The handle [http://hdl.handle.net/1887/19818](http://hdl.handle.net/1887/19818) holds various files of this Leiden University dissertation.

**Author:** Umanski, Daniil  
**Title:** The use of technological support in communication disorders: how development of computer-based tools can refine the treatment of motor speech disorders  
**Date:** 2012-09-18
CHAPTER 2

Background

Speaking comes by nature; silence, by understanding.

A German saying
2.1 Speech motor control and its disorders

Speech is perhaps the most complex and extraordinary masterpiece of intricate movement skills which humans have acquired. The term ‘Speech Motor Control’ refers to the systems and strategies involved in the coordination and production of speech. Speech movements are not innate, and require practice for an extended period of time, before they mature and become skillful (Smith and Goman, 2004). Speech motor skill is evident in highly organized, both temporally and spatially, movement sequences, performed in an adaptive, cost-effective, and purposeful way (Van Lieshout et al., 2004).

Speech production has been modeled both theoretically (Dell, 1988; Levelt, 1999; Guenther, 2003; Kalveram, 2001) and computationally (Tourville and Guenther, 2011; Roelofs, 2000), with models varying in focus and detail of the underlying linguistic and sensorimotor processes. For example, Dell’s account is based on an interactive model of phonological encoding, resulting in an ordered set of discrete phonemes, so that each of the selected phonemes is translated into an articulatory code for the control of the speech muscles (Dell, 1986).

The model proposed by Levelt (1999) shares the idea of a linearly ordered string of speech motor plans. According to the model, once the target word has been retrieved from the mental lexicon, the process of phonological encoding results in abstract, syllabified phonological words, which are then incrementally translated into articulatory motor programs in the process of phonetic encoding (Levelt and Wheeldon, 1994). A crucial assumption of Levelt’s theory is that speakers have access to a ‘mental syllabary’ - a repository of syllabic gestures, which contains the articulatory scores for at least the high-frequency syllables of the language (Cholin et al., 2006).

Evidence for a mental store of syllabic programs converged from three lines of investigation. First, researchers have observed that syllables exhibit an exponential distribution throughout a language. In fact, 500 syllables from English, Dutch, and German, i.e. less than 5% of the entire syllable inventory in those languages, succeed to produce approximately 80% of all speech in those languages (Schiller et al., 1996). Second, a number of psycholinguistic studies have shown that participants produced high-frequent syllables with shorter response latencies than low-frequent syllables (Schiller, 1997). Third, clinical studies have consistently reported more accurate productions of high-frequency syllables in patients with apraxia of speech (Staiger and Ziegler, 2008).

Other models emphasize the interaction of motor and sensory components of speech production (Guenther, 1995; Shaiman and Gracco, 2002; Schmidt and Wrisberg, 2008). It is generally accepted that movement control is grounded on sensorimotor interaction, and that the brain uses feedforward and feedback (auditory, tactile, and proprioceptive) information in a flexible and generative manner, adapted to the context of ongoing performance (Van der Merwe, 1997).
Background

During motor learning, the control mode is assumed to be largely based on feedback control, which allow movement accuracy to be perpetually refined. After skills acquisition, control is assumed to rely mostly on feedforward projections, and feedback control may only be necessary if the brain’s predictive models need to be adjusted to novel circumstances (Guenther, 2003; Max, 2004).

There is still much debate in the literature regarding the nature and architecture of speech motor plans, as well as the tangled interactions between linguistic and sensorimotor processes (Caruso and Strand, 1999; Ziegler, 2009; McNeil, 2008). However, it is evident that an intact speech motor system generates a series of articulatory movements, producing an acoustic signal, such that an intended linguistic message of the speaker can be adequately comprehended by a listener. With all its incredible complexity, for most adult humans, speech is a robust and rather effortless activity, performed with a high degree of automaticity and adaptability.

