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

The use of technological support in communication disorders.

How development of computer-based tools can refine the treatment of Motor Speech Disorders.

LUCL
Linguistics in Leiden

Leiden University Centre for Linguistics

P.N. van Eyckhof 3

NL-2311 BV Leiden

phone: +31 71 527 2125

email: lucl@hum.leidenuniv.nl

The investigations in this dissertation were supported by a Mosaic grant (no. 017.004.028) by the Netherlands Organization for Scientific Research (NWO).

Cover illustration by Anneke Walch.

Printed and bound by Ipskamp Drukkers.

ISBN:
NUR: 896

The use of technological support in communication disorders.

How development of computer-based tools can refine the treatment of Motor Speech Disorders.

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 18 september 2012
klokke 15.00 uur

door

Daniil Umanski

geboren 26 September 1979
te Leningrad (St' Petersburg), Rusland

Supervisor:	Prof. Niels O. Schiller
PhD Committee:	Prof. Lisa Cheng Prof. Vincent van Heuven Prof. Wolfram Ziegler Dr. Yiya Chen Dr. Crit Cremers Dr. Willemijn Heeren Dr. Anja Lowit

Contents

CHAPTER 1

Introduction

> The limits of my language
> are the limits of my world.
>
> *Ludwig Wittgenstein*

1.1 Communication and its disorders

Communication is the dynamic foundation of human interaction, and the drive behind our development as a social species. We exclaim, describe, argue, joke, tease, explain, protest, question, laugh, and express all other urges of our inner world through communication. It is a deep need of a human being to communicate, as well as a basic human right. For communication to be successful, people must be able to generate, exchange, and understand messages of one another. Although we all use many different forms of communication, natural language and speech are the predominant channels of interacting with our environment.

For those people who have difficulties with processing language, or with producing and comprehending speech, communication can be a significant challenge. Speech and language difficulties may be temporal or permanent; they may be due to congenital impairments, or result from brain injury or illness. The general term referring to these conditions is *communication disorders*. Individuals with communication disorders may face difficulties with social inclusion, with access to services and information, resulting in an overall decreased quality of life. Particularly, children who have difficulties speaking may need a great deal of support to overcome the challenges on their way towards learning to communicate. In order to provide this kind of support and, more generally, improve the quality of life of people with communication disorders, clinical professionals are trained in the area of speech-language therapy.

The scope of communication disorders which speech therapists treat may differ between countries and health insurance systems. Disorders in which speech and language functions are affected directly may include: hearing impairment; voice and resonance disorders; articulation disorders; fluency disorders; aphasia; Motor Speech Disorders; specific language impairment (SLI) and language delay. In certain cases, speech and language functions may become affected as a consequence of the following conditions: cognitive impairment and disability; behavioral and emotional disabilities (e.g., autism, attention deficit); psychiatric disabilities (e.g., schizophrenia, the dementias); structural abnormalities, including congenital (e.g., cleft palate) and acquired abnormalities (e.g., laryngectomy); cerebral palsy and other neuro-motor impairment; swallowing disabilities.

Clinicians make use of a large variety of methods, techniques and tools for treating patients with communication disorders, among them technology-based methods. This thesis explores a number of ways in which technology can support the work of speech-language therapists in dealing with one type of communication disorders, namely Motor Speech Disorders (MSD) (dysarthria, apraxia, and stuttering). In this chapter, we shortly introduce the field of speech-language therapy, the role of technological innovation in that field, and outline the research and development activities described in this thesis.

Speech and Language Therapy

Originally, the term *Logopedics* was used to identify the professional area of treating communication disorders. Currently, the term most commonly used in Europe is *Speech Language Therapy (SLT)*, and in the United States, Australia and Canada - *Speech Language Pathology (SLP)*. Throughout this thesis, when referring to professionals in the field of communication disorders, the terms 'speech clinician' and 'speech therapist' will be used interchangeably.

In the process of treating patients with Motor Speech Disorders, a significant aspect of the work of speech therapists consists of teaching, training, rehearsing and sustaining new speech and language skills with their patients. In order to perform these tasks, clinicians (and respectively, their patients) may have the following needs:

» Have the means to demonstrate the desired speech targets to their patients , in a consistent and systematic way.

» Have access to detailed and objective information about the speech productions of their patients.

» Have access to standardized diagnostic and treatment methodologies.

» Have the means to engage patients in repeated drill exercises, especially with young children.

» Have the means to effectively measure treatment results, and follow treatment progress in the course of time.

» Have methods to enable autonomous training, and management of communication outside the clinic.

» Have possibilities to effectively exchange clinical experience with other specialists, and share best-practice information.

The need for systematic, effective, objective, engaging and autonomous features of treatment delivery have led researchers to consider the ways in which technological tools can support the process of providing speech language therapy. The inertia of technological innovation in the field has been driven by a number of considerations. First, the large amount of time spent on mass-practice drills with patients may not be an efficient use of clinician's time, which is, by default, in shortage. Clinical time could be better spent if patients could perform the routine drill elements of their treatment semi independently, under the guidance and supervision of the clinician. Second, computer-based methods allow the creation of conceptually novel tasks, which may go beyond the paper-based training methods (Hubbard, 2006).

1.2 Technological innovation in SLT

Technological developments have had a considerable impact on the area of health care. Speech-language therapy practice is relatively not heavily based on technology, although increasingly more technological developments have been introduced into the field over the last decade. Broadly speaking, technological innovation in the area of communication disorders can be associated both with the field of assistive technology (AT), as well as with Computer-Aided Language Learning (CALL).[1] AT devices are tools for enhancing the independent functioning of people who have physical limitations or cognitive impairments. CALL includes a wide range of applications and tools for learning languages. Technology in the field of SLT can be roughly categorized into several types:

Prosthetic devices aim at the rehabilitation of hearing functions, include hearing aids and cochlear implants.

Altered Auditory Feedback devices manipulate the user's speech signal, so that the altered signal is fed back to the person throughout the act of speech. AAF devices are mostly used by individuals who stutter for the enhancement of speech fluency.

Augmented Alternative Communication (AAC) devices are used to supplement or replace speech for individuals with impairments in the production of spoken language. AAC is used by people with a wide range of speech and language impairments, such as cerebral palsy, autism, amyotrophic lateral sclerosis and Parkinson's disease.

Computer Aided Speech Language Therapy (CASLT) tools provide speech clinicians with a platform for teaching and training speech and language skills of individuals with speech impairments, which is the most common form of technological support in the treatment of MSD.

Sign language learning systems provide support for the education of sign languages, as well as automatic sign language recognition and translation.

Communication stimulating systems are designed to assist individuals with intellectual impairment and autism to explore communicative interactions in a safe and controlled environment, as well as to promote collaborative and social skills.

Tele-health systems utilize advanced telecommunications technologies for providing remote speech-language therapy services in the treatment of fluency, voice, swallowing and childhood speech and language disorders.

---

[1] In this context we do not include technologies underlying various *research* instruments in the study of communication disorders (such as Electromagenetic Articulography (EMA), Ultrasound Tongue Imaging (UTI) or brain imaging techniques), but rather refer to technologies employed in the work of speech therapists.

New approaches to innovation in SLT

While some scientific fields have been traditionally involved in the support of communication disorders (such as language technologies, phonetics and clinical linguistics), new approaches and combinations are steadily taking form. In recent years, there has been an increasing interest in applying methods from the field of computational linguistics (CL) to the management of communication disorders. For example, advanced language models have been utilized for improving automatic speech recognition of speakers with dysarthria (Sharma and Hasegawa-Johnson, 2010), Natural Language Generation techniques have been applied to support narrative capabilities of children with complex communication needs (Black et al., 2010), and n-gram language prediction models were shown to improve the rate of binary switch typing on Augmentative and Alternative Communication (AAC) devices (Roark et al., 2010).

