatical encoding procedures deposit their interim results in the Syntactic Buffer. Phonological encoding entails retrieving or building a phonetic or articulatory plan for each word and for the utterance as a whole. The output of this processor is a phonetic or articulatory plan, which is the input to the next processing component, namely, the Articulator (Levelt, 1989). The articulatory plan is a “program for articulation” (Levelt, 1989:12). It is not yet overt articulation, but an internal representation of how the utterance should be articulated. Levelt (1989:12) refers to this representation as “internal speech” and to articulating as “the execution of the phonetic plan by the musculature of the respiratory, the laryngeal and the supralaryngeal systems”. The phonetic plan is stored in the Articulatory Buffer, since internal speech may be ahead of articulatory execution. During motor execution sets of muscles are used in a coordinated way in order to achieve overt speech. According to Levelt (1989) the articulatory plan is relatively independent of context, but execution adapts to varying circumstances or articulation in order to achieve roughly the same articulatory goal.

According Levelt (1989), the speaker is able to monitor correctness of his production through the Speech-Comprehension System. This system can monitor both internal and overt speech. Trouble in internal speech can be detected before the utterance has been completely articulated. The Speech-Comprehension System’s output is parsed speech which entails a representation of the input speech “in terms of its phonological, morphological, syntactic, and semantic composition” (Levelt, 1989:13). The Speech-Comprehension System can detect form errors during production or if speech differs from the intention, in other words, differences between overt speech and the intended speech target.

Although Levelt (1989) proposes different stages to be involved in the progression from an intention to communicate to overt speech or articulation, he does not go into much detail regarding the motor planning and or programming stages of speech production. However, Levelt’s (1989) theory of speech production has potential to

explain speech and language deficits from an information processing perspective. The framework of speech sensorimotor control proposed by Van der Merwe (1997) describes the various linguistic and motor stages of speech production in detail and will be reviewed in the following section.

#### **2.2.4 Van der Merwe's four-level framework of speech sensorimotor control**

Van der Merwe (1997) proposed a framework for the sensorimotor control of speech in which the speech production process is depicted as consisting of four stages. This model is different from traditional models of speech production that propose that the speech production process consists of only three stages, namely linguistic encoding (semantic, syntactic and phonological), articulatory programming and execution (Itoh & Sasanuma, 1984). The most important difference is the distinction which Van der Merwe (1997) makes between motor programming and motor planning of speech. Within Van der Merwe's framework, motor planning and programming are viewed as two separate stages in the speech production process. The addition of another component to the speech production process has important implications for the study of speech motor control, since it implies that temporal specification and control can be exerted on more than one level of the speech production process.

Van der Merwe (1997) ascribes the control of each stage of the speech production process, to a coalition of specific neural structures. The components in the framework proposed by Van der Merwe (1997) will be discussed in detail, since this framework includes the most significant components of various models which have fallen short in explaining normal and pathologic speech motor control. Van der Merwe (1997) relates the various stages involved in speech production to specific neuroanatomic regions.

##### **2.2.4.1 Intention to communicate verbally**

The first "event" in the speech production process, yet not one of the formal stages proposed, is the intention of the speaker to communicate verbally. This aspect of speech production is posed to be controlled by the frontal-limbic formations of the forebrain. The limbic system generates the emotional motivation to act. The

intention to communicate is, however, distinct from the actual initiation of speech movements.

#### **2.2.4.2 Linguistic-symbolic planning**

The first formal stage in the speech production process as postulated by Van der Merwe (1997) consists of the linguistic-symbolic planning of the utterance. During this stage, selection and sequencing of the phonemes take place, governed by the phonotactic rules of the language. This stage entails the syntactic, lexical, morphological and phonological planning of the intended utterance. It is evident from this description that this stage of the speech production process is non-motoric in nature. It further implies that the linguistic and motor planning of speech are performed at two distinct levels of the speech production process. A deficit at the level of linguistic-symbolic planning would thus result in phonologic errors, such as, PPs that occur in certain types of aphasia, for example, CA (Van der Merwe, 1997). The linguistic-symbolic stage proposed by Van der Merwe (1997) appears to be similar to the function performed by the Formulator proposed by Levelt (1989).

#### **2.2.4.3 Motor planning**

In order for the phonemes to be actualized at the articulatory level, a transformation of the phonemic representation of the utterance to a code that can be interpreted by the motor system has to take place. During this phase the motor goals for the actualization of the utterance are specified. Van der Merwe (1997) presents a hypothetical description of motor planning and this will be reviewed briefly.

Motor planning commences with the feedforward of the invariant phonological units in sequence to the motor planning system. Van der Merwe (1997) postulates that during the motor planning of the utterance the motor goals for each phoneme are specified in terms of spatial and temporal characteristics. The invariant core motor plan with these spatial (place and manner of articulation) and temporal specifications of movements for each phoneme is recalled from the sensorimotor memory where it is stored. Van der Merwe (1997) emphasizes the fact, however, that although the recalled core motor plan is invariable, adaptations to this core motor plan need to be

made during the motor planning of speech, depending on the phonetic context in which it is to be produced. Certain factors thus necessitate adaptations to be made to the temporal and spatial parameters specified in the core motor plan. Van der Merwe (1997) postulates these factors to include aspects, such as, the phonetic or sound context (Borden & Harris, 1984), coarticulation possibilities (Borden & Harris, 1984; Kent & Minifie, 1977), phonetic and linguistic influences on segmental duration and changes in speech rate (Gay, 1981; Kelso *et al.*, 1983).

The core motor plan for each phonological unit is presumably similar to the GMP proposed by Schmidt (1982, 1988). The adaptation which takes place depending on the context in which production takes place is presumably similar to the parameterization which is suggested by Schmidt (1982, 1988). Schmidt (1982, 1988) does not specify, however, how language formulation processes and linguistic planning of the utterance occurs, nor does he specify the influence of specific speech contexts, although he mentions that spatial and temporal parameters are adjusted depending on the context of production.

Van der Merwe's model thus postulates that adaptation of the core motor plan for each phoneme within the context of the planned unit has to be made. This then includes adaptation of spatial specifications to the phonetic (sound) context and rate of production and also adaptation of temporal specifications to segmental duration, speech rate, coarticulation potential and interarticulatory synchronization (IAS). All this is done within the boundaries of equivalence, guided by knowledge of the acoustic effect of movements and a representation of the acoustic configuration to be reached. During adaptation of the core motor plan response feedback is not yet available since the movement has not been actualized at the moment of planning. However, internal feedback and predictive stimulation presumably guides adaptation (Van der Merwe, 1997). The motor system takes the initial conditions into account which is consistent with the dynamic systems view of motor control and also with Schmidt's (1982, 1988) proposal that the initial conditions form part of the information from which the motor response schema is derived.

The different subroutines that constitute the adapted motor plan (such as lip rounding and velar lift) are then specified and temporally organized. Van der Merwe (1997)

states that motor goals, such as, lip rounding, jaw depression or glottal closure need to be specified. Interarticulatory synchronization is also planned for a particular phoneme. These temporally arranged structure-specific motor plan subroutines are then systematically fed forward to the motor programming system. This step then concludes the motor planning phase. Van der Merwe (1997) emphasizes the fact that motor planning is articulator-specific and not muscle-specific.

It is evident from the motor planning stage that specification of temporal and spatial parameters of movement is crucial for obtaining on-target acoustic output. Temporal parameters, such as, the duration of movements of the articulators and IAS are specified during the motor planning of speech in order to obtain a desired acoustic output and consequently to achieve an accurate perceptual goal. If the motor planning stage is disrupted, distorted articulation could result, for example, either due to aberrant timing between the movements of the articulators or because of aberrant spatial goal specification or achievement. Examples of other speech characteristics which are the result of difficulty regarding the motor planning of speech include, slow speech or struggling behavior. The aforementioned speech characteristics which result from difficulty regarding the motor planning of speech are characteristic of AOS (Van der Merwe, 1997).

#### **2.2.4.4 Motor Programming**

Traditional models of speech production do not distinguish between motor planning and motor programming of speech, but use these two terms as synonyms. This is where the framework proposed by Van der Merwe (1997) differs from most traditional models of speech production. Whereas motor planning of speech refers to the planning of the temporal and spatial goals of the articulators, motor programming refers to the selection and sequencing of motor programs for the movements of the individual muscles of these articulators (including the vocal cords). During motor programming of speech these muscle specific programs are specified “in terms of spatial-temporal and force dimensions such as muscle tone, rate, direction and range of movements” (Van der Merwe, 1997:16). During this phase, sensory feedback is potentially available to update motor programs, while internal feedback controls programming. All the neural structures involved at this stage are supposedly involved

in internal feedback. Repeated initiation and feedforward of co-occurring and successive motor programs and integration with respiration for speech concludes the motor programming of speech (Van der Merwe, 1997).

Motor planning is thus a phase prior to motor programming which refers to the planning of motor goals of the articulators (spatial, as well as temporal), while motor programming is more specific and refers to the temporal and spatial specifications for each individual muscle. To achieve accurate temporal and spatial movements of the articulators, the activity of the muscles also needs to be temporally and spatially controlled. DeLong (1971) states that for the accomplishment of synchronized movement the appropriate muscles need to be selected, these muscles then need to be activated/inhibited in the correct temporal relationship and lastly the correct amount of excitation for each muscle needs to be applied. Temporal and spatial specification of muscle movement is thus performed during the motor programming of speech.

#### **2.2.4.5 Execution**

Execution refers to the actual realization of speech on the articulatory level. At this stage the temporal and spatial parameters of speech have been specified during motor planning and programming and are realized on the acoustic level. During the execution of movement closed-loop tactile-kinesthetic feedback is supposedly available for control and acoustic feedback is also implemented. Although response feedback is available, it is not necessarily constantly utilized during speech production of the mature speaker (Van der Merwe, 1997).