Motor speech disorders

As a result of certain developmental or acquired conditions, the delicate coordination involved in speech motor control may become impaired. The group of such impairments is typically termed ‘Motor Speech Disorders’ (MSD). Although there is no consensus in the literature as to precisely which disorders are included in this group, the term generally implies dysarthria, apraxia of speech, and fluency disorders, namely stuttering and cluttering (Kent, 2000). Fluency disorders, being recognized as complex, multi-causal conditions, are not considered exclusively as motor speech disorders, but have been extensively studied from the speech motor skill perspective (Peters et al., 2000).

Dysarthria is a motor speech disorder resulting from central or peripheral nervous system damage, characterized by an abnormal neuromuscular activity, such as paralysis, spasticity or atrophy. People with dysarthria have difficulties with respiration, phonation, resonance and articulation (Du y, 2005). Consequently, the control for speech movements is reduced and can result in incomprehensible speech on the levels of intelligibility, audibility and naturalness due to nasalized speech, very poor articulation and rate control problems.

Apraxia of speech (AOS) is considered to constitute an impairment of the speech motor programming or the phonetic encoding stage of spoken language production (Du y, 2005), although the pathomechanism underlying AOS is still not well understood. AOS is thought to result from lesions to the dominant hemisphere, and more specifically to anterior areas, including Broca’s area (Ziegler, 2008). Speakers with apraxia of speech produce phonemic and phonetic errors, their speech is laborious, halting, with false starts and with trial-and-error groping movements, and is often characterized by a reduction of prosodic contrast (McNeil et al., 2004).
2.2. Speech motor skill learning

Childhood Apraxia of Speech (CAS) is a neurological motor speech disorder of unknown origin, which is thought to interfere with motor planning and programming of speech movements, resulting in moderate to severe deficits in speech intelligibility (Caruso and Strand, 1999). CAS has been associated with language deficits, dysarthria and cognitive deficits. Characteristics of CAS include difficulty with producing and timing speech movement sequences, with movement transitions between phonemes and syllables, vowel errors, lack of proper stress pattern (monotonous or expressionless), as well as reduced diadochokinetic rates (McNeil, 2008).

Stuttering is a complex communication disorder in which the flow of speech is disrupted by involuntary repetitions, prolongations of sounds, and silent blocks. It is generally accepted that stuttering emerges from a complex interaction among factors including genetics, language processing, emotional/social aspects, and speech motor control (Wingate and Howell, 2002). These factors may play different roles among individuals, as well as change in their expression over different periods of development (Smith and Zelaznik, 2004).

From a motor control perspective, stuttering has been described as a disorder in the timing and coordination of the systems involved in speech production (Peters et al., 2000). Caruso and Strand (1999) argue that viewing stuttering as a motor speech disorder may offer certain benefits for clinicians, such as facilitating the development of treatment protocols based upon principles of motor learning. Van Lieshout et al. (2004) propose that stuttering reflects a limited, or compromised level of skill for preparing and performing the motor actions required for the production of fluent speech. The elegance of this view is in rendering the rather static notion of a disorder into a more dynamic concept which involves a range of skill. The concept of motor skills entails the possibility of refinement, through a process of skill learning.

2.2 Speech motor skill learning

The process of learning and maintaining productive speech motor skills is referred to as speech motor training, which aims at teaching patients to produce correct patterns of speech, through gradually shifting production from conscious control of learned speech movements to an over-learned, automatic level. According to principles of motor learning (Schmidt and Lee, 2005), in order for new learned movements to become over-learned and automatized, they must be practiced systematically, and adequate feedback must be provided. Motor learning is known to be influenced by several factors: (1) the design of a treatment program; (2) the provided feedback; (3) the motivation of the learner.