Original applications may not be limited to the linguistic domain. For example, Brundage (2007) examined the potential of using the cognitive aspects of interactions in virtual reality environments for the assessment and treatment of stuttering. Light and Lindsay (1991) have applied insights from cognitive science (the limitations of working memory, the knowledge structures of long-term memory) for improving the design of AAC systems. Developments of this kind point out the advantage of a wide-angle view on the process of innovation in SLT, which concurrently considers the abilities of several expertise areas in advancing development, as well as the prospect of integrating different solutions into broader, more systematic treatment platforms.

Challenges for technological innovation in the field

A number of challenges are encountered in the process of developing technological innovation in the field of communication disorders. First, the inherent diversity among disorder manifestations and individual patients makes it difficult to develop generic tools and products. Since innovative tools in this field often rely on speech and language technologies, their development involves a considerable investment of resources, which becomes harder to justify if eventually products can not be generalized and widely used (Ruiter et al., 2010).

Second, innovation efforts in the field of communication disorders occur rather sporadically, as there are only few dedicated research and development structures to advance these efforts. This stands in stark contrast to other fields of assistive technology (e.g., physical rehabilitation) for which ongoing research activities are organized within special interest groups, research consortia, specialized centers and laboratories. This situation is, perhaps, the main obstruction in propelling innovation in the field. Therefore, this thesis strives to complement efforts towards establishing an ongoing scheme of innovative development for supporting the management of communication disorders.

1.3 Project overview

The aim of this thesis is to apply and examine technological innovation in the delivery of treatment for patients with Motor Speech Disorders (MSD). This aim is pursued from two complementary perspectives. Firstly, this thesis explores the question of how new computer-based tools can refine current methodologies in the treatment of MSD. On a second, and broader level, this thesis aims at examining the necessary ingredients for a durable process of innovation in the field, in terms of an interdisciplinary research and development framework.

The purpose of activities on the first level is to investigate the challenges and opportunities of technology-based innovation, when considering a range of steps in the delivery of treatment procedures. Specifically, the work in this thesis addresses the following three main phases in the treatment delivery of MSD:

The preparation of treatment programs. In the first stages of MSD treatment delivery, speech clinicians need to plan and design the treatment program according to the needs of individual patients. Within a treatment program, linguistic items must be selected which best target the relevant aspects of the speech disorder (background information on this topic is provided in Chapter 2.2). Assisting clinicians with the selection of speech treatment targets involves two steps - obtaining relevant linguistic materials, and providing the means for an effective generation of targets.

In Chapter 3 we address the first step, and demonstrate how techniques of computational linguistics can be brought into service for obtaining valuable linguistic materials. We obtain the syllabic inventory of Dutch by means of automatic syllabification of the spoken language data. Consequently, we analyze the suspected discrepancy between the spoken and canonical syllabic inventories, and prove insights on the applicability of the two data sets as speech training materials. The content of this chapter is partially extracted from the following publication:

> Umanski, D., Sangati, F., and Schiller, N. O. How spoken language corpora can refine current speech motor training methodologies. In *Proceedings of the ACL 2010 Student Research Workshop*, p. 37–42.

The work described in Chapter 4 demonstrates how theoretical advancements in the field of Apraxia of Speech (AOS) can be applied to the development of a new computer application for allowing clinicians to generate customized treatment materials for their patients. We outline research evidence for the benefits of a systematic manipulation of speech targets based on syllabic parameters for the treatment of AOS. Based on these insights, an application for generating word-lists according to clinically relevant parameters is developed and evaluated.

Learning and practicing speech motor skills. Throughout the treatment process, much time is spent on practicing speech motor skills. The challenge of clinicians in engaging patients in training consists of a systematic demonstration of the desired speech targets, providing adequate feedback on the quality of speech productions, and stimulating patients to perform a large amount of exercises (background information on this topic is provided in Chapter 2.3).

In addressing these challenges, in Chapter 5 we examine the possibilities of using a computer game to support the training of timing skills in speech of children with MSD. To this end, we investigate what are the elements necessary for constructing a training system, in terms of exercise methodology, the required technology, and usability factors. We implement timing accuracy exercises within a computer game, and evaluate the feasibility of the game as a supplementary therapeutic tool. The content of this chapter is partially extracted from the following publication:

Umanski, D., Schiller, N. O., Kogovšek D. & Ozbi M. Development of a voice-based rhythm game for training speech motor skills of children with speech motor disorders. In *Proceedings of the 2010 8th International Conference on Disability, Virtual Reality and Associated Technologies. p. 255–262.*

The autonomous management of communication. After a treatment period, individuals may need to maintain the acquired skills, and manage their communication independently. One example of technology supporting this process is the utilization of Altered Auditory Feedback (AAF) devices by some adults who stutter, in order to retain improved fluency in everyday speech. The limitations and challenges associated with current AAF procedures are described in Chapter 2.4.

In Chapter 6, we address these limitations, and propose a refinement to the current AAF procedures for the enhancement of speech fluency. For this purpose, an *adaptive* feedback procedure is proposed, which utilizes digital signal processing techniques for real-time analysis of the speech signal, and the dynamic activation of auditory feedback. The proposed AAF procedure is evaluated with adults who stutter, in terms of fluency enhancement, as well as comfortability with the auditory cues. The content of this chapter is partially extracted from the following publication:

Umanski, D. and Schiller, N. O. How do adaptive auditory feedback procedures affect the speech of adults who stutter. *Stem-, Spraak- en Taalpathologie 17 (supplement), p.82*

Methods and approach

In addressing each of these steps, the process of development is grounded on the frameworks of Instructional Systems Design (ISD) (Peterson, 2003), and User-Centred Design (Gould and Lewis, 1985; Hersh, 2010). According to the principles of these models, the following basic steps are involved in the process:

Analysis Phase involves the identification of a problem, a need, or an improvement opportunity, in regard to a specific methodology or procedure in the treatment of MSD. This phase may include observations of clinical practice, interviews with clinicians and patients, and literature research.

Design Phase involves the consideration of alternative strategies for solving the identified problem, while considering the areas of expertise, whose methods can contribute to attain the solution. This phase may include the specification of requirements, preparation of user scenarios, use cases, task flows and navigation structure.

Development Phase involves the implementation of the proposed solution strategy, usually on a level of a prototype, where the core functionality can be tested. It is considered a good practice to present users with early prototypes for evaluation, and iterate by re-designing in order to account for problems identified in user testing.

Evaluation Phase involves the testing of the developed solution, either by evaluating the performance of the developed techniques directly, or by means of a study with the user group. This step proceeds with the analysis of quantitative and qualitative results, in order to generate insights about the feasibility of the proposed solution, and provide recommendations for further research and development.

Quite naturally, upon reading about ‘new methods for speech therapy’, some readers might expect to find reports about the effects of these methods on the speech impairment of patients. However, the methods proposed in this thesis are not comprehensive clinical interventions, but rather experimental and complementary solutions. Therefore, before these methods can be fully developed and undergo efficacy testing, preliminary studies are necessary to evaluate the feasibility of the developed methods as valuable therapeutic tools. By definition, the purpose of a feasibility study is to evaluate and analyze the potential value of a proposed new method, in order to determine if the investment of resources in this method is likely to generate a desirable result. This is important to keep in mind when interpreting the reported results. The objective of this thesis is not to provide conclusive evidence for specific effects of the proposed methods, but rather to delineate the limitations and potentials of these methods.

Finally, this thesis is concerned with the ways in which future efforts of innovation in the field may become more productive, efficient and sustainable, by examining the ingredients necessary for an effective process of development. Through the studies described in this thesis, we hope to provide evidence for the benefits of a collaborative approach to providing support for the field. Furthermore, in order to contribute to the ongoing efforts of creating structured frameworks for technological innovation in the field of SLT (Cucchiarini et al., 2008), we will propose a model of interdisciplinary collaboration, and outline a possible workflow for such a framework.