#### **2.2.5 Conclusion regarding models and theories of speech production**

From the above review of prominent models and theories of speech production, it is evident that speech can be viewed from a motor control perspective due to the fact that speech is essentially a learnt motor skill. However, speech production cannot be isolated from the language processes which occur in the brain, since “speech is constrained not only by the task dynamics of its production system and the articulatory-acoustic relations of the vocal tract, but also by its service to language”

(Kent *et al.*, 1996:33). A model or theory of speech production must thus account for the conversion of an abstract language code into movement parameters resulting in overt speech.

No matter from which perspective the speech production process is viewed, it involves complex parallel processing by various components to finally achieve perceptually accurate speech. From the models and theories discussed above, it is evident that speech and language processing is not isolated from external influences, however, and that certain factors exert an influence thereon. Van der Merwe (1997) emphasizes this by saying “contextual factors affect the dynamics of motor control by exerting an influence on the mode of coalition of neural structures involved during a particular phase and on the skill required from the planning, programming, and execution mechanisms”. Van der Merwe (1997) identifies specific contextual factors which exert an influence on speech and language processing. The contextual factors proposed by Van der Merwe (1997) will be discussed in more detail further on in this chapter.

### **2.3 FACTORS INFLUENCING INFORMATION PROCESSING**

Before discussing factors or contexts which exert an influence on speech and language processing specifically, it is necessary to peruse factors influencing information processing in general, since they are relevant during execution of any given task. Consequently these factors might also impact on speech and language processing. These factors are important, since they might influence the processing load, which McNeil *et al.* (1991a:35) refer to as “the idea that the more complex or difficult the task, the greater is the processing load and the outlay of effort”. Factors influencing information processing and consequently the processing load, will consequently need to be considered when investigating speech production in both normal speakers and speakers with neurogenic speech and language disorders to determine their effect on normal and pathologic speech and language processing. Factors influencing information processing can either be related to the task or context or inherent to the speaker.

## **2.3.1 Factors influencing information processing related to the task and context**

### **2.3.1.1 Stages of learning**

According to Shea *et al.* (1993) performance changes as a movement becomes better learnt. The manner in which the brain processes information, also changes from when the task was first introduced compared to after the task has been practiced several times. Shea *et al.* (1993) mention three stages of learning, namely, the cognitive stage, associative stage and autonomous stage. During the cognitive stage, high demands are placed on sensation and perception processing. The person needs to determine the objective of the skill and take into account the environmental cues that control and regulate the movement. At this stage, demands on response execution are low, since it is too soon to concentrate on refinement of the movement.

During the associative stage the skill is performed and refined and concentration or attention shifts to the task and response execution. The information processing load appears to become reduced and the person becomes better able to attend to other stimuli which are not related to the task. The final stage of learning is referred to as the autonomous stage. This stage results in a nearly automatic kind of performance where the person can attend to other tasks while performing the primary task, for example, driving a car and having a conversation at the same time. Information processing activities thus change as a result of practice (Shea *et al.*, 1993).

### **2.3.1.2 Automaticity**

Speaking is an intentional activity serving the purpose the speaker wants to realize (Levelt, 1989). An intentional activity is believed to be under central control, however (Bock, 1982). Levelt (1989) poses that a speaker invests his attention on matters, such as, his state of motivation, his obligations, what has previously been uttered or what has happened previously, and so forth. The question arises regarding the extent to which the processing components are under central or executive control. If a component is not centrally controlled, its functioning is implied to be automatic in nature. An automatic process is executed without attention or conscious awareness.

It runs on its own resources and does thus not share processing capacity with other processes (Levelt, 1989). Automatic processes are believed to be quick and even “reflex-like” and the structure of such a “process is ‘wired-in’, either genetically or by learning (or both)” (Levelt, 1989:20). Automatic processes can run in parallel without mutual interference, since they do not share resources. Each processor can work on different bits and pieces of the “utterance under construction” (Levelt, 1989:24). This is referred to as incremental processing (Kempen & Hoenkamp, 1987) and is based on the concept of automaticity, since only automatic processors can work without sharing access to attentional resources.

The concept of automaticity is important in the study of speech production, since most of the processing stages involved in normal speech production are generally viewed as occurring in a fairly automatized fashion (Bock, 1982; Kent, 1990; Levelt, 1989). Although the speaker needs to think about what he is going to say, he is mostly unaware of the linguistic and motor processing which occurs in translating thought into overt speech. Levelt (1989:22) states that the processes of the Formulator and Articulator are “probably largely impenetrable to executive control even when one wishes otherwise”. When the speed at which these processes need to take place to achieve fluent speech which is produced at a rate of approximately fifteen phonemes per second (Levelt, 1989) is considered, it becomes evident that conscious processing by these two components (the Formulator and the Articulator) is not feasible.

Borden and Harris (1984) also underscore the fact that although a person is generally conscious of the message he wants to convey, as well as the search for the appropriate words and feelings towards the topic or listener, a person is seldom aware of the processes involved in sound production as such. According to these researchers, sound production only comes to one’s attention when attempting new or unfamiliar words or in unfamiliar circumstances, for example, when speaking with a new dental appliance. Novel responses are thus not yet as automatized as over-learned responses.

In contrast to automatic processing, controlled processing places demands on attentional resources, with the implication that only a limited number of things can be attended to at a time (Levelt, 1989). Controlled processing requires capacity in working memory (Kent, 1990). The processes which are placed in working memory

(to be discussed later on) require a certain level of awareness. In Levelt's (1989) model of speech and language processing, the Conceptualizer requires highly controlled processing, because communicative intentions can vary in infinite ways. According to Levelt (1989) message construction and monitoring are thus subject to controlled processing. In this sense a speaker can attend to his own internal or overt speech and is aware when self-corrections are made. However, only a few concepts or bits of internal speech are available for conscious processing in working memory at a time, however.

Kent (1990) proposes that speakers exhibit flexibility regarding the employment of automatic or controlled processing. Marginal forms of control are hypothesized to be present as evidenced by the fact that a speaker can interrupt his speech when an error is detected in order to correct it. It is proposed that other global aspects of processing, for example, speaking rate, loudness and articulatory precision can also be controlled by the same executive signal. More attention needs to be allocated during the performance or implementation of these parameters according to Levelt (1989). According to Kent (1990), controlled processing is likely to be employed in more challenging situations. Aspects proposed to specifically influence and presumably increase the processing demands during speech production specifically, will be discussed in more depth further on in this chapter.

### **2.3.1.3 Movement time**

The information processing load is increased when the time available to complete a movement is reduced. When executing rapid movements, all information processing must be completed before the movement is begun (Shea *et al.*, 1993). Speech production is the result of "rapidly changing vocal tract configurations" (Smith, 1992a:233). In certain situations a speaker might be required to execute speech at a faster than usual rate due to time constraints, for example. Schmidt (1975) proposes that rapid movements need to be preprogrammed. A motor program implies a set of prestructured commands that are able to control the movement from beginning to end. These programs are presumably stored in memory and once retrieved do not require active information processing related to the construction of an action plan. The existence of these motor programs in memory significantly reduces the information

processing load, since active information processing related to action plan construction does not occur (Shea *et al.*, 1993).

As discussed, the concept of motor plans or programs (Schmidt, 1982, 1988; Van der Merwe, 1986, 1997) has also been proposed in various models of speech production. A faster than normal speaking rate will presumably place higher demands on the speech production mechanism and will consequently influence both speech and language processing. Speaking rate as a context for speech production will be discussed in more depth further on in this chapter.

#### **2.3.1.4 Movement complexity**

An interval of time in which the motor control system is prepared according to the demands and constraints of the situation/context precedes the intended action. In other words, preparation of the motor control system is required when performing voluntary coordinated movement. Certain actions and circumstances require more preparation than others. The task itself, the situation and personal factors influence the time it takes to prepare the motor control system. The complexity of the action to be performed influences the amount of time the person requires to prepare the motor control system. The number of parts to a movement in turn determines the movement complexity. Furthermore, the more accurate the movement must be the longer the preparation time (Magill, 2001). From the above, it is evident that processing is influenced by movement complexity, since this determines the preparation time the motor control system requires.

#### **2.3.1.5 Environment**

Regarding general motor skills, the environments in which movements are executed can be classified on a continuum ranging from closed to open environments (Magill, 2001; Rose, 1997; Shea *et al.*, 1993). Conditions in a closed environment are relatively stable and ample processing time is available in this context. Processing demands are generally placed on sensation-perception and response selection. Open environments exhibit continuously changing conditions and information processing

has to occur at a fast rate and anticipation needs to be employed for response selection (Shea *et al.*, 1993).

In the same way that the nature of the environment can influence processing demands involved in motor actions, the conditions under which speech needs to be produced can influence speech and language processing. Although environments for speech production cannot be described as open or closed per se, certain contexts in which speech is produced are also more challenging than others and consequently influence the processing demands. An example of an “environment” in speech production might be related to the phonetic environment in which a phoneme is produced. The presence of a compromised speech motor system due to, for example, a neurologic insult might also be considered as causing changing conditions for speech production. This is a condition inherent to the speaker, however, and not brought about by the environment. All the other contexts proposed for speech production might also be seen as different environments for speech production. The contexts proposed to influence speech and language processing will be discussed further on in this chapter.

### **2.3.2 Factors influencing information processing inherent to the individual**

Certain factors regarding information processing are inherent to the person or speaker. These factors might differ between persons and might be affected in persons with damage to the central nervous system.