Learning, in its broad sense, should be viewed not only through the mastery of behaviors which are being treated, but also through the degree of generalization and maintenance of the acquired skills (Schmidt and Lee, 2005). Generalization refers to the transfer from learned behavior to related, but untrained targets,
Background

as well as to different, unrelated targets. Maintenance refers to the preservation of acquired skills in the long-term. Since therapy contact time is always limited, the goal of treatment procedures is to encourage generalization from a relatively small set of trained targets to a larger communicative context, and to promote the maintenance of skills over time (Odell, 2002). Therefore, an insightful planning of treatment programs is of great importance.

Selection of speech treatment targets

One of the first considerations clinicians make when planning an intervention for a certain patient, is the selection of treatment targets. Treatment targets are specified on several levels. First, the aspect of speech motor control is identified, which impairs the patient’s speech production, and will be targeted in treatment. Next, an inventory of speech items needs to be selected, in which the impaired speech motor aspects are challenged. The selection of speech items can be guided by a number of linguistic parameters, as well as by considering the idiosyncratic patterns of a patient. Some treatment approaches target speech elements which are stable and consistent within a patient. Others choose to work on the elements for which production is most impaired. Among the general linguistic parameters, clinicians may consider the following:

The choice of the speech target unit. Various linguistic units (phonemes, syllables, words, phrases) are utilized for training, according to a certain treatment methodology. Although the syllable is considered as the core unit of speech motor programming in Levelt’s model (Levelt and Wheel- don, 1994), psycholinguistic evidence suggests that speakers plan their speech on multiple levels (Schiller et al., 2002).

The articulatory complexity of speech items. On a basic level, a certain hierarchy is presumed, by which speech sounds vary in difficulty, with vowels considered most easy, followed by single consonants, and consonant clusters being the most complex. However, in recent work, Ziegler (2009) presented a non-linear probabilistic model of the phonetic code which involves units from a sub-segmental level up to the level of metrical feet. The model is verified on the basis of accuracy data from a large sample of apraxic speakers, and thus provides a quantitative index of a speech segment’s motor complexity.

The frequency of speech items. Although the frequency of speech tokens in a language tends to correlate inversely with their articulatory complexity, frequency and neighborhood density variables (number of words that are phonologically similar to a target word), have been reported to have a consistent influence on speech production (Vitevitch, 2002). For example, a study by Anderson (2007) revealed that young children are
2.2. Speech motor skill learning

more likely to stutter on low frequency words than high frequency words. The authors suggest that neighborhood and frequency variables not only influence the fluency with which words are produced in speech, but also have an impact on the type of dysfluencies. With respect to AOS, studies reported a significant effect of syllable frequency on production accuracy of speakers with AOS (Staiger and Ziegler, 2008; Laganaro, 2008). Odell (2002) suggests that controlling for the frequency of speech targets is expected to a ect the learning process of new motor gestures.

The embedding of speech targets

Finally, it needs to be specified whether and how the selected speech items will be embedded in larger production units. For example, in the treatment of AOS, clinicians typically target individual sound segments and embed them in larger units such as syllables, words or phrases. Most often, target phonemes are framed within words, as their meaningfulness and functional relevance are thought to stimulate speech production (Odell, 2002). Clinicians may embed speech targets within different units throughout treatment phases. For example, one method for training stress-patterning (Tjaden, 2000), involves two levels of intervention. First, the patient is required to produce sequences of syllables with a certain stress pattern (i.e., DaDa or daDa). Later, in order to shift the trained skills towards real speech, a patient is asked to produce real words (i.e., Honesty or reHEARsal), with the corresponding stress pattern.

Another important choice is whether speech targets should be embedded within real words or non-words. Some evidence has been reported for the advantage of using real words. For example, Kahn et al. (1998) found that acquisition by their apraxic patient was significantly higher when the target was embedded in real words rather than in non-words. However, the use of non-words can better challenge the motor component of speech production, since the effects of higher-order linguistic processing levels are minimized (Namavivayam and van Lieshout, 2008). Furthermore, it has been suggested that for some young children, the task of producing non-words might be too abstract, and demotivate them to perform speech motor exercises (Yaruss and Logan, 2002).