Personal motivation

The urge for dealing with technological innovation for the support of speech and language therapy arises from my personal experience with both domains. Being a person who stutters, I am aware of the importance of advancing speech therapy methodologies. Coming from a background of computer science and interaction design, I am aware of the many opportunities of applying technology for enhancing the ability of people to communicate. The intersection of these experiences has driven me to identify innovation potentials in various methods of treatment delivery, and work on new solutions to problems in the field.

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 Goffman, 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, suffice 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).

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 (Duffy, 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 (Duffy, 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).

Childhood Apraxia of Speech (CAS) is a neurological motor speech disorder of unknwon 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 a 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, 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 Wheeldon, 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 quantitive 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 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 affect 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., DAdada or daDAda),[1] while 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 (Namasivayam 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.

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 offer sufficient computational resources for integrating AAF applications, opening a perspective for low-cost fluency enhancing solutions.

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.

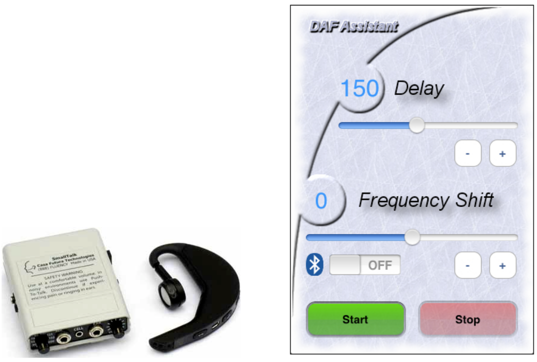


Figure 2.2: AAF hardware device (left) & a mobile AAF application (right)

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 affecting 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).

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.

CHAPTER 3

Utilizing techniques of Computational Linguistics for obtaining speech treatment materials

> Words – so innocent and powerless as
> they are, as standing in a dictionary,
> how potent for good and evil they
> become, in the hands of one who
> knows how to combine them!
>
> *Nathaniel Hawthorne, 1847*

3.1 Introduction

In this chapter, we wish to demonstrate how techniques of Computational Linguistics can be employed for refining the linguistic material needed for generating customized treatment targets. We obtain the syllabic inventory of Dutch by means of automatic syllabification of the spoken language data. Consequently, we analyze the suspected discrepancy between the spoken and canonical syllabic inventories, and prove insights on the applicability of the two data sets as speech training materials.

Motivation for obtaining a syllabic inventory

Traditional speech motor treatment programs are based on a rather static inventory of speech items, and speech clinicians usually do not have access to a methodological way of selecting treatment targets for training. In order to provide clinicians with effective computer-based tools for compiling customized treatment targets and exercises for their patients, a database of linguistic materials is necessary, which is annotated with clinically relevant parameters. As described in Chapter 2.2, clinicians may utilize different linguistic units for creating speech exercises, such as syllables, words, non-words or phrases. Particularly, the possibility to generate custom non-word forms can be a valuable supplement for the clinician's toolbox.

For example, in SMT exercises,[1] children are asked to repeat non-word syllable sequences modeled by the clinician. Since non-word repetition has been proposed as a reliable index of phonological memory (Sahlen et al., 1999), non-word reading task has been used in the assessment and training of phonological awareness skills of children with childhood apraxia of speech (CAS) (Moriarty and Gillon, 2006; McNeill et al., 2009). Since a parametric generation of non-word forms requires a pool of syllables to be combined, it would be beneficial to have access to the syllabic inventory of a language, so that treatment targets can be selected based on relevant syllabic combinations.

This importance of obtaining a detailed syllabic inventory is supported by empirical findings from clinical linguistics. For example, Staiger and Ziegler (2008) report that syllable frequency and syllable structure play a decisive role with respect to articulatory accuracy in the spontaneous speech of patients with apraxia of speech (AOS). In a recent study with AOS patients, Schoor et al. (2012) have shown that syllable learning effects spread to unlearned syllables with a specific structural kinship with the training syllables, i.e., with a position-true overlap on syllable constituents of the training and the trans-

---

[1] Speech Motor Training (SMT) is a common methodology employed in the Netherlands for teaching speech motor skills to children who stutter, as well as to children with other motor speech disorders. The complexity of syllable sets varies systematically in terms of number of syllables (1-4), the syllable structure (CV, VCV, VCCV, CVCCVC, etc.), and the number of voiceless consonants in the set (Riley and Ingham, 2000).

fer syllables. More specifically, the inclusion of coarticulatory adjustments was crucial for the training success. Considering the clinical perspective, (Schoor et al., 2012) argue for the benefits of treatment protocols based on syllabic targets, because at the phonetic encoding level the phonemes of syllables and words are not context-independent entities, but are rather embedded in the structural framework of a syllable, including coarticulations and transitions. This reasoning can be extended when considering practicing with non-word syllable sequences, and suggest the benefit of embedding exercises within a prosodic structure. Studies have demonstrated that performance on non-words repetition task is influenced by the prosodic structure of the non-word (Roy and Chiat, 2004), the phonotactic probability of the constituent phonemes (Edwards et al., 2004), and the articulatory complexity of the nonword (Archibald and Gathercole, 2006). In particular, investigating the effects of coarticulation on non-word repetition, (Archibald et al., 2009) report that valid coarticulatory cues (across syllabic units), as well as within-word stress information facilitated non-word repetition performance.

In summary, in order to allow a systematic assembly of treatment targets based on non-word syllable sequences, it is necessary to obtain a syllabic inventory which contains detailed information on both intra-syllabic and inter-syllabic levels. The inner composition of syllables is useful in order to allow a parametric search for structurally related syllables, such as obtaining units with certain phonological neighborhood relationships. Apart from listing single syllable types of a language, it would be beneficial to obtain a list of syllable pairs which have been uttered consecutively. The specification of 'syllable- tactics' (inter-syllabic constraints), such as frequencies of syllable combinations or stress patterning, can be useful for placing the exercise in a framework of natural speech production Odell (2002).

Accessing spoken language

When aiming at constructing a syllabic inventory of a spoken language, several linguistic sources can be accessed. The most readily available solution is scanning a dictionary or a lexicon of a language, which most often include the canonical syllabification of words. However, these data lack any contextual information about language use, such as frequency of occurrence for various syllable types. Next, one can rely on the widely available corpora of written language, such as the often used CELEX corpora for English, German, and Dutch (Baayen et al., 1996). These corpora do include frequency information for the occurring word forms and syllable types, but corpora which are mostly based on newspaper text may not be faithfully representative of the patterns in spoken language. Therefore, corpora which transcribe natural spoken language use may be a more appropriate source for obtaining syllabic inventories, both for the academic study of speech production phenomena, as well as for preparing speech training materials.

As one embarks on surveying the transcribed utterances in a corpus of spoken language, one faces a choice between two representations of spoken words – the *canonical* pronunciation of words (as specified in dictionaries), or the *actual* phonetic realization of words (spoken word forms), as they are actually uttered by speakers. Speakers most often deviate from the canonical pronunciation, producing segment reductions, deletions, insertions and assimilations, throughout spontaneous speech (Mitterer, 2008; Greenberg, 1999). Statistical analysis of a manually annotated subset of the Switchboard corpus indicates that pronunciation variation observed in spoken American English is highly structured at the level of the syllable, particularly when prosodic stress accent (i.e., syllable prominence) is taken into account (Greenberg, 1997, 2003). Interestingly, the authors point out that the phenomena of pronunciation variation should not be viewed as a side-effect of spontaneous speech production, but rather as reflecting exceedingly high-level processing, as specific patterns of pronunciation are likely to be indicative of the speaker's projection of the listener's internal knowledge model (Greenberg et al., 2002).