#### **2.3.2.1 Attention**

Since the earliest days of investigating human behavior, the study of *attention* has been of great interest. It is one of the most significant limitations influencing human learning and performance (Magill, 2001). Attention is also an important concept in the study of speech and language processing. As discussed, the amount of attention which needs to be allocated decreases as the task becomes over-learned/automatic. This is evident from fluent speech production where the speaker does not need to think about speech production as such. However, certain speaking situations/contexts, as will be discussed, place higher demands on speech and language processing and

require allocation of more attentional resources. Therefore, this concept needs to be incorporated and understood in a discussion of speech and language processing.

Shea *et al.* (1993:312) define attention as “The direction of mental energy or the allocation of resources to important stimuli and ignoring irrelevant ones; the process by which we notice important, meaningful, or relevant information and ignore unimportant stimuli”. McNeil *et al.* (1991a:30) note that attention is synonymous to “resources, capacity, or effort”. Magill (2001) emphasizes the fact that attention can be directed toward perceptual, cognitive and/or motor activities.

Limitations exist regarding the number of activities which can be attended to at one time (Kahneman, 1973; Magill, 2001; Stelmach, 1982). Attention theories propose that attention limits are the result of the limited availability of resources that are needed to carry out information-processing functions. We thus have limited attention resources to do all the activities that we may attempt at one time. One is able to perform several tasks simultaneously as long as the resource capacity limits of the system are not exceeded. If these limits are exceeded, performance of one or more of these tasks will deteriorate (Just & Carpenter, 1992; Magill, 2001; Stelmach, 1982).

Magill (2001) states that theorists have opposing views regarding the nature of resource limitations. The one group poses that there is one central resource pool from where all attentional resources are allocated. The latter is known as central resource capacity theories (Kahneman, 1973). The other group proposes multiple sources for resources and is known as multiple resource theories (Wickens, 1992). These two theories regarding resources will be discussed briefly in the following section.

#### **2.3.2.1.1 Central resource capacity theories**

Central resource capacity theories propose that there is a central reserve of resources from which all activities draw when being performed. An example of the central resource theory is proposed by Kahneman (1973). According to Kahneman (1973), the amount of available attention can vary depending on certain conditions related to the individual, the task at hand and the situation. The available attention is viewed as

a general pool of effort, which involves the mental resources necessary to carry out activities. Attention can be allocated to several activities at the same time. The allocation of resources is determined by factors such as the characteristics of the activities, as well as the allocation policy of the individual. This in turn is influenced by situations internal and external to the individual. This central pool of available resources is the available capacity and can fluctuate according to the arousal level of the person (Magill, 2001).

Arousal refers to the general state of excitability of a person which involves physiological, emotional and mental systems. A too high or too low arousal level will influence the available attention capacity. To exhibit the maximum attentional resources, the person must be at an optimal arousal level (Magill, 2001). A person with a compromised central nervous system might not always be at an optimal arousal level and this could lead to fluctuations in the ability to effectively allocate attention (McNeil *et al.*, 1991a). It has also been proposed that subjects with aphasia might have difficulty with effective allocation of resources (McNeil *et al.*, 1991a).

#### **2.3.2.1.2 Multiple resource theories**

Multiple resource theories propose that we have several attention mechanisms which each have limited resources. Each mechanism is related to a specific information processing activity and is limited as to how much information it can process simultaneously. Here each resource pool is specific to a component of skill performance. Wickens (1980, 1992) proposed the most popular theory regarding the multiple resource theories. According to Wickens (1980, 1992), three different sources are available as resources for processing information. These include the *input and output modalities* (for example, vision, limbs and speech systems), the *stages of information processing* (for example, perception, memory encoding, and response output) and the *codes of processing information* (for example, verbal codes and spatial codes). Depending on whether two tasks demand attention from a common resource or from different resources, they can be performed simultaneously. In other words, when a resource is shared between two simultaneously performed tasks, performance will decrease compared to when the two tasks compete for different resources.

### **2.3.2.1.3 Attention and context**

From the foregoing discussion one can conclude that the allocation of attentional resources is dependent on the context in which actions are executed. The higher the processing demands induced by the context, the more attentional resources will be needed for accurate performance in the specific context (McNeil *et al.*, 1991a). On the other hand, McNeil *et al.* (1991a:35) state “when the task is more automatic, the processing load is smaller and fewer resources, less attention, and less effort are required for its successful completion”.

Depending on the context in which speech is produced, demands can be placed on the processing involved in any of the stages in the speech production process. Since speech and language are intertwined, increased processing demands on one level of processing might also affect processing at other levels of the speech production process. Specific “contexts” for speech production (Van der Merwe, 1986, 1997) which are believed to increase processing demands during speech and language processing will be discussed further on in this chapter. The way in which persons with different speech and language disorders react to the increased processing demands has the potential to inform about the underlying nature of the specific disorder.

### **2.3.2.2 Memory**

Memory plays an important role in virtually any activity we perform. Consequently memory storage and retrieval influences learning and performance (Magill, 2001). We are continually confronted by situations which require the use of memory to produce a response. Magill (2001:143) describes memory as “(a) our capacity to remember or be influenced by past experiences; (b) a component of the information processing system in which information is stored and processed”. Shea *et al.* (1993:52) refer to memory as the “system that enables us to retain information over time”. In the speech production process, memory is important regarding both speech and language processing. Regarding language processing, rules regarding language content, form and use are learnt and stored as implicit linguistic knowledge (Paradis,

1998). These rules are then applied during message formulation and comprehension. Regarding speech production, the motor plans/programs are also presumably stored in and retrieved from memory when the speaker wants to convey his message orally (Van der Merwe, 1997).

Baddeley (1986, 1992) proposed that memory comprises two functional components namely, *working memory*, also known as *short-term memory* and *long-term memory*.

### **2.3.2.2.1 Working memory or short-term memory**

Working memory is active in all situations which require the temporary use and storage of information and the execution of memory and response production processes. Information is stored for a short time in working memory. Critical information processing occurs in this part of memory and it is important in decision-making, problem solving, movement production and evaluation, as well as in long-term memory functions (Magill, 2001). It provides essential processing activity needed for the transfer of information into the long-term memory. Working memory also serves as an interactive workspace where information retrieved from long-term memory is integrated with information in working memory (Baddeley, 1992; Just & Carpenter, 1992; Magill, 2001; Rose, 1997). Working memory contains all the information we attend to and are conscious of at a specific point in time (Levelt, 1989; Rose, 1997). According to Levelt (1989) working memory is active during the processes involved in conceptualizing and monitoring of speech.

Working memory is important in the study of speech and language processing, since aspects of speech and language processing which need to be attended to are placed in working memory. When speech and language processing demands increase, certain aspects might require more attentional resources and presumably occupy space in working memory. According to Levelt (1989) parsed internal speech is also represented in working memory. A person thus has access to his internal and overt speech. In her information processing model of verbal formulation and speech production, Bock (1982) proposes that in normal fluent adult speech, syntactic, semantic and phonological processing do not require capacity in working memory, since these are proposed to be automatic processes (Kent, 1990; Kent & McNeil,

1987). As was mentioned, however, a normal speaker is flexible in his deployment of controlled processing (Kent, 1990; Levelt, 1989). In other words, controlled or conscious processing might be needed and implemented under circumstances of increased processing demand and consequently require space in working memory.

#### **2.3.2.2.2 Long-term memory**

Magill (2001:147) describes long-term memory as “a component of the structure of memory that serves as a relatively permanent storage repository for information”. Three systems in long-term memory have been proposed, namely, procedural, episodic and semantic memories. In more recent years, researchers have tended to describe only two types of memory, however, namely declarative and procedural memory (Anderson, 1987). Procedural memory relates specifically to storing information about motor skills. It provides knowledge about how to perform a skill. The person might be able to perform the skill, but not be able to describe verbally how he performed it.

Regarding memory for speech production, Paradis (1998) states that implicit linguistic knowledge is subserved by procedural memory and is acquired “automatically”. The person is thus not aware of the processes involved in acquiring implicit linguistic knowledge, which refers to the phonology, morphology, syntax and lexicon of a language. According to Paradis (1998), acquisition of L1 occurs in this manner. Whilst speaking, a person is generally not aware of the grammatical aspects of language, for example, sentence construction. He merely utters a grammatically correct sentence due to his implicit linguistic competence (Paradis, 1998). According to Paradis (1995a), damage to this language system results in aphasia proper.

Semantic knowledge refers to general knowledge and includes factual and conceptual knowledge which develops from experiences (Shea *et al.*, 1993). Schmidt (1975) refers to semantic memory as memory for abstract generalization of a movement and labels this “schemas”. Shea *et al.* (1993:63) refer to semantic memory as a “person’s general background memory about words, symbols, concepts and rules”. Episodic memory refers to knowledge about personally experienced events and information about the time they were experienced, for example, when recalling the first time you

drove a car, information from episodic memory would be retrieved. Semantic and episodic knowledge can be verbalized. It entails knowing “what to do”. These two memory components are referred to as declarative knowledge and can be verbalized. It is important to distinguish between knowing “what to do” and “how to do” it when relating the three memory systems of long term memory with processes underlying motor control. In a specific situation a person might know “what to do”, but might not be able to perform the action successfully. This means that problems exist in attaching the appropriate parameter values to the selected motor program (Magill, 2001).

Paradis (1995a, 1998) states that acquisition of L2 occurs using more conscious strategies and this is known as explicit linguistic knowledge. According to Paradis (1995a), explicit linguistic knowledge is subserved by declarative memory. Paradis (1995a:6) states that explicit linguistic knowledge is “learned consciously (possibly but not necessarily with effort), is available for conscious recall, and is applied to the production (and comprehension) of language in a controlled manner”.