To conclude, the process of selecting treatment targets for MSD patients should, ideally, by guided by factors based on principles of motor learning, advances in psycholinguistic theory, and on evidence from clinical studies. Since the described factors have been shown to influence motor skills learning, clinicians may benefit from having the means to systematically manipulate these parameters, in order to realize the optimal conditions of practice and feedback. In addressing this question, Chapters 3 and 4 of this thesis will describe instances of innovative solutions for assisting clinicians with compiling speech treatment programs.

1 Capital letters denote the stressed syllable
2.3 Computer aided speech language therapy

Having dealt with the design of a treatment program, a clinician must now address the remaining two aspects of motor learning, pertaining to the situation of practice itself — providing a meaningful form of feedback, and sustaining the motivation of patients to practice. Addressing these challenges have lead clinicians and developers to examine the possibilities of using computer-based systems to support the process of practicing speech and language skills.

Consequently, computer aided speech language therapy (CASLT) games have been developed in order to engage children in therapy by seamlessly integrating practice into an enjoyable process of gameplay. CASLT games may provide an opportunity to create fun and motivational forms of exercise. Ideally, such games are designed to stimulate children to perform many repetitions, explore the limits of their abilities, and solve motor problems (Sandlund et al., 2009b). By combining visual feedback on specific speech parameters with motivating gameplay scenarios, computer games may provide a training situation which appeals to current theories of motor control and learning (Wulf, 2007).

Visual feedback and exercise structure

An advantage of interactive games is the possibility to design tasks which stimulate specific motor control aspects, such as amplitude, timing, precision, smoothness, or the coordination of several motor aspects, as well as providing knowledge of performance (KP) feedback on these aspects. KP provides information about the nature of the movement pattern itself, while knowledge of result (KR) is given after the movement or exercise is completed, and provides information about the movement outcome in relation to the goal, often in terms of the spatial or temporal deviation (Schmidt and Lee, 2005).

This novelty is often realized through the application of biofeedback principles. Biofeedback is a powerful methodology, which involves the simultaneous measurement and display of a physiological process in real-time, enabling a person to increase awareness and control of this process (Maryn et al., 2006). Applied to speech production, computer programs typically aim to visualize a certain feature of the acoustic speech signal in real-time, in order to allow the user to gain better control over that feature.

The inherent structure of computer games offers some cardinal advantages for integrating treatment procedures. First, tasks in computer games are presented in a consistent manner, while at the same time, adaptation and dynamic changes in those tasks take place according to the real-time performance of players (Hourcade, 2008). The combination of consistency and adaptability forms an excellent framework for learning procedures, as tasks remain relevant and engaging throughout the treatment (Donker and Reitsma, 2004). Second, computer games typically involve a hierarchical structure of tasks. This feature also
promotes sustained motivation, as players remain challenged while progressing through levels of difficulty (Cordova and Lepper, 1996).

When learning new motor skills, the motivation to practice is a necessary component for effective learning (Schmidt and Wrisberg, 2008). Particularly intrinsic motivation is recognized as a powerful catalyst in promoting the learning of new skills (Ryan and Deci, 2000). Studies stress the importance of the learner’s attitude towards the practice situation in achieving progress in a variety of motor learning contexts (Green and Bavelier, 2008). Practicing in environments that have a meaning for learners improves motor learning, motivation, and generalization of new skills to novel environments (Graybiel and Saka, 2004).

Typically, computer games for speech therapy are based on a microphone input. Signal processing techniques are then applied to the acoustic signal in order to derive meaningful speech or voice parameters. For example, within the recent ‘Comunica’ project, a comprehensive set of speech training tools has been developed (Vaquero et al., 2006). The phonation skills module presents games for the training of five speech skills (voice activity, intensity, breathing, tone and vocalization). Another module is oriented at training the articulation abilities of the patient in isolated words and short sentences.