In connected speech, syllable boundaries may differ from a word's canonical syllabification due to morphophonological processes such as inflection, derivation or cliticisation (Booij, 1996). Furthermore, spoken syllables change as a result of suprasegmental processes: a syllable can be intensified, reduced, or lengthened, as a results of stress, or its position within a word. In American English, the phonetic identity of vocalic syllable nuclei as well as the probability of deletion in the coda are largely associated with stress accent (Greenberg et al., 2003). Studies of pronunciation variation suggest that the syllabic inventory of natural speech may be rather distinct from the canonical syllabary represented in a lexicon of a language. Therefore, it would be beneficial to analyze the suspected discrepancy between the spoken and canonical syllabic inventories, and to evaluate in how far the two data sets may be appropriate as materials for speech training.

Both types of syllabic inventories have their potential merits and disadvantages in the context of speech training. Since the phonetic realizations of word forms represent the way people produce speech more faithfully than canonical forms, spoken word forms may constitute a more natural speech material, especially when considering the inter-syllabic level, where the effects of contextual assimilation across word boundaries are captured within the spoken syllabary. The importance of contextual assimilation information can be demonstrated in studies of synthetic speech, where the absence of coarticulatory cues across word boundaries has been shown to lead to a decrease in sentence recall, as compared to naturally produced speech (Paris et al., 2000; Werner et al., 2004). As far as their disadvantage, spoken word forms may include very irregular phonological patterns, which may not be appropriate to be used as speech training material. The canonical syllabification of word forms represents the other side of the tradeoff – while canonical syllable types are, by default, regular and normative, they may not capture the context-dependant patterns of connected speech.

In order to engage with the analysis of both kinds of syllabic inventories, we need to obtain the corresponding data sets. The growing availability of annotated language corpora presents new opportunities for retrieving valuable linguistic data. However, phonetic annotations of spoken word forms in a corpus typically report the phoneme sequence as it was uttered by the speaker, and provides no information on syllable boundaries. Therefore, we turn to examine how a syllabic inventory can be obtained from a corpus of spoken language through an automatic syllabification of spoken word forms.

Automatic syllabification techniques

Several principles have been proposed to model the process of automatic syllabification, ranging from predefined sets of rules to data-driven models. Rule-based automatic syllabification methods attempt to enforce one or more phonotactic principles, i.e., maximum onset, sonority sequencing, legality theory, optimality theory (Bartlett, 2007). In contrast, data-driven methods utilize language corpora and statistical machine learning techniques to rely on the patterns inherent in the data itself. Supervised automatic syllabification have been mainly performed on English, German and Dutch. Although different works in the literature are not always easy to compare, due to different corpora and different evaluation metrics, state of the art systems include Hidden Markov Models (Bartlett et al., 2008), Joint N-Gram Models (Schmid et al., 2007), Syllabification by Analogy (Marchand et al., 2009), Context-Free Grammars (Tian, 2004), and Neural Networks (Daelemans and van den Bosch, 1992).

3.2 Methods

In order to obtain the syllabic inventory of spoken Dutch, a study on automatic syllabification of spoken word forms has been carried out. The Corpus Gesproken Nederlands (CGN) is a large corpus of spoken Dutch[2]. The CGN contains manually verified phonetic transcriptions of 53,708 spoken forms, sampled from a wide variety of communication situations. A spoken form reports the phoneme sequence as it was actually uttered by the speaker, but includes no annotation on how the phoneme sequence is segmented into syllables.

We define the task of syllabification as the assignment of syllable boundaries to the phonetic transcription of a given word. Two methods for dealing with the syllabification task were proposed, the first based on an n-gram model defined over sequences of phonemes, and the second based on statistics over syllable units. Both algorithms accept as input a list of possible segmentations of a given phonetic sequence, and return the one which maximizes the score of the specific function they implement. The list of possible segmentations is obtained by exhaustively generating all possible divisions of the sequence, satisfying the

---

[2] See http://lands.let.kun.nl/cgn/

condition of keeping exactly one vowel per segment. Both methods are trained on statistics over the canonical pronunciation data of the CGN lexicon.

The first method is a reimplementation of the work of Schmid et al. (2007). The authors describe the syllabification task as a tagging problem, in which each phonetic symbol of a word is tagged as either a 'B' (indicating a syllable boundary after the phone) or as 'N' (indicating no syllable boundary). Given a set of possible segmentations of a given word, the aim is to select the one, viz. the tag sequence $\hat{b}_1^n$, which is more probable for the given phoneme sequence p_1^n, as shown in equation (3.1). This probability in equations (3.3) is reduced to the joint probability of the two sequences: the denominator of equation (3.2) is in fact constant for the given list of possible syllabifications, since they all share the same sequence of phonemes. Equation (3.4) is obtained by introducing a Markovian assumption of order 3 in the way the phonemes and tags are jointly generated

$$\begin{aligned}
\hat{b}_1^n &= \arg\max_{b_1^n} P(b_1^n | p_1^n) && (3.1)\\
&= \arg\max_{b_1^n} P(b_1^n, p_1^n)/P(p_1^n) && (3.2)\\
&= \arg\max_{b_1^n} P(b_1^n, p_1^n) && (3.3)\\
&= \arg\max_{b_1^n} \prod_{i=1}^{n+1} P(b_i, p_i | b_{i-3}^{i-1}, p_{i-3}^{i-1}) && (3.4)
\end{aligned}$$

The second syllabification method relies on statistics over the set of single syllables and bi-grams (two consecutive syllables) in the training corpus. Broadly speaking, given a set of possible segmentations of a given phoneme sequence, the algorithm selects the one which maximizes the presence and frequency of its segments. More precisely, for each given segmentation t_1^m, we compute four different scores $P_1(t_1^m)$, $P_2(t_1^m)$, $P_3(t_1^m)$, $P_4(t_1^m)$ summarized as follows:

$$\begin{aligned}
P_1(t_1^m) &= \sum_{i=1}^{m} exists(t_i) && (3.5)\\
P_2(t_1^m) &= \sum_{i=1}^{m-1} exists(t_i, t_{i+1}) && (3.6)\\
P_3(t_1^m) &= \sum_{i=1}^{m-1} freq(t_i, t_{i+1}) && (3.7)\\
P_4(t_1^m) &= \sum_{i=1}^{m} freq(t_i) && (3.8)
\end{aligned}$$

The function $exists(t_i)$ in equation 3.5 returns 1 if the segment t_i has been observed in the training corpus at least once, and 0 otherwise (analogously for bisegments of equation 3.6), and $freq(t_i)$ in equation 3.8 returns the total number of occurrences of t_i in the training data (analogously for bisegments in equation 3.7).

Next, we define a total order over the set of possible segmentations of a given phoneme sequence. For every two segmentations T and T' we compare the alpha scores in a priority fashion as follows:

1. We give absolute priority to the number of segment types observed in the training data ($Ꝑ_1$). If this number is different in the two variations, we consider the segmentation with highest $Ꝑ_1$ as the more appropriate one.
2. Only if $Ꝑ_1$ is equal in the two variations we rely on $Ꝑ_2$ which corresponds to the number of bisegment types observed in the training data.
3. If necessary, the same idea is iterated for the remaining alpha scores, giving priority to bisegment token frequencies ($Ꝑ_3$) over single fragment token frequencies ($Ꝑ_4$).

Given the set of possible syllabifications $T_1; T_2; :::; T_k$ of a given phoneme sequence, we will select the one which is maximal in respect to this order: $\hat{T} : \hat{T}$ ½ T_i for i 2 $f1;2;:::;kg$.[3]

3.3 Results

The first step involved the evaluation of the two algorithms on performing syllabification of canonical word forms. Four corpora comprising three different languages (English, German, and Dutch) were evaluated: the CELEX2 corpora for the three languages, and the Spoken Dutch Corpus (CGN). In this step, the algorithms are trained on a subset of the lexicon with syllabification data provided, and are then tested on another subset of the lexicon, with the syllabification data omitted.

A 10-fold cross validation on each of the corpora was performed to evaluate the accuracy of both methods. The evaluation is presented in terms of percentage of correct syllable boundaries, and percentage of correctly syllabified words.[4]

Table 3.1 summarizes the results obtained. For the CELEX2 corpora, both methods produce almost equally high scores, which are comparable to the state of the art results reported in Adsett and Marchand (2009). For the CGN, both methods demonstrate high scores, with the phoneme-level method showing an advantage, especially with respect to correctly syllabified words.