### **2.3.3 Conclusion regarding factors influencing information processing**

From the above discussion it is evident that specific aspects related to the task and environment/context in which a task is executed, exert an influence on information processing. These aspects include the stage of learning, automaticity, movement time and task complexity. Furthermore, specific aspects inherent to the individual are involved in information processing, namely, attention and memory. In the next section, contexts proposed to influence speech and language processing in particular will be discussed.

## **2.4 CONTEXTS FOR SPEECH PRODUCTION**

Context-sensitivity is an integral part of speech production (Van der Merwe, 1997:6). In Van der Merwe’s (1997) framework of speech sensorimotor control, it is hypothesized that “contextual factors affect the dynamics of motor control by exerting an influence on the mode of coalition of neural structures involved during a particular phase and on the

skill required for the planning, programming and execution mechanisms”. Van der Merwe (1997) states that certain variations of a specific contextual factor might necessitate more complex control strategies than others. The context of speech production thus influences the processing demands, which in turn necessitate the allocation of more attentional resources and consequently more conscious control. Different motor tasks, for example, exhibit different levels of activity of certain neural structures, as has been observed by neurophysiologists (Schultz & Romo, 1992). It has also been found that unfamiliar and precise fine movements require more sensory input and thus greater involvement of the sensory areas, than well-learned or ballistic movements (Brooks, 1986). Van der Merwe (1997) concludes that context therefore influences the control system.

As discussed in chapter one, Van der Merwe (1997) acknowledges a variety of contexts in her framework of speech sensorimotor control. These include voluntary versus involuntary (automatic) speech, sound or phonological structure, motor complexity of the utterance, length of the utterance, familiar versus unfamiliar utterances and speech rate. Van der Merwe (1997) states that the role of the various contextual factors will have to be determined by research and goes on to say that both treatment and research results will be influenced by variation in contextual factors. The contexts for speech production identified by Van der Merwe (1986, 1997) will be discussed in the following section.

#### **2.4.1 Speaking rate**

One temporal characteristic of speech production is speech rate. This is a temporal variant which causes drastic changes in the speech production process (Gay, 1981; Kelso *et al.*, 1983) and also has perceptual consequences. Speech physiologists state that an increase in speech rate demands substantial modifications of system control compared to normal speech rates (Abbs, 1973; Gay & Hirose, 1973; Gay, Ushijima, Hirose & Cooper, 1974). An increase in speech rate can be accomplished by reducing pauses between phrases, by increasing the rate of words within a phrase or by reducing word or syllable durations (Ludlow *et al.*, 1987).

Movement duration of the articulators also changes with an adjustment in speaking rate. This implies that during the motor planning of speech, the core motor plan needs to be adapted according to speaking rate. Consequently, Van der Merwe (1997) views speaking rate as a context of speech production. When speech rate changes, the context thus changes which necessitates adaptation of the core motor plan to the context of increased speaking rate (Van der Merwe, 1997). The core motor plan is adjusted in terms of the temporal and spatial characteristics needed for the acoustic realization of the specific phoneme/(s) in context.

When studying the speech of persons with neuropathology, an alteration of speaking rate is often employed as a means of assessing motor facility. The ability to accomplish rate changes with ease is a characteristic of the normal speech production mechanism, whereas an impaired motor speech system will be limited regarding this ability (Kent & McNeil, 1987). Furthermore, studying the effect of changes in speaking rate can tell us more about the disorders in persons with neurogenic speech and/or language disorders, since an increase in speech rate places higher demands on the speech production system and consequently on motor control (Kent & McNeil, 1987; McNeil *et al.*, 1990a). However, linguistic aspects of speech production are also presumably influenced by an increase in speaking rate (Fossett, McNeil & Pratt, 2001).

It could be argued that speech production at a faster than normal rate requires greater motor skill and consequently more attentional resources, and controlled processing needs to be exerted. Resources might, however, be more easily exceeded when rate has to be increased in persons with difficulty regarding speech and/or language processing, causing deficits to become more evident under circumstances of increased processing demands. By studying the effect of speech rate alterations on specific parameters of speech production, more can thus be learnt about the motor control of speech in different subject groups under circumstances of increased processing demand, than by investigating these aspects at the usual self-selected rate alone.

#### **2.4.2 Level of voluntary initiation of actions**

Kelso and Tuller (1981:229) mention voluntary versus automatic speech as a contextual factor in speech production. These researchers distinguish between “planned” and

“voluntary” acts as opposed to “habitual” or “automatic” acts. Kelso and Tuller (1981) refer to the phenomenon that persons with AOS can perform certain “habitual” actions adequately in certain contexts, whilst other times, they are unable to do so in response to a clinician’s request. Buckingham (1979) states that spontaneous conversation is a less voluntary action than when a person is requested to speak. In this sense the level of voluntary initiation can be viewed as a context for speech production (Van der Merwe, 1986).

From an information processing perspective, one could argue that more attentional resources are needed for voluntary initiation of actions than when they are performed “automatically” or in response to a specific stimulus. Capacity in working memory is thus presumably required for the voluntary initiation of the action. Persons with AOS are known to have difficulty with the voluntary initiation of utterances (Wertz, LaPointe & Rosenbek, 1984), indicating that voluntary initiation is more difficult for persons with difficulty regarding the motor planning of speech.

### **2.4.3 Familiarity: Novel versus automatic production**

The more familiar a person is with a skill, the less feedback is utilized for its control and the more fluent the skill becomes (Ashton, 1976). Execution of an unfamiliar movement is slower, probably because it cannot be completely planned in advance (Allen and Tsukahara, 1974). The more familiar a person is with a specific movement pattern, for example, production of a specific word, the better the person can anticipate upcoming events in the movement sequence or utterance. Variability also decreases as the movement becomes more skilled with practice, implying that the planning and execution of the movement are more precise and easier to accomplish than when producing a novel sequence or word (Sharkey & Folkins, 1985).

Novel actions require conscious attention and consequently demand more attentional resources due to the need for controlled processing. Production of a foreign word for the first time, for example, requires conscious attention and consequently greater attentional resources. As actions, or in this case speech, are practised, the movements become “automatized” and attentional resources are not required for their execution (Levelt, 1989; Magill, 2001). Although certain aspects of speech production might still require

conscious processing, as proposed by Levelt (1989), the greater part of speech production occurs in a fairly automatic fashion under normal circumstances in normal speakers.

#### **2.4.4 Sound structure**

Sound structure can be viewed as a contextual factor for speech production (Van der Merwe, 1986). Speech production is a learnt skill and it appears as if certain speech sounds and combinations thereof require greater skill and are more difficult to produce than others (Calvert, 1980; Oller & MacNeilage, 1983). The fact that certain speech sounds are acquired earlier in childhood (Ingram, 1976) and that the first syllable to be produced consists of a consonant and a vowel or reduplication thereof, might indicate that this structure is easier to produce than other sounds and structures (Van der Merwe, 1986). Rosenbek, Kent and LaPointe (1984) view the consonant-vowel-consonant structure as the easiest to produce. On the other hand, Seddoh *et al.* (1996b) proposed that words with closed syllables might be more difficult to produce than those with open syllables for persons with AOS. Utterances with consonant clusters, for example, CVCCC combinations (for example, “desks”) are also known to be extremely difficult to produce (Calvert, 1980).

#### **2.4.5 Motor complexity**

Van der Merwe (1986) states that certain factors, which are not related to the sound structure of an utterance, can also increase motor complexity. These include coarticulation possibilities and greater variation of articulatory characteristics, for example, sequential stretching and rounding of the lips. It is important to determine which contexts influence speech production in persons with neurogenic speech disorders, since persons with AOS have been noted to experience difficulty with the production of certain sounds or combinations thereof (Wertz *et al.*, 1984). The motor complexity of an utterance can most probably not be ascribed to a single factor. A certain combination of sounds might pose higher demands to the linguistic-symbolic planning of an utterance, for example, require more conscious processing and might also be motorically more difficult to produce (Van der Merwe, 1986).

Kent and Rosenbek (1983) state that most dimensions of apraxic disturbances are influenced by increased syllabic or phonetic complexity and that this aspect of apraxic behavior is important for theoretical understanding of AOS. The observation that certain sounds are more difficult to produce than others in persons with AOS underscore the fact that sound structure exerts an influence on motor performance and can consequently be seen as a context for speech production. A motorically complex utterance will presumably increase the demands regarding both motor and linguistic processing. From an information processing perspective, increased movement complexity will presumably require more controlled processing for correct production.

#### **2.4.6 Length of the utterance**

The length of the utterance is partially determined by the sound structure, but in this case length also refers to single words versus more words or even sentences. Longer utterances, just as longer movement sequences (Magill, 2001), would presumably take longer to plan. This has been found experimentally by Klapp, Anderson and Berrian (1973). In children it has been demonstrated that during imitation tasks, utterance length contributes to the number of incorrect responses produced (Miller, 1973; Montgomery, Montgomery & Stephens, 1978; Smith & Van Kleeck, 1986).

Strand and McNeil (1996) investigated the effect of length and linguistic complexity on temporal acoustic parameters of speech in persons with AOS. They found that persons with AOS had consistently longer vowel and between-word segment durations than normal speakers in all conditions. The persons with AOS also produced longer vowel and between-word segment durations in sentence contexts than in word contexts. Strand and McNeil (1996) concluded that the differences in duration in sentence production versus word or word-string production implied different mechanisms for executing motor programs for varying linguistic stimuli.

As mentioned previously, an interval of time in which the motor control system is prepared according to the demands and constraints of the situation/context precedes the intended action. The number of parts of the movement increases the movement complexity (Magill, 2001). In the same sense the number of sounds to be articulated

might increase speech and language processing demands and consequently complexity of production.