![Screenshot from ‘Comunica’ games](image.png)

**Figure 2.1:** Screenshots from ‘Comunica’ games. Players control the height of the butterfly with the pitch of their voice (left). Players receive visual feedback on their productions of vowel targets (right).

To conclude, the process of engaging patients in the training of speech motor skills involves a number of challenges, pertaining to a systematic demonstration of the desired speech targets, providing adequate feedback, and stimulating patients to perform a large amount of exercises. We address these questions in Chapter 5, where we examine the possibilities of using a computer game to support the training of one kind of speech motor skill, namely the speech timing accuracy skills of children with MSD.
2.4 Technology in the management of stuttering

Biofeedback applications

Dedicated computer games developed for training speech skills directly related to stuttering are rare. One example is a game designed for a study on EMG biofeedback treatment (Block et al., 2004). In this study, 12 children were instructed to keep muscle activity, as detected by EMG electrodes and displayed on a computer screen, below a certain level. As a reward, children practiced with a computer game based on EMG feedback, called 'speech muscle tennis'. In this game, two players move the bats up and down by tensing and relaxing speech muscles. The participants reported that the treatment had an effect on their stuttering, and they enjoyed using a computer to learn to stutter less.

Few biofeedback applications have been developed for adults. The Computer-Aided Fluency Establishment and Trainer (CAFET) was a program which monitored vocal volume through a microphone, as well as respiration through a chest strap (Tellis, 1996). The program helped to train relaxed, continuous breathing, and 'gentle onsets', where speakers have to increase their vocal volume gradually and smoothly. The Modifying Phonation Intervals (MPI) program is a computer based application, in which adults who stutter attempt to reduce the frequency of short phonation intervals (PI), which are assumed to be functionally related to stuttering (Ingham et al., 2001).

Altered Auditory Feedback methodologies

A technology which enjoys a strong revival in the management of stuttering in recent years is AAF - Altered Auditory Feedback (Bakker, 2006). AAF can be defined as a manipulation of one's speech signal, in which the altered signal is fed back to the person throughout the act of speech. Two most common forms of AAF are delayed auditory feedback (DAF), whereby speakers hear their own voice with a short time delay, and frequency altered feedback (FAF), whereby the frequency spectrum of the speaker's voice is shifted up or down, resulting in an altered pitch.

The enhancing effects of DAF on the fluency of people who stutter were discovered decades ago (Goldiamond, 1965). Later, the group led by Howell has demonstrated similar effects with FAF (Howell et al., 1987). Since then, numerous studies have investigated both DAF and FAF, with convergent evidence that these conditions tend to reduce stuttering by 50-80% in some people who stutter (Lincoln et al., 2006). Modern micro-electronic technologies have enabled the manufacturing of miniaturized AAF devices (Stuart et al., 2003), while current generations of smart-phones or sufficient computational resources for integrating AAF applications, opening a perspective for low-cost fluency enhancing solutions.
2.4. Technology in the management of stuttering

Fluency enhancing effects

The fluency enhancing effects of AAF have been thoroughly demonstrated for reading tasks, although most authors have reported a considerable individual variability among their participants (Lincoln et al., 2010). Only a few studies have examined the effects of AAF on stuttering during spontaneous speech, with rather divergent findings (Antipova et al., 2008). Two studies reported positive results which remained stable over time (Armson et al., 2006; Van Borsel et al., 2003), while the results of others were inconclusive (Ingham et al., 1997; Zimmerman et al., 1997; Armson and Stuart, 1998). The effects of AAF procedures have not yet been well established beyond the laboratory and the clinic, neither has the effect of a prolonged exposure to AAF been investigated systematically (Lincoln et al., 2006). In a study with the SpeechEasy device, measuring fluency levels in naturalistic settings, Pollard et al. (2009) reported no group treatment effects for the nine adult participants.