---

[3] All ties are broken even.

[4] Note that recall and precision coincide since the number of boundaries (one less than the number of vowels) is constant for different segmentations of the same word.

Corpus	Number of words	Phonemes-based method		Syllables-based method	
		Boundaries	Words	Boundaries	Words
CGN_Dutch	86,551	98.62	97.15	97.58	94.99
CELEX2_Dutch	282,905	99.12	97.76	99.09	97.70
CELEX2_German	311,951	99.77	99.41	99.51	98.73
CELEX2_English	88,047	98.86	97.96	96.37	93.50

Table 3.1: Summary of syllabification results on canonical word forms, in terms of percentage of correct syllable boundaries, and percentage of correctly syllabified words.

The next step is applying the syllabification algorithms on the spoken word forms. The process of evaluating syllabification of spoken word forms is compromised by the fact that there exists no 'gold' annotation for this task in the corpus. Therefore, the next step involved applying both methods on the data set and comparing the two solutions. The results revealed that the two algorithms agree on 94.29% of syllable boundaries and on 90.22% of whole word syllabification. Based on the high scores reported for canonical word forms syllabification, an agreement between both methods most probably implies a correct solution. The 'disagreement' set represents the class of ambiguous cases, typically involving a consonant cluster, where syllable boundaries can be assigned in multiple ways.[5]

Through a manual evaluation of the disagreement set, we were able to conclude that both methods weaken in their performance when tested on spoken word forms, revealing an advantage for the phoneme-based method. For the syllable-level method, the reason could be the drop in syllable coverage in respect to the training corpus. Since this method is trained on statistics over canonic syllables present in the lexicon, it runs into trouble if the set of spoken syllables differs from the training set. A possible explanation for the weakness of the phonemes-level method is that it relies entirely on sequences of single phonemes, without utilizing higher organizational units. This seems an apparent limitation, considering that humans rely on multiple levels of representation in the comprehension and production of a speech signal (Guenther, 1995). Motivated by the high agreement score between the two algorithms, we have compiled the syllabic inventory by means of syllabifying the sub-set of spoken word forms in the CGN, for which the two methods are in agreement.

---

[5] As an example, consider the following pair of possible syllabifications, on which the two methods disagree: [**or-sprɔŋk-lə-kə**] vs [**or-sprɔŋ-klə-kə**], pronunciation variants of the Dutch word *oorspronkelijke*, meaning original, pristine.

3.4 The obtained syllabic data

As a result of the automatic syllabification procedure, 832,236 single syllable tokens were retrieved from spoken (phonetic) word forms, comprising a total of 11,054 syllable types. Furthermore, an inventory of all bi-syllabic types in the corpus was retrieved. Bi-syllabic types are pairs of syllables which were uttered consecutively within a word, as well as syllable pairs across word boundaries within the same utterance[6]. A total of 722,814 tokens were encountered in the corpus, of which 194,152 were distinct bi-syllabic types.

In order to evaluate in how far the obtained spoken syllabaries (both single and bi-syllabic) are distinct from canonical forms, we have compiled parallel inventories, by considering the canonical syllabification of the same word forms in CGN which were used for retrieving the spoken syllabaries. In the discussion which follows, we term the set of spoken syllables – SPO, and the set of corresponding canonical syllables – CAN.

For the purpose of comparing the two syllabaries, it is useful to consider their intercession and difference sets. The intersection set includes all types which are common to both sets. A difference set contains all types, which are present in the one set, but not the other. Table 3.2 shows the 10 most frequent single syllable types in respect to the different sets. A list of the top-100 syllables and their frequencies for both the single-syllable types and the bi-syllabic types is provided in the same format in Appendix A.

SPO ∩ CAN			SPO \ CAN		CAN \ SPO	
Type	Freq. SPO	Freq. CAN	Type	Frequency	Type	Frequency
[də]	4.15	3.89	**[dɑd]**	0.33	**[gɔːt]**	0.020
[tə]	2.52	1.89	**[ɪg]**	0.32	**[wʌuw]**	0.016
[ja]	2.45	2.51	**[əd]**	0.26	**[bʌuwt]**	0.005
[xə]	1.86	1.60	**[og]**	0.17	**[tənts]**	0.003
[ɛn]	1.66	1.90	**[mɛd]**	0.11	**[ɟuw]**	0.003
[jə]	1.36	1.49	**[fel]**	0.11	**[prɔx]**	0.003
[ɪn]	1.23	1.27	**[wɑd]**	0.10	**[pɑː]**	0.003
[dɑt]	1.19	2.07	**[hɛb]**	0.10	**[sxʌuwt]**	0.002
[di]	1.11	1.26	**[əŋ]**	0.10	**[kɔrbz]**	0.002
[ət]	1.08	1.16	**[nɔɣ]**	0.09	**[tʌuw]**	0.002

Table 3.2: The top 10 syllable types and their frequencies for the intersection and difference sets of the spoken (SPO) and canonical (CAN) syllabaries. Frequency expressed as percentage of all syllable tokens in the set.

---

[6] Utterance boundaries were determined by orthographic punctuation marks ("." "?" "!").

The frequency distribution for single spoken syllables, as well as single canonical syllables, as can be seen in Figure 3.1, exhibits a Zipf curve, a result consistent with earlier reported findings (Schiller et al., 1996). According to our statistics, 4% of unique syllable types account for 80% of all encountered syllables, and 10% of unique types account for 90% of all syllables, respectively. The distribution of frequencies for bi-syllabic types supports the notion that speakers recurrently use a set of over-learned, high-frequency syllabic combinations (Cholin et al., 2006), albeit to a lesser degree than the distribution of single syllables. Results reveal that 10% of unique bi-syllabic types account for 64% of all encountered bi-syllabic tokens, and 22% of unique bi-syllabic types account for 75% of all tokens, respectively.

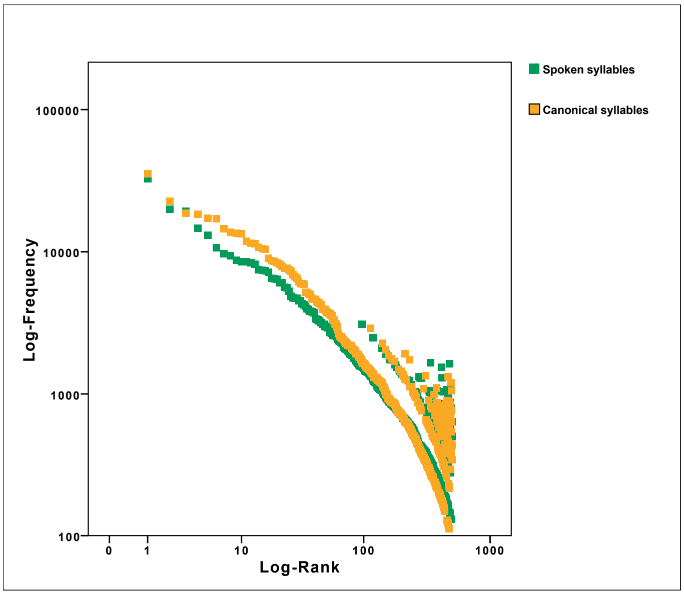


Figure 3.1: Syllable frequency distribution for the spoken and the corresponding canonical syllables in the CGN. The x-axis represents 625 ranked frequency bins. The y-axis plots the total number of syllable tokens extracted for each frequency bin.