#### **2.4.7 Linguistic complexity**

The influence of linguistic aspects on speech production has also been proposed. In a recent study by Maner *et al.* (2000), the influence of sentence length and complexity on speech motor performance was investigated. Specifically these researchers investigated the effect of increased linguistic demands on articulatory movement stability in both children and adults. This was done by analyzing lower lip movement stability, which rendered a spatiotemporal index reflecting the stability of lip movement over ten productions of a specific phrase. The index reflected contributions of spatial variations (for example, variation in amplitude of movement) and temporal variations (for example, change in timing of peak displacements). A higher index reflected greater variability in the normalized movement waveforms. The phrase these researchers used for analysis was spoken in isolation and then embedded in sentences of varying complexity.

Maner *et al.* (2000) found that the spatiotemporal index was increased significantly for the phrase when it was spoken in the complex sentences compared to being spoken in isolation (baseline condition). Furthermore, the spatiotemporal index values of the children were consistently higher than those of the adults across conditions. The adults thus had more stable production systems and were not as easily influenced by the increased processing demands. These researchers concluded that their findings rendered “novel evidence that speech motor planning, execution, or both are affected by processes often considered to be relatively remote from the motor output stage” (Maner *et al.*, 2000:560).

Studies of phonological and syntactic processing in typically developing and disordered children, as well as work on the effect of linguistic variables on speech fluency in stutterers and non-stutterers have also been conducted in an attempt to shed light on the interaction between language processes and speech motor performance (Maner *et al.*, 2000). Kamhi, Catts and Davis (1984) found that increases in language complexity had effects on the accuracy of target sound production in normally

developing children aged 22-34 months. Furthermore, these researchers concluded that phonological performance was influenced by changes in language complexity more often in younger children than in older children.

In another study by Masterson and Kamhi (1992) imitated and spontaneous speech tasks were examined in three groups of children, namely, language-learning disabled, reading disabled and normal language-learning. These researchers sought to determine if processing demands in one component would lead to a decrease in performance of another component. Specifically they wanted to establish if increased syntactic complexity would lead to a decrease in semantic/phonological complexity and/or accuracy and vice versa. From the results of their study, Masterson and Kamhi (1992) concluded that for later acquired speech and language skills, increasing processing demands at one level influenced performance at another level. Specifically these researchers found that elementary sentences were produced with higher phonemic accuracy than complex sentences. Gordon and Luper (1989) investigated differences in the number of dysfluencies of three-, five- and seven-year-old non-stuttering children as syntactic complexity was varied using three different syntactic constructions. All three age groups produced a significant complexity effect for the passive sentence construction form. These findings render evidence that linguistic variables affect speech production.

Research in the area of stuttering has also focused on the relationship between syntactic complexity and fluency. If fluency is influenced by syntactic complexity, it would imply that motor speech processes are negatively affected by a linguistic variable, namely increased syntactic processing demands. Gordon, Luper and Peterson (1986) examined the effects of increased syntactic complexity on fluency in five-year-olds with normal fluency during a sentence imitation and sentence-modeling task. A significant effect of syntactic complexity on fluency was only found during the sentence-modeling task. In a later study on normally developing children and children who stutter, effects of syntactic complexity on fluency were also found during imitation tasks (Bernstein Ratner & Sih, 1987).

Silverman and Bernstein Ratner (1997) also reported that normal dysfluencies and errors in repetition accuracy increased in stuttering and non-stuttering adolescents

when the syntactic complexity increased. However, stuttering frequency was not affected by changes in syntactic complexity. The researchers concluded that although syntactic effects were strong for young children, they were minimal for older speakers. It was still evident, however, that linguistic factors impacted on motor performance in the subjects in their study.

The above mentioned studies render support for the assumption that interactions occur between linguistic complexity and motor speech performance. When linguistic or syntactic complexity increases, the processing demands are presumably increased regarding both linguistic and motor processing. This presumably requires allocation of more attentional resources, more controlled processing and consequently can impact on the execution level of speech production if resources are exceeded.

#### **2.4.8 First versus second language in bilingual speakers as a context for speech production**

The study of bilingualism has attracted attention from several disciplines. These include, amongst others, psychologists who investigate the effect of bilingualism on mental processes, sociologists who treat bilingualism as an element in cultural conflict, educationists who are concerned with bilingualism as it relates to public policy, sociolinguists interested in the ways in which language is used in society, and linguists who are interested in bilingualism as an explanation for certain changes in a language (Romaine, 1995).

Speech production in bilinguals has also been of interest to researchers in the field of speech-language pathology, because of the potential it has to shed light on the underlying mechanisms involved in bilingual speech and language processing (Paradis, 1990, 1992, 1995b). Determination of the localization of two or more languages in the brain and the neural substrate subserving these languages is another aspect which is attempted by such investigations (Albert & Obler, 1978; Ojemann, 1983; Ojemann & Whitaker, 1978; Paradis, 1977, 1989, 1993; Whitaker, 1989). Aphasia in bilingual speakers is also often investigated in an attempt to learn more about the recovery and disruption of language in bilinguals or polyglots in the presence of neurologic lesion (Paradis, 1977).

Speech production in bilingual speakers can be viewed as a context for speech production from a number of perspectives. Speech production in L2 might pose higher demands to both linguistic and motor processing, causing speech production in this language to be more difficult or effortful. The reason for this is that L2 can be viewed as a fairly novel speaking context which is also less automatized compared to L1, since the speaker does not use this language as often as L1. Furthermore, in the case of where L2 was acquired after L1 had been established (coordinate bilingualism), L2 was probably acquired using more conscious metalinguistic strategies, rather than being acquired via more automatic processes (Paradis, 1995a). Language processing in L2 will thus presumably require greater resources and controlled processing, due to the “novelty” and less “automatized” nature of L2. Processes other than conceptualization and monitoring as proposed by Levelt (1989), will thus also require more controlled processing. The speaker will, for example, need to think consciously about word selection for sentence production resulting in increased demands regarding linguistic-symbolic planning of the utterance.

Speech production in L2 might also increase processing demands regarding the motor planning of an utterance. Regarding motor planning or processing of an utterance in L2, sounds in L2 which are not a part of the sound repertoire of L1 will also be novel, less automatized and consequently require more controlled processing for their production. Even if the sounds in L1 and L2 are similar, the increased attention required for formulating and producing utterances in a less automatized language, might also impose demands on the motor planning of the utterance. For these reasons, speaking in L2 is proposed as a context exerting an influence on speech and language processing. The influence of these increased processing demands might impact on the execution level of speech production, since resources between language and motor domains are presumably shared (Strand & McNeil, 1996).

Speech production in L1 versus L2 in bilingual speakers as a context for speech production is particularly important to study in persons with breakdown at various levels of the speech production process. The resource capacity of persons with neurologic lesions might be more easily exceeded when the processing demands are increased with speech production in L2, since more than normal resources already need to be allocated to the levels of the speech production process where difficulty exists. The consequence

of increased processing demands might be visible in the temporal parameters of the acoustic speech signal.

Temporal control is inherent to speech motor control and the temporal parameters of speech production are measurable in the acoustic speech signal. Temporal parameters of speech production are often studied in an attempt to infer about the motor processes underlying speech production (Keller, 1990). Studying the effect of these increased processing demands imposed by speaking in L2 thus has the potential to inform about the nature of speech and language disorders in persons with breakdown at various levels of the speech production process. Furthermore, information can be obtained as to how these persons perform in the presence of increased processing demands imposed by a language context (L1 versus L2). The latter in turn will influence the planning of assessment and treatment procedures for these persons. The study of speech production in persons with neurogenic speech and language disorders will be discussed in more depth in chapter three.

To delineate the study of specific temporal parameters of speech production in L1 versus L2 in bilingual speakers, it is necessary to examine some of the concepts and theoretical issues related to bilingualism. After this, speech as a motor skill and the parameters of speech production which are studied in an attempt to learn more about the higher level processes which occur in the brain during speech and language processing will be discussed.

## **2.5 BILINGUAL SPEECH AND LANGUAGE PROCESSING**

### **2.5.1 Defining bilingualism**

Several definitions of bilingualism have been proposed. Considering that the degree to which a person is proficient in L2 can differ regarding comprehension, or expression, or both, it becomes clear that bilingual speakers can function at various ends of a continuum of proficiency (Romaine, 1995). Mackey (in Romaine, 1995:12) consequently “considers bilingualism as simply the alternate use of two or more languages”.

### 2.5.2 Types of bilingualism

Weinreich (1968) distinguishes three types of bilingualism, namely, coordinate, compound and sub-coordinate bilingualism. In coordinate bilingualism the person learns his two languages in separate environments, resulting in the words of the two languages being kept separate with each word having its own specific meaning (Junqué, Vendrell & Vendrell, 1995; Paradis, 1995a; Romaine, 1995). An example of this would be a person whose home language is Afrikaans, but who learned English later in his life, as a second language, at school.

Compound bilingualism entails that the two languages were learnt in the same context and were thus used concurrently whilst being learnt. This results in a fused representation of language in the brain causing two words to be tied to the same mental representation. A single concept with two different verbal labels, one in each language, thus exists. The two languages are consequently interdependent. An example of this would be a person who grew up in a bilingual home (Junqué *et al.*, 1995; Paradis, 1995a; Romaine, 1995).

A third type of bilingualism is a subtype of coordinate bilingualism and is known as sub-coordinate bilingualism. Sub-coordinate bilingualism implies that the bilingual speakers interpret words of their weaker language through words of their stronger language. If English were the weaker language, the English word “book” would, for example, evoke the Afrikaans word “boek” in an Afrikaans/English bilingual speaker. This type of bilingual speaker has a primary set of meanings in L1 with another linguistic system attached to them (Romaine, 1995).