![Figure 2.2: AAF hardware device (left) & a mobile AAF application (right)](image)

Theoretic accounts

Despite many decades of research on AAF methods, the observed phenomenon of fluency enhancement is not yet explained. According to the group led by Kalinowski (who has commercialized the SpeechEasy device), the effects of DAF and FAF on stuttering are attributed to the presence of a second speech signal, which is hypothesized to link speech perception with speech production by activating 'mirror neurons', thus "temporarily restoring the integrity of the neural pipeline that is compromised" (Kalinowski and Saltuklaroglu, 2003). Howell's group, however, presents an alternative account, by which AAF enhances fluent speech by acting a timekeeping process that controls execution rate directly. Howell argues that AAF does not create a second speech signal, but rather a second rhythmic signal. This concurrent rhythmic signal presumably changes the cerebellar timekeeping operation (Howell, 2004).
Background

Methodological challenges

The lack of an accepted theoretical explanation for the effects of AAF, can, at least partly, be ascribed to the fact that AAF research involves a number of methodological pitfalls, leading to a difficulty in a consistent interpretation of fluency enhancement phenomena. First, the choice of experimental task and setting has been shown to play a role (Van Borsel et al., 2007). On the one hand, the wish to keep variables under control has led most researchers to employ reading tasks. On the other hand, the social relevance of AAF methods demands a better understanding of the effects of AAF on spontaneous speech.

Second, it has been established that an individual speaker reacts differently to various AAF settings, such as the length of delay, or the amount of frequency shift (Lincoln et al., 2006). Ideally, it would be useful to examine the 'optimal' effects of AAF procedures, which would involve a different, customized setting for each individual. However, such experimental design would not be replicable. On the other hand, a study design which delivers standard settings to all participants is prone to overlook the full potential effects on each individual.

Finally, a debatable methodological aspect is the evaluation of fluency levels in individuals who stutter. Commonly used measures are percentage of stuttered syllables (%SS), and the percentage of discontinuous speech time (PDST). Fluency data based on the PDST is presented in Ludlow and Braun (1993); Grosser et al. (2001); Natke and Kalveram (2001). PDST is a measure of dysfluency which takes into account the durational impact of dysfluent events on speech (Natke et al., 2001), introduced by Starkweather in 1993. It is believed that PDST is a sensitive indication of fluency levels, since it expresses the time aspect in the experience of dysfluent speech. A longer block, for example, would reasonably have a larger impact on the experience of the speaker and listener alike, compared to a short syllable repetition. However, with %SS measures, both events would influence the overall dysfluency measure equally.

Limitations of current AAF procedures

One obvious weakness of AAF procedures is that they provide no audio signal at the moment when phrases are initiated. At that moment, the audio signal contains nothing but the silent interval which typically precedes a new phrase. The relevance of this limitation is highlighted by evidence that in spontaneous speech, stuttering is more likely to occur at the beginning of sentences (Koopmans et al., 1991). Words are more likely to be stuttered when they are at the beginning of a sentence than when they appear at the end of the sentence (Jayaram, 1984). If that is the case, AAF procedures may not provide support at a rather critical moment for speakers who stutter (Lincoln et al., 2006).

Support for this notion comes from a recent study by Saltuklaroglu et al. (2009). The authors were interested in the relative distribution of stuttering events across utterances under AAF and choral speech conditions. They used a reading
task which was broken down into smaller timed trials, requiring participants to initiate speech at the beginning of every trial. Their results show that AAF can be highly effective after speech is initiated but does relatively little to aid with initiation (84% versus 23% reduction, respectively).

Moreover, some evidence exists that AAF procedures are less effective during spontaneous speech, where more frequent phrase initiations are required, than during reading (Armson et al., 2006). The notion that DAF and FAF do not effectively assist the initiation of sentences can also be found in usability reports elicited from AAF users (Lincoln and Walker, 2007). Addressing these limitations, Chapter 6 examines the development of adaptive feedback procedures, with the idea of selectively targeting those regions of speech which are at higher risk of being dysfluent.