An insightful comparison between the syllabaries can be made through examining the frequencies of various syllable structures. The Dutch language is characterized by a heterogeneous phonological structure of its syllables, which can assume a wide range of patterns due to the occurrence of consonant clusters (Vroomen and Gelder, 1994). However, the frequencies obtained from the CGN reveal a relatively consistent syllabic structure in spoken Dutch – over 85% of syllable types are of a simple syllabic structure. Table 3.3 shows the distribution of the 12 most frequent syllable structures found in the spoken syllabary (SPO), the canonical syllabary (CAN), and in the CGN lexicon (LEX),

Syllable structure	Lexicon	CAN	SPO
Simple			
[C V]	43.8	35.4	39.6
[C V C]	28.4	32.3	29.7
[V C]	4.2	14.5	14.7
[V]	2.3	4.7	3.3
Sub-total	78.8	86.9	87.4
Complex			
[C V C C]	7.3	5.4	4.9
[C C V]	5.8	2.5	2.8
[C C V C]	4.9	2.2	2.6
[V C C]	0.5	1.2	0.9
[C C V C C]	1.3	0.8	0.8
[C V C C C]	0.5	0.3	0.2
[C C C V]	0.3	0.2	0.2
[C C C V C]	0.4	0.1	0.2

Table 3.3: The relative frequency of various syllable structures found in the spoken syllabary (SPO), the corresponding canonical syllabary (CAN), and in the CGN lexicon (LEX), expressed in percentage from all syllable types in each set.

expressed in percentage from all syllable types in each set (note that while both LEX and CAN refer to canonical syllable types, the CAN counts pertain to actual instances of occurrence, while the LEX counts refer to syllable frequency within the dictionary, independent of their frequency of occurrence).

A similar widespread occurrence of simple syllabic structures has been reported for spoken English – over 83% of syllable tokens in the Switchboard corpus are of a simple structure (Greenberg, 1997). In the CGN, this pattern is almost identical for the spoken and canonical syllabaries (87.4& and 86.9% respectively). Overall, the two syllabaries exhibit a very similar frequency distribution of syllabic structures, with a slight tendency for CVC forms to reduce to CV structure (a decrease of 2.6% in CVC forms with an increase of 4.2% in CV forms for the spoken syllables in relation to canonical ones).

In order to further examine the degree of similarity between the spoken and canonical syllables, Table 3.2 summarizes the number of distinct types and the number of total tokens in the two sets, as well as in their intercession and difference sets, for single and bi-syllabic types respectively. One can estimate the divergence between the spoken and canonic syllabaries by examining the proportion between the size (the number of types) of the difference set in respect to the sets examined. In the case of single syllables, for the spoken syllables set, this proportion is: $5254/11054 = 0.47$, meaning that nearly half of the obtained types are unique to the spoken syllabary. However, this account does not consider the frequency of occurrence of syllable types. Therefore, a more precise evaluation examines the ‘weight’ of each set (the total number of tokens). This account reveals a proportion of $44,615/832,236$, meaning that only 5.36% of all uttered spoken syllables are not found in the canonical syllabary. Indeed, when examining the 100 most frequent syllables, we find that 80 of them are present in both syllabaries.

	Single-syllable types		Bi-syllabic types	
Syllable set	# Types	# Tokens	# Types	# Tokens
SPO	11,054	832,236	194,152	722,814
CAN	6572	908,076	134,753	850,163
SPO $\cap$ CAN	5800	787,621	78,621	467,655
SPO $\setminus$ CAN	5254	44,615	115,531	255,159
CAN $\setminus$ SPO	772	4904	56,132	113,965

Table 3.4: Number of distinct syllable types and of total syllable tokens, for the set of spoken syllables (SPO), the corresponding set of canonical syllables (CAN), their intersection and difference sets.

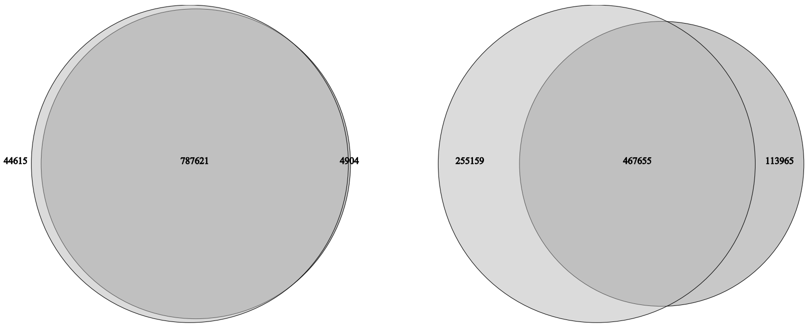


Figure 3.2: Venn diagrams of the spoken and canonic sets for single syllables (left), and bi-syllabic tokens (right). In both diagrams, the spoken set is represented in the left circle, and the canonic set in the right.

In the case of bi-syllabic types, the convergence is less pronounced. In terms of number of types, the proportion is: $27,691/48,404 = 0.57$, meaning that more than half of the bi-syllabic types are unique to the spoken inventory. When frequency is considered, the proportion is: $255,159/722,814 = 0.35$, meaning that roughly a third of all spoken syllable pairs are not found in the corresponding canonical syllabary. The divergence is further illustrated by the finding that from the 100 most frequent bi-syllabic types, only 57 are found in both the spoken and canonical sets.

3.5 Discussion and conclusions

Having defined our aim as obtaining appropriate linguistic materials for compiling customized non-word speech treatment targets, we argued for the clinical relevance of obtaining a detailed syllabic inventory of the spoken language. Considering the phenomena of pronunciation variation and contextual assimilation in connected speech, we have anticipated a divergence between a syllabary obtained from syllabification of the actual phonetic realization of words, and the one derived from the canonical pronunciation of the same words. Therefore, our goal in this project was extracting the spoken and canonical syllabaries from a corpus of spoken Dutch, examining the consistency of both syllabic inventories, and providing insights on their applicability as speech training materials.

In our study on the Dutch Spoken Corpus, syllabic inventories were obtained by means of automatic syllabification of the spoken language data. We have tested two syllabification methods, the first based on an n-gram model defined over sequences of phonemes, and the second based on sequences of syllabic units. Our results have shown that both methods were very successful in syllabifying canonical word forms. Therefore, the high agreement rate between the two algorithms when tested on spoken word forms indicates a high probability for a correct solution. The syllabic inventory has been compiled by syllabifying the set of words for which the two methods agree.

Describing the obtained syllabic data, we have compared the set of syllables derived from spoken word forms (SPO) with a set of corresponding canonical syllables (CAN), derived from the canonical syllabification of the same word forms. To this end, we summarized the frequency distribution for both single syllables, as well as bi-syllabic types, confirming previous psycholinguistic evidence for the reliance of speakers on a relatively small set of high frequency syllabic types (Cholin, 2008). Next, the relative occurrences of syllabic structures was compared, indicating that the two syllabaries exhibit a very similar frequency distribution of syllabic structures. Finally, a direct comparison of single syllable types in the SPO and CAN syllabaries has revealed a considerable convergence of the two sets, especially for the high-frequency syllable types. Spoken syllables which are not found in the canonical inventory belong to the class of very low-frequency types, and together constitute only 5.36% of all uttered syllables.

The comparison of bi-syllabic types indicates a divergence of the two sets, with roughly a third of all spoken syllable pairs not found in the corresponding canonical syllabary. The observed divergence can be attributed to the influence of contextual assimilation in syllable pairs across word boundaries, a phenomenon which is captured within the spoken bi-syllabic types, but cannot be represented within canonical types, since syllable pairs are derived from context-independent specifications in the lexicon.

The observed similarity between the spoken and canonical sets for single syllable types, in terms of syllable frequency distribution, the distribution of syllabic structures and identity of syllabic types, indicates that the set of canonical syllables form a good approximation of the spoken syllabary. Therefore, for the generation of monosyllabic speech treatment targets, the canonical syllabary can be seen as an appropriate material. For the purpose of compiling customized multi-syllabic non-word sequences, materials based on bi-syllabic types may be more appropriate, since these data capture the patterns of pronunciation on the inter-syllabic level. However, due to the observed divergence between the spoken and canonical sets, a strategy must be sought which combines the advantages of both syllabaries. One possibility is to extract the intersection set of the two syllabaries, which assures that syllable pairs are both natural (in so far as having been uttered by speakers with certain frequency), as well as normative (in so far as the constituent syllables are present in the lexicon with certain frequency). An additional step to optimize the acquisition of syllables for treatment materials is filtering the various categories of speech sources available in the corpus. For example, the CGN corpus is structured to include several categories of recordings, varying in the degree of formality of the communication context. In order to prepare materials for speech training, it is possible to include in the syllabification procedure only these categories where a normative speech style is expected, such as news broadcasts, lectures, language lessons.