When studying any aspect of bilingualism, it is important that the subjects exhibit the same level/type of bilingualism, since this could influence speech and language processing and consequently the demands imposed by the two languages and the results obtained in a specific study. If one were to study specific parameters of speech production, the level of bilingualism could thus influence speech production in the two languages. In other words, the type of bilingualism could influence the ease with which L2 is produced.

### 2.5.3 Crosslinguistic influence

The term “crosslinguistic influence” is used to refer to the influence of one language of a bilingual speaker on the other language during speech production (Sharwood-Smith & Kellerman, 1986:1). Foreign accent is an example of crosslinguistic influence at the level of pronunciation. In this instance, a bilingual speaker associates a phoneme of L2 with one in his primary language and subjects it to the phonetic rules of his primary language. This can result in under-differentiation, over-differentiation, re-interpretation or substitution (Romaine, 1995).

Romaine (1995) explains these abovementioned four phenomena as follows. *Under-differentiation* occurs when one language makes a distinction between sounds, which is not made in the other language. English, for example, distinguishes between the vowels of “sit” (/ɪ/) and “seat” (/e/), whereas French has only one sound in this area of vowel space, namely, the /ɛ/ as in “petit”. A French/English bilingual speaker might then under-differentiate these two sounds in English and replace both with the French /ɛ/-sound. *Over-differentiation* results from the “imposition of phonological distinctions made in one language on sounds in the second one”, for example, the carryover of vowel system length in one language onto another language where it is not needed (Romaine, 1995:53). *Reinterpretation* occurs when the bilingual speaker is misled by the written form of the word and applies the pronunciation of L1 which is elicited by the specific written form, for example, words with the double consonant in English pronounced by an Italian. The English word “patty” will then be pronounced /pattɪ/. *Substitution* occurs when a bilingual speakers replaces a sound of L2 with a sound from L1, because the original sound does not occur in L1. An example of this is English speakers in South Africa who replace the Afrikaans glottal /ʔ/-sound with a /ɰ/-sound, because English does not have a glottal /ʔ/-sound in its phonetic repertoire (Romaine, 1995).

Phonotactic patterns between languages thus differ, with the result that pronunciation of a sound not included in a person’s L1 might initially be more difficult to produce on an articulatory level, because the motor pattern for that sound has not been specified in the brain. It is evident that each language has specific characteristics

(temporal and spatial parameters) assigned to each phoneme for production of that phoneme. Van der Merwe (1997) refers to the temporal and spatial parameters which are assigned to a phoneme as the core motor plan. If L1 does not have the specifications for production of the specific phoneme in L2, it is thus presumably replaced with a corresponding phoneme in L1 or pronounced with the specifications of the L1 phoneme.

From an information processing perspective, production of sounds not included in a person's L1 can thus be seen as novel and consequently less automatized. This might lead to increased processing demands imposed by speech production in L2. From a motor control perspective, retrieval of core motor plans of L2 speech sounds which do not occur in L1, as well as adaptation of these motor plans to the phonetic context, will contribute to increased motor complexity of L2 utterances. All the operations which occur during the motor planning and programming of speech production as proposed by Van der Merwe (1997) will presumably be more difficult in L2, since L2 speech sounds are novel and less automatized compared to L1 speech sounds.

Cross-linguistic influence can also affect other aspects of speech production, such as, prosody or even the pragmatic level of language (Romaine, 1995). Bilingual speakers can, for example, transfer the stress patterns of L1 to L2, sometimes causing misinterpretation or unintelligibility. Romaine (1995) sites the example of French/English bilinguals who tend to give equal stress to every syllable when speaking English, since it is characteristic of French speech timing.

#### **2.5.4 The study of specific aspects of bilingual speech production**

Temporal parameters of speech sounds in L1 compared to L2 have been investigated. This includes, for example, the study of VOT. Caramazza, Yeni-Komshian, Zurif and Carbone (1973) found that bilingual speakers whose two languages have different VOTs may produce VOT values in at least one of the languages which are intermediate in value to those of monolingual speakers. It has also been found that bilingual speakers, whose two languages have different VOTs, perceive VOT differently compared to monolinguals. Watson (as cited in Romaine, 1995) proposed that this compromise reduces the processing load involved in mastering two phonetic

repertoires. The bilingual speaker thus stays within the boundaries of acceptable production, but the values do not completely match the values of the monolingual speaker in either of the languages. When a language is acquired later in life, the speaker does not “necessarily establish distinct phonetic categories for the sounds in that language which differ from those of their first language” (Romaine, 1995).

VOT has been found to exhibit inherent language-universal features, as well as learnt, language-specific characteristics (Smith, 1978). For example, the short-lag category of stop consonants in Spanish appears to differ somewhat from the English short-lag category (Lisker & Abramson, 1964) and the Swedish long-lag stop consonants exhibit somewhat greater durational values than English long-lag stops (Fant, 1960). In languages such as Dutch and Afrikaans, aspiration of stops is not as common as in English (Lisker & Abramson, 1964) and different VOT will thus presumably exist for these languages.

Durational aspects of specific speech segments might also be language-specific. It is well-established that the duration of segments varies depending on the phonetic context (Kent *et al.*, 1996; Van der Merwe, 1997). For example, duration of a vowel which precedes a voiced consonant is one and a half times that of the same vowel preceding a voiceless consonant (Kluender, Diehl & Wright, 1988; Peterson & Lehiste, 1960). In this regard, Kent *et al.* (1996:217) state that although the aforementioned finding appears to be present in different languages, it appears “especially pronounced in English, suggesting a learned phenomenon in addition to physiologically based conditioning”.

In another study related to speech production in bilinguals, Lubker and Gay (1981) examined the amount of lip rounding for rounded vowels in Swedish and English speakers. They hypothesized that Swedish subjects would have greater anterior-posterior displacements than English speakers, since the Swedish vowel space is more crowded than the American English vowel space, especially regarding rounded vowels. Decreasing the length of the labial segment leads to an upward shift in vowel formants. This would in turn infringe upon the space of another vowel in the Swedish system, thereby creating the chance of perceptual confusion.

From the results of their study, Lupker and Gay (1981) concluded that motor control of the lip rounding gesture in speech production is a language dependant, learnt behavior which is more important to some languages than to others. Swedish speakers move their lips further, initiate onset of movement earlier with greater velocity and precision of goal achievements than speakers of American English. The study by Lubker and Gay (1981) points to the fact that spatial-temporal aspects of speech movements are language-specific and need to be learnt. This underscores the fact that production in L2 might be motorically more complex and consequently exert greater processing demands.

Klein *et al.* (1995) have also proposed that speech production in L2 might be motorically more difficult. Regarding speech production in L2, using positron emission tomography, these researchers found that the articulatory demands of L2 might require additional processing. The latter finding was deduced from activation of the left putamen during articulation in a repetition task in L2. Subcortical activation sites were not evident during speech production in monolinguals. Klein *et al.* (1995) postulated that activation of the left putamen was presumably the result of increased articulatory demands which were imposed by speech production in a language which was learned later in life.

From the above discussion it is evident that production of speech sounds which are not part of one's L1 repertoire would most probably pose higher motor demands to the speech production mechanism, since the motor plans for their production are novel and less familiar. If L2 is not used as often as L1, these motor plans will also not be as automatized as those for sounds in L1. More conscious processing is thus necessary for production of these sounds and consequently allocation of more attentional resources to a process (speech production) which is generally executed automatically. Even if the sounds of L1 and L2 are similar, the less familiar language (L2) might still pose higher linguistic processing demands and these might in turn also place higher demands on motor processing. The influence of these increased processing demands might be manifested in the parameters of speech production as measured in the acoustic speech signal. Speech production in L2 might, for these reasons, hypothetically speaking be more difficult. Difficulty with speaking in L2 might be especially evident in populations with neurogenic speech and/or language

disorders, since these subjects display less flexibility to adapt to increased processing demands (Kent & McNeil, 1987).

### **2.5.5 Language processing and automaticity in bilinguals**

Positron emission tomography studies are often employed to study the neuronal processes that underlie linguistic performance in normal unilingual subjects. Activity-related regional cerebral blood flow is measured in order to make inferences about the neural substrates that underlie specific functions of unilingual and also bilingual language processing activities. The rationale for using positron emission tomography activation studies involves the assumption that any task places specific processing demands on the brain (Klein *et al.*, 1995). These demands result in changes in neural activity in various functional areas of the brain causing changes in the local blood flow to these areas (Raichle, 1989). Language is the most studied process in brain-imaging research (Haxby, Grady, Ungerleider & Horwitz, 1991)

Positron emission tomography findings regarding primary sensory processing and motor output have generally been uncontroversial, although localization of higher cognitive functions and their interpretation have been subject to more debate (Demonet, Price, Wise & Frackowiak, 1993; Liotti, Gay & Fox, 1994). These higher cognitive functions include aspects, such as areas which are activated during phonological and semantic tasks. Positron emission tomography studies have rendered support for the hypothesis that some speech tasks are more automatic than others. Klein *et al.* (1995:31), from the results of their study, concluded that “two pathways are distinguished by the degree to which the task at hand is learned or automatic”. Word generation requires a non-automatic pathway, while repetition plays an important role in the automatic pathway regarding verbal response selection. Regarding bilingualism and automaticity, Ojemann and Whitaker (1978) suggested that a less extensive cortical area subserves a language which has become more automatized, whereas a language in which one is less fluent (L2) is subserved by a more extensive cortical area. However, according to Paradis (1995b), this study has been questioned regarding methodological considerations.

On a receptive level, different strategies for sentence processing have been suggested in languages which differ typologically, for example, English versus Chinese (Romaine, 1995). Other researchers have also found different strategies used in processing on a receptive level in different languages (MacWhinney, Bates and Kliegl, 1984). From the results of their study, Klein *et al.* (1995) concluded, however, that “the same neural processes subserve second-language performance as subserve first” regarding the particular task which they used in their study. These researchers emphasize, however, that speaking in L1 and L2 might differ regarding the cognitive demands which the two languages place on the speech production mechanism, even though the same brain regions are active in both cases.

### **2.5.6 Neuroanatomical organization in bilinguals**

Another area of research which would support different processing strategies in bilinguals relates to the study of the cerebral localization of languages (Whitaker, 1989). If the two languages of a bilingual speaker are subserved by different neural structures, one could assume that their processing strategies might also differ. The study of the cerebral localization of languages has been undertaken by various means, for example, mapping sites in the brain where electrical stimulation alters naming in bilingual individuals (Ojemann and Whitaker, 1978; Ojemann, 1983). Although it has been proposed that bilinguals have their languages (implicit linguistic competence) less asymmetrically represented in their cerebral hemispheres than unilinguals, research has not supported this proposal (Paradis, 1990, 1995b). Recent positron emission tomography evidence has rendered support for the claim that the two languages of a bilingual speaker are not geographically separated within the brain, but are subserved by the same neurological substrate (Klein *et al.*, 1995). Paradis (1995b) also underscores the fact that all clinical studies indicate that implicit linguistic knowledge is subserved by areas of the left hemisphere in bilinguals to the same degree as in unilinguals.

Increasing evidence has been gathered to indicate increased involvement of the right hemisphere in pragmatic and paralinguistic aspects of language (Paradis, 1995b). Thus, although most researchers currently generally agree that there are not separate loci for different languages in the brain, it has been suggested that the right

hemisphere plays a greater role in the acquisition of L2 (Paradis, 1995b). Although not verified, it has been found that bilingual speakers rely to a greater extent on pragmatic aspects to interpret messages in their weaker language (Albert & Obler, 1978). Paradis (1998) explains this by stating that late bilinguals can compensate for gaps in their implicit linguistic knowledge by relying on controlled declarative memory. The latter is based on metalinguistic knowledge, as well as right hemisphere based pragmatic competence. Metalinguistic knowledge is acquired consciously and is stored explicitly. In contrast to this, implicit linguistic competence is acquired incidentally, stored implicitly, without conscious control and used automatically. These two types of memory can be neurofunctionally, although not neuroanatomically different according to Paradis (1995b) and might thus be differentially affected after brain damage. This then might be the reason for findings indicating different localization of different languages in the brain.

### **2.5.7 Conclusion regarding bilingual speech and language processing**

The above discussion indicates that different processing strategies might be employed for different languages during comprehension activities, even though the neural substrate underlying more than one language in the brain appears to be similar across languages. Production in L1 and L2 also appears to be subserved by the same neural substrates, although activation of the left putamen during repetition tasks in L2 indicates that speech production in L2 might impose additional articulatory demands. It thus appears as if speech production in L2 might be motorically more difficult as was proposed by Klein *et al.* (1995). Speech production in L2 might consequently heighten the processing demands and result in greater difficulty regarding the accomplishment of perceptually accurate speech.

One way in which to obtain information regarding the processes which occur in the brain during speech production is to study the manifestation of these processes in the spatial and temporal parameters of the acoustic speech signal. Since speech is a motor skill, it is necessary to discuss some concepts related to the study of speech as a motor skill. However, it is also imperative to bear in mind that speech is the result of both linguistic and motor processing which occur in the brain. The impact of increased processing demands might be visible in the temporal and spatial parameters of speech

production as measured in the acoustic speech signal and consequently these parameters of speech production are often studied in an attempt to learn more about the higher level speech and language processing which occurs in the brain.

## **2.6 SPEECH PRODUCTION AS THE MANIFESTATION OF LANGUAGE**

Hodge (1993:128) defines speech as “the acoustic representation of language, that results from highly coordinated movement sequences produced by the actions of the speech mechanism”. Mlcoch & Noll (1980:201) define speech production as “a process in which internal thought is progressively externalized into a series of muscular contractions resulting in a particular acoustic output”. No matter how the speech production process is defined, it is one of the most complex human behaviors to analyze (Borden & Harris, 1984). In this regard, Gracco (1990:3) states, “Speaking is a complex action involving a number of levels of organization and representative processes” Because these higher level speech and language processing functions are largely inaccessible, researchers must infer the nature of these processes by examining the behavioral manifestations of neural processing (Borden & Harris, 1984). Since temporal and spatial parameters of speech production can be measured on an articulatory level and consequently can be used to make inferences about the language processes which precede the acoustic realization of language, it is necessary to take a closer look at speech as a motor skill.

### **2.6.1 Speech as a motor skill**

Speech is a motor skill with a perceptual goal namely, generating sound patterns to convey a message. Air from the lungs is used to produce different speech sounds. These speech sounds are then further modified depending on the phonetic context in which they are produced (Borden & Harris, 1984). The breath stream is regulated as it passes from the lungs to the atmosphere. To accomplish the perceptual goal, the movements of the respiratory system, larynx and articulators (supralaryngeal) need to be coordinated to reach the desired acoustic output (Smith, 1992a). The movements of these structures are the result of muscle contractions due to nerve impulses, all of which are controlled in the nervous system (Borden & Harris, 1984).

Hirose (1986:61) states "the speech production process can be viewed as a fine motor skill which must be regulated in terms of sequence and duration with great accuracy, speed and rhythmicity". For the accomplishment of coordinated movement, Lashley (as cited in Moll, Zimmermann & Smith, 1977) postulates that there are two major aspects that need to be integrated, namely the temporal and spatial parameters of speech production. Movement is thus the result of signals which are ordered in time and space. An understanding of the spatial aspects of a system is necessary to determine the operation of that system. The "spatial relationships of structures...are important in the development of hypotheses about how movements occur" (Moll *et al.*, 1977:111). However, Moll *et al.* (1977) emphasize the importance of knowledge regarding the temporal characteristics of speech for understanding coordinated motor output.

The spatial and temporal parameters of movement for each articulator and consequently each muscle necessary for movement of the specific structure need to be specified and coordinated with every other muscle/articulator involved in the specific movement. It is thus evident that timing and coordination of speech movements constitute an integral part of speech motor control (Keller, 1990). Abbs (1988) also emphasizes the fact that coordination, including both intra- and interarticulatory coordination or timing, is the essence of speech motor control (Abbs, 1988).

The importance of control of the temporal and spatial parameters of speech production becomes evident when one considers the large number of degrees of freedom of the speech production system and (Kent *et al.*, 1996) the phenomenon of motor equivalence. The presence of these two phenomena thus needs to be considered when studying speech as a motor skill.

### **2.6.1.1 Degrees of freedom**

Considering the number of muscles and movements that need to be coordinated during speech production, the question arises as to how the nervous system controls the many muscles and joints involved in producing a given pattern to generate a specific acoustic signal (Smith, 1992a). This is referred to as the degrees of freedom problem. The degrees of freedom “of any system reflect the number of independent elements or components of the system” and “arises when a complex system needs to be organized to produce a specific result” (Magill, 2001:44).

The degrees of freedom problem is also inherent to speech production. Multiple movements of various articulators need to be temporally and spatially synchronized to accomplish production of a specific sound/(s) with a characteristic acoustic output. According to Gracco (1990), there are more than 70 different muscular degrees of freedom in the production of speech. Kent *et al.* (1996:8) state “The tongue, lips, jaw, velum, larynx and respiratory systems all possess several possible types of movement with respect to range, direction, speed and temporal combinations with one another”. This is further complicated by the fact that they can combine their movements in various ways. Kent *et al.* (1996) state that it is important that a theory or model of speech motor control account for the way in which the nervous system solves this control problem. The fact that the movements of so many muscles and structures need to be coordinated during speech production underscores the importance of control of the temporal and spatial parameters involved in these movements.

### **2.6.1.2 Motor equivalence**

Despite the fact that temporal and spatial parameters of movement for each muscle and articulator need to be specified and coordinated, the desired acoustic output can be achieved by varying movements of the muscles and articulators involved. This is known as motor equivalence. A characteristic of general motor skills is the fact that a variety of component movements can produce the same action and therefore obtain the same goal (Abbs, 1981; Folkins, 1981; Hughes & Abbs, 1976; Kelso & Tuller, 1983; Netsell, 1984; Sharkey & Folkins, 1985). Similarly, the same phoneme can be

produced with widely varying articulatory movements (Smith, 1992a) depending on the inherent characteristics of the speaker, the speaking rate, amount of stress employed and the nature of the surrounding speech sounds. Thus depending on the phonetic and/or other contexts (for example, speaking rate) in which a phoneme is produced the movements of the articulators will vary (Van der Merwe, 1997), although the specific phoneme will maintain certain acoustic characteristics to be perceived as the intended phoneme.

Motor equivalence is characteristic of all motor skills and point towards the fact that coordination of motor skills is “flexibly accomplished by the nervous system, perhaps to ease what might be impossible, namely, achieving a given functionally significant goal in exactly the same way each time” (Abbs, 1988:160). From this statement it is evident that even in normal speakers, variability regarding the movement parameters is present from one production to the next, even when the context is held constant. Critical boundaries for motor equivalence exist, however, and deviation outside these boundaries will result in sound distortion or even perceived substitution (Van der Merwe, 1986, 1997).

In order to remain within the boundaries of motor equivalence considerable timing adjustments need to occur in the presence of multiple degrees of muscle and movement freedom (Gracco, 1988:4628). In order to shed light on temporal control, the temporal parameters of speech production have consequently been a prominent subject of investigation. One reason for the study of temporal parameters of speech production is because of their perceptual prominence (Kent & Rosenbek, 1983). These temporal parameters are often investigated by means of acoustic analysis, since this allows for objective measurement of temporal aspects of speech production. The study of the temporal parameters of speech production will be discussed in more detail in the following section.