An important specification on the inter-syllabic level which was not taken into account in the current study is the prosodic stress accent (i.e., syllable prominence). Although it has been shown that high-quality prosodic annotation can be obtained by non-experts when enough training is provided (Buhmann et al., 2002), manual prosodic annotation is costly, and most large-scale spoken corpora do not currently include this level of detail. Once prosodic annotation becomes available for large portions of spoken corpora, future attempts to compile a syllabic inventory for a given language should include the prominence context of syllables. Such data can further refine the procedure of generating parametrized non-word speech targets, since stress assignment for target syllable combinations can be based on the patterns found in spoken language.

The collected syllabic data may also be used for promoting ongoing psycholinguistic research. The finding that over 85% of syllable types in spoken Dutch are of a simple syllabic structure, similar to earlier reports for English (Greenberg, 1997), may support the notion that syllables play an important role in the structuring of speech, since simple syllabic structures are likely to facilitate the process of syllabic segmentation and aid listeners in the decoding of a continuous speech stream. Furthermore, by examining the relations between the single syllable inventory and the set of bi-syllabic combinations, new questions about 'syllable- tactics' in speech could be formulated. For example, one could study whether the frequency of syllable pairs is associated with the frequency of the single syllables comprising it.

Finally, it is important to note that the results and implications summarized in this chapter may not obviously hold for other languages than Dutch. Although the analyses pertaining to syllable structure and syllable frequency distributions revealed similar patterns to those reported for English, a generalization to other languages is not straightforward, and language-specific investigations with appropriate spoken corpora are necessary. Moreover, the methodology of such investigations will depend on the quality and detail of annotations available in spoken language corpora, which may considerably differ from one language to another.

Computational linguistics techniques can be utilized to uncover the statistical regularities of spoken language and provide a quantitative analysis of these patterns. With the work presented in this chapter we wish to demonstrate how the results of such investigations can be harvested for refining the materials used for generating speech treatment targets. In the next chapter, we wish to address the subsequent step in these efforts – the design of an intuitive interface for speech clinicians for allowing the compilation of customized speech targets.

CHAPTER 4

Computer-aided generation of customized speech treatment materials

All my life I've looked at words as though
I were seeing them for the first time.

Ernest Hemingway

4.1 Introduction

In this chapter, a practical tool for assisting speech clinicians with compiling customized treatment materials for their patients is described. The design of a computer application, based on current knowledge about the pathomechanism of apraxia of speech (AOS), is outlined, and its feasibility as an efficient clinical tool is evaluated through a usability study with speech therapists.

Selection of speech targets in AOS treatment programs

An insightful design of speech treatment programs for apraxia of speech (AOS) patients must address a clinically relevant choice of speech items to be used in exercises throughout the program.[1] Ideally, treatment targets should be structured according to the units on which the apraxic speech mechanism operates, as well as considering the structural properties of these units (Ziegler, 2008). Recent research findings suggest that AOS treatment targets should not be confined to the level of individual phonemes, and involve higher level speech motor units, such as syllables or metrical feet (Odell, 2002). This view receives strong support from evidence on the pathomechanism in AOS (Aichert and Ziegler, 2004; Staiger and Ziegler, 2008). The authors have conducted a comprehensive investigation into the factors which influence the proportion of speech errors of AOS patients. The primary factors identified were: (1) syllable frequency, (2) syllable position, (3) metrical structure, (4) articulatory complexity.

In an investigation of patients with mild and moderate impairment due to AOS, Aichert and Ziegler (2004) found that low frequency syllables (which were embedded in two-syllable words) were more prone to errors than high frequency syllables, a result also reported by Laganaro (2008). In addition, Aichert and Ziegler (2004) discovered an influence of syllable position on the production accuracy of consonant clusters. Rates of simplification errors varied, depending on the position of a cluster (either at syllable onset, the coda, or at the boundary between two syllables). Further, an influence of the metrical content of an utterance on the error patterns in speakers with AOS has been found. Aichert and Ziegler (2009) compared the production accuracy of two-syllabic German words with stress on the first syllable (trochees) with words stressed on the second syllable (iambs). Trochaic words were generally produced more accurately than iambic words.

Another factor influencing the error probability in patients with AOS is articulatory complexity. Patients tend to make more errors on speech units with complex internal structures (i.e., containing consonant clusters), and they tend to replace complex units by less complex ones. Aichert and Ziegler (2008) investigated whether the training of less complex syllables will generalize to benefit

---

[1] For a review on the process of selecting treatment targets, see Chapter 2.2.

the production of related, and more complex syllables (for example, whether exercising syllables like /nas/, /kas/, and /nast/ would improve the production of /knast/). This turned out to be the case, and the authors conclude that patients with AOS may re-learn complex syllables by exercising their less complex relatives (Riecker et al., 2008).

Following the indication that the architecture of speech motor plans spreads below and above the level of the syllable (Ziegler, 2009), relevant parameters for specifying treatment targets should regard the nature of certain phonemes (manner and place of articulation, the presence of voice, etc.) comprising a speech item (Odell et al., 1990), as well as the metrical pattern by which multiple syllables are combined into words.

Towards customized treatment programs

The cumulation of research evidence indicates that treatment protocols may benefit from a systematic manipulation of treatment targets according to the factors linked to the pathomechanism of AOS. Furthermore, treatment programs may achieve an optimal impact if speech targets are selected *individually* for each patient, in accordance with the specific needs of that patient. Currently, there are no available computer-aided tools that allow speech clinicians to compile individual treatment materials. Considering the strong time constraints in speech clinics, a computer-based application can render the process more time saving, efficient, and flexible. Moreover, a digital database of linguistic materials can be easily extended and updated.

The Clinical Neuropsychology Research Group (EKN) at Klinikum Bogenhausen in Munich has created a software application which allows for an automated construction and analysis of speech materials in German (Aichert et al., 2005). The program is based on materials from the CELEX database for German (i.e., 365.530 word types). However, since the application was designed mainly for research activities, it comprises a variety of rather specific and complex functions (such as the determining of syllable onset and rime frequencies), making it unsuitable for clinical purposes, in terms of practicality and usability. Considering this shortcoming, we have teamed up with the EKN in order to create an effective and 'user-friendly' computer program for speech clinicians.

As a first step, the EKN team has compiled (through a selective retrieval from the CELEX database) a set of words which are suitable for serving as treatment targets in clinical programs (Riegel and Buedel, 2010).[2] Next, words were manually annotated for parameters implicated in the pathomechanism of AOS, as described above, in 4.1. This work has provided the necessary linguistic materials for the next step – the construction of a new computer application allowing the compilation of customized speech targets for treatment programs of AOS.

---

[2] Excluding very low-frequency words, and words with more than 5 syllables.

4.2 Generating customized treatment targets

The *phoNavigate* application has been designed and developed in a tight collaboration with the EKN team. It has been decided to start out with a relatively simple functionality, allowing clinicians to generate treatment materials based on a limited scope of parameters and configurations. The application consists of two modules - 'Material Generation', where words are selected according to the described parameters, and 'List Processing', where the resulting list of lexical items can be further formatted and stored. We provide an overview of each module, and its relevant functionalities.