### **2.6.2 Temporal parameters of speech production**

The study of temporal aspects of speech can aid in gaining insight into the motor control strategies used by the central nervous system to accomplish accurate speech production (Keller, 1990). Furthermore, spatiotemporal control is essential for achievement of

coordinated movement (Kent & Adams, 1989). Abbs and Connor (1989) pose that a primary role of the motorsensory system for speech and most other motor tasks, is the generation of precisely timed and measured multiple muscle contractions. Information regarding temporal control can thus be helpful when compiling models of normal and pathological speech motor control or in testing hypotheses proposed by such models.

Because time is an integral part of speech motor control, the effect of various factors on timing has been studied in an attempt to gain insight into the motor control of speech by the brain. As has been mentioned, speech production is context-sensitive with the implication that various contexts might influence timing in various ways. Speech production exhibits many temporal parameters, which can be perceived or measured acoustically. These temporal parameters are presumably influenced by the context in which they are produced.

Two temporal parameters of speech production which are often studied due to the fact that they have potential to inform about the nature of motor speech processes includes durational aspects of a segment/segments and the timing amongst movements of different articulators, known as IAS.

### **2.6.2.1 Segmental duration**

Segmental duration is a feature of language that represents the length of a given segment of speech, for example, a phoneme. Segmental duration is typically measured acoustically and is believed to reflect principles of speech timing (Forrest & Weismer, 1997). Importantly, different speech sounds have different intrinsic acoustic durations. Diphthongs and “long” vowels, for instance, are longer than the “short” and unstressed vowels. Similarly, continuous consonants, such as, fricatives are longer than stop consonants (Borden & Harris, 1984; Kent *et al.*, 1996).

Segmental duration is furthermore, dependent on the context in which a phoneme is found (Borden & Harris, 1984; Kent *et al.*, 1996). Vowels, for example, are longer before voiced consonants, than before voiceless consonants, as can be illustrated by comparing the words “leaf” and “leave” (Borden & Harris, 1984; Kluender *et al.*, 1988). Other factors that influence segment durations include speaking rate, the

phonetic context, position of the word in an utterance (for example, at the beginning or at the end of the utterance), the type of speech material (for example, isolated words versus connected speech and casual versus formal speech styles) and idiosyncratic speaker characteristics (for example dialect, age, gender and vocal tract length) (Forrest & Weismer, 1997).

It is thus evident that each speech sound has an ideal range of durations which is necessary for accurate perceptual realization thereof. This durational range is dependant on the inherent *characteristics of the phoneme* itself and the phonetic *context* in which it is to be produced. The duration of phonemes needs to be specified during the motor planning of speech (Van der Merwe, 1997; Walsh, 1984) and needs to be within the boundaries or limits of equivalence in order to be perceptually accurate (Van der Merwe, 1997). The ideal duration can thus vary to a certain extent, provided that it stays within the boundaries of equivalence, since too great a change in duration can sometimes result in a change of the meaning of a word in a language, such as, Afrikaans for example. To accomplish this ideal duration of a phoneme or syllable on the acoustic level within the boundaries of equivalence, timing is thus of considerable importance.

### **2.6.2.2 Interarticulatory synchronization**

From the fact that each speech segment has an ideal durational range, it becomes evident that considerable timing control needs to be exerted for the critical acoustic configuration to be reached and consequently for achievement of on-target speech. During the motor planning of speech, the spatial and temporal parameters of movements need to be specified (Van der Merwe, 1997). According to Van der Merwe (1997), these temporal and spatial specifications of movements constitute the motor goals. She emphasizes the fact that motor planning of speech is articulator-specific and that IAS needs to be planned for the production of a particular phoneme. The spatiotemporal features of the core motor plan are thus invariant. Invariant spatiotemporal features of the motor plan are similar to the generalized motor program proposed by Schmidt (1975, 1988).

During the motor planning of speech, the core motor plan is adapted according to the phonetic context and consequently motor goals regarding articulatory synchronization also need to be adjusted depending on the phonetic context. Adaptation of the spatiotemporal parameters of movement is presumably similar to Schmidt's (1975, 1988) proposal of parameters (absolute timing and forces of actions) which need to be scaled according to the context in which a movement is produced. Movements of the various articulators for speech production are thus precisely timed (Kent & Adams, 1989). Apart from the fact that the movement parameters for each articulator need to be temporally controlled (intra-articulatory synchronization), the timing between the movements of the various articulators needs to be accurately controlled. This is referred to as *interarticulatory timing or synchronization*.

Interarticulatory synchronization of various articulators has been studied, for example, synchronization of upper and lower lip movements (Gracco & Abbs, 1986; Tseng *et al.*, 1990) and lingual-laryngeal phasing (Ziegler & von Cramon, 1986). One form of IAS which is often measured in normal and pathologic speakers is VOT. Accurate voicing requires precise timing of a supralaryngeal event (oral articulation) and a laryngeal event (vocal fold vibration for voicing onset), and consequently VOT is viewed as a form of IAS (Tyler and Watterson, 1991; Van der Merwe, 1986).

- **Interarticulatory synchronization of supralaryngeal and laryngeal speech movements: Voice onset time**

Tyler and Watterson (1991:56) state "VOT is a temporal characteristic of stop consonants that reflects the complex timing of glottal articulation relative to supraglottal articulation...VOT is a reliable, relatively easy measurement to make and is thought to reflect a complex aspect of supralaryngeal-laryngeal coordination". VOT is defined by Borden and Harris (1984:289) as "the interval of time between the release of a stop-plosive, voiced or unvoiced, and the onset of voicing of the following vowel". The temporal relationship between these two events is usually determined in milliseconds (ms) and specific temporal boundaries of VOT exist for voiced and voiceless sounds respectively, to be perceived as either voiced or voiceless. For voiced sounds, VOT can range from -180 ms, implying that voice onset occurred before the release of air (voicing lead) to 25 ms after this release

(voicing lag). In voiceless sounds, voicing needs to be initiated between 40 ms and 120 ms after the release of air (+40ms to +120 ms) (Cooper, 1977; Zlatin, 1974).

It is thus evident that accurate timing plays a crucial role in speech production, since deviant VOT, for example, can also change the linguistic meaning of a word (for example, "pig" and "big"). Timing of articulatory movements relative to each other thus needs to be carefully specified in the motor plan of a specific utterance in order to obtain the desired acoustic output (Van der Merwe, 1997). Itoh and Sasanuma (1984) and Löfqvist and Yoshioka (1981) emphasize this by saying that VOT is a temporal aspect of speech production that is less variable than other temporal parameters and needs to be carefully controlled. Furthermore, the fact IAS also needs to be adapted to the context of production underscores the importance of accurate timing of this parameter (Van der Merwe, 1997).

## **2.7 CONCLUSION**

From the above discussion, it is evident that speech is a motor skill enabling measurement of specific speech production parameters. These spatial and temporal parameters need to be specified during the motor planning and programming of speech production as proposed by Van der Merwe (1997) in the four-level framework of speech sensorimotor control. Motor planning, programming and execution need to occur within the boundaries of motor equivalence to prevent distortion of the acoustic speech signal. It is important to bear in mind, however, that the speech production process cannot be considered without reference to the linguistic or language processes which precede the acoustic realization of the intended speech target.

Certain contextual factors might negatively influence both motor and linguistic processing and can impact on the temporal and spatial parameters of speech production. Speech production in certain contexts might thus prove to be more complex or difficult than in others due to increased processing demands. Difficulty in certain contexts might be more readily experienced by persons with neurogenic speech and/or language disorders since they already exhibit difficulty regarding one or more of the levels in the speech production process. The resources in these persons might be more easily

exceeded due to the fact that more than normal resources already need to be allocated to the deficient processes.

The study of the influence of speech production in L1 versus L2 as a context for speech has not been undertaken, although the influence of several other contexts has been studied. An investigation of the influence of speech production in L1 versus L2 is important for inferring about the processing demands imposed by speech production in L2 in bilingual speakers. The effect of the increased processing demands is important when studying the speech of persons with neurogenic speech and/or language disorders at distinct levels of the speech production process to gain information about the nature of these disorders and their reaction to increased processing demands. Models and theories of speech production can aid in explanation of normal and pathological speech and language processing in bilingual speakers with neurogenic speech and/or language disorders.

## **2.8 SUMMARY OF CHAPTER TWO**

In this chapter prominent models and theories of speech production were reviewed in an attempt to delineate the processes and stages involved in the speech production process. These models or theories included the dynamic systems model (Kelso & Tuller, 1981), motor schema theory (Schmidt, 1975, 1982, 1988), the information-processing perspective on speech production proposed by Levelt (1989) and the four-level framework of speech sensorimotor control proposed by Van der Merwe (1997). From this discussion, it was evident that contextual factors exert an influence on speech and language processing and consequently on motor control. Contextual factors inherent to the individual and those related to the task, influencing information processing were reviewed. After this, specific contexts which might influence speech and language processing were reviewed. Speech production in L1 versus L2 in bilingual speakers was proposed as a context for speech production, since speech production in L2 might increase processing demands.

Concepts relevant to bilingualism were discussed to gain insight into bilingual speech and language processing. The importance of viewing speech as the manifestation of

language processing by the brain was emphasized and manifestation of these processes in the temporal and spatial parameters of speech production was discussed, since speech is in essence also a motor skill. Furthermore, the fact that timing constitutes an integral part of speech motor control led to a review of specific temporal parameters which can be measured acoustically.

From this discussion, the relevance of studying the influence of speech production in L1 versus L2 as a context for speech production on specific temporal parameters of speech in persons with neurogenic speech and language disorders becomes evident. The study of the aforementioned could render important information regarding normal and pathological speech motor control in the presence of increased processing demands.