Material Generation

The 'Material Generation' module consists of several menus which correspond to linguistic parameters considered most relevant in the treatment of AOS (see Figure 4.1). A user indicates the number of syllables that words may contain. Selection is made from 1 to 4 syllables, also allowing for a selection of multiple sizes. Next, articulatory complexity can be determined by choosing 'simple', 'complex', or 'any'. A word is considered 'simple' only if all syllables comprising the word are of a simple structure (CV, VC, CVC, V), and, there are no consonant clusters at syllable boundaries (as in CVC.CVC, CVC.CV, or VC.CV). When a single syllable number is selected, an appropriate menu appears which allows for the specification of the word stress parameter.[3] The menu includes all possible stress patterns[4], according to the number of syllables. An additional menu allows users to indicate the desired word frequency rank. For simplicity, words are divided into 'high' and 'low' frequency ranks.[5]

The generated material can be further refined by specifying the set of permitted word-initial phonemes. To that end, an interactive table of phonemes is depicted (rendered as German International Phonetic Alphabet (IPA) symbols), which allows users to select, or de-select individual phonemes, as well as phoneme groups defined by German phonology.[6] The grid of phoneme toggle buttons, together with 'include-all' and 'exclude-all' buttons, provide the user with a flexible and fast mechanism for selecting arbitrary sets of phonemes.

The order in which menus and buttons are operated is irrelevant. Upon input gestures from the user, the application re-calculates and displays the results for the current configuration, providing a 'real-time' preview of the obtained material. Results are displayed as a list of words satisfying the current selection criteria, arranged in a 5-column table, together with a count of the retrieved items. At any point, the user can reset the interface to the initial configuration, at which no words are selected, and start afresh.

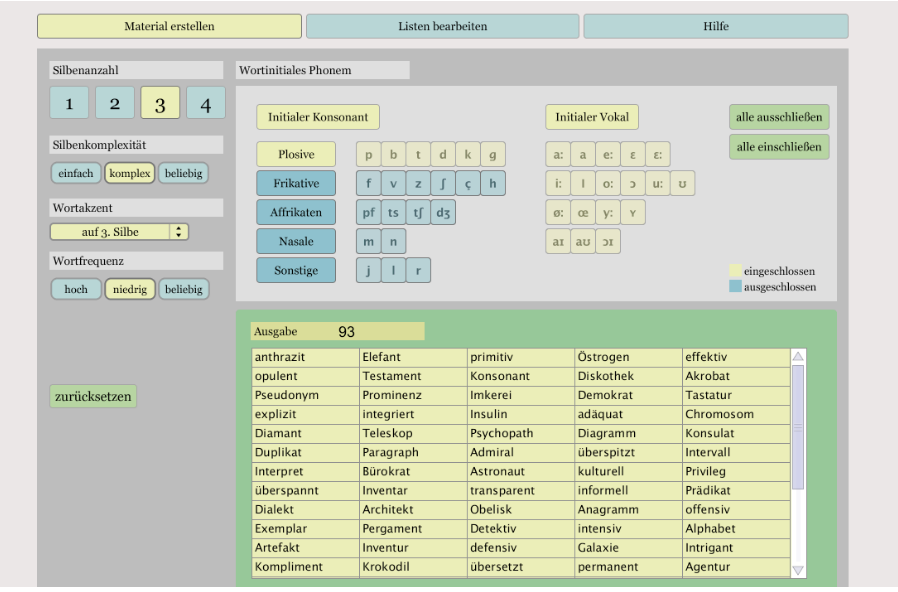
---

[3] No such menu is presented when mono-syllabic words are selected.

[4] 1st, 2nd, 3rd, or 4th syllable stressed.

[5] High and low word frequency ranks are divided by a threshold of 10 words per Million.

[6] Vowels, Vocals, Plosives, Fricatives, Affricate, Nasals, and Others.



Silbenanzahl: 3

Silbenkomplexität: komplex

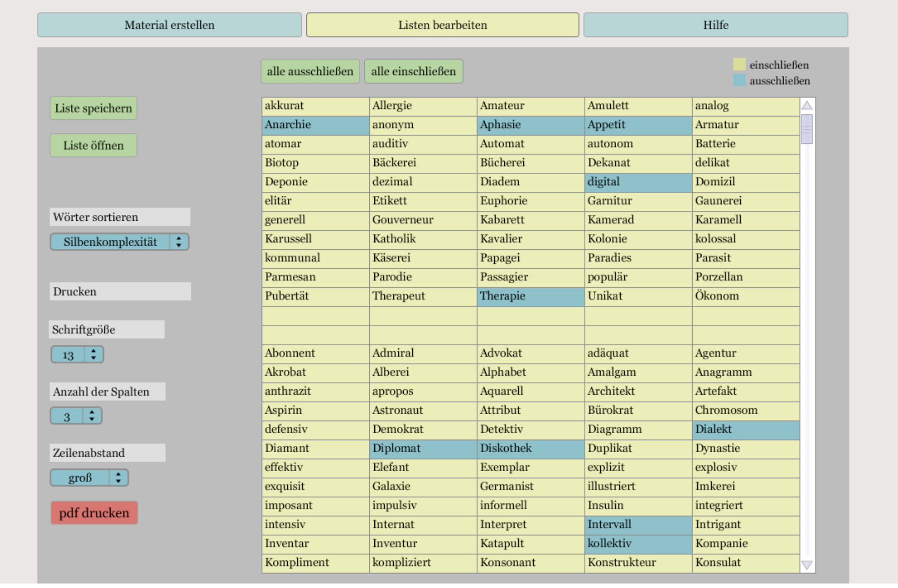
Wortakzent: auf 3. Silbe

Wortfrequenz: niedrig

Ausgabe 93

anthrazit	Elefant	primitiv	Östrogen	effektiv
opulent	Testament	Konsonant	Diskothek	Akrobat
Pseudonym	Prominenz	Imkerei	Demokrat	Tastatur
explizit	integriert	Insulin	adäquat	Chromosom
Diamant	Teleskop	Psychopath	Diagramm	Konsulat
Duplikat	Paragraph	Admiral	überspitzt	Intervall
Interpret	Bürokrat	Astronaut	kulturell	Privileg
überspannt	Inventar	transparent	informell	Prädikat
Dialekt	Architekt	Obelisk	Anagramm	offensiv
Exemplar	Pergament	Detektiv	intensiv	Alphabet
Artefakt	Inventur	defensiv	Galaxie	Intrigant
Kompliment	Krokodil	übersetzt	permanent	Agentur

Figure 4.1: A screenshot of the 'Material Generation' module



Wörter sortieren: Silbenkomplexität

Schriftgröße: 13

Anzahl der Spalten: 3

Zeilenabstand: groß

akkurat	Allergie	Amateur	Amulett	analog
Anarchie	anonym	Aphasie	Appetit	Armatur
atomar	auditiv	Automat	autonom	Batterie
Biotop	Bäckerei	Bücherei	Dekanat	delikat
Deponie	dezimal	Diadem	digital	Domizil
elitär	Etikett	Euphorie	Garnitur	Gaunerei
generell	Gouverneur	Kabarett	Kamerad	Karamell
Karussell	Katholik	Kavalier	Kolonie	kolossal
kommunal	Käserei	Papagei	Paradies	Parasit
Parmesan	Parodie	Passagier	populär	Porzellan
Pubertät	Therapeut	Therapie	Unikat	Ökonom
Abonnent	Admiral	Advokat	adäquat	Agentur
Akrobat	Alberei	Alphabet	Amalgam	Anagramm
anthrazit	apropos	Aquarell	Architekt	Artefakt
Aspirin	Astronaut	Attribut	Bürokrat	Chromosom
defensiv	Demokrat	Detektiv	Diagramm	Dialekt
Diamant	Diplomat	Diskothek	Duplikat	Dynastie
effektiv	Elefant	Exemplar	explizit	explosiv
exquisit	Galaxie	Germanist	illustriert	Imkerei
imposant	impulsiv	informell	Insulin	integriert
intensiv	Internat	Interpret	Intervall	Intrigant
Inventar	Inventur	Katapult	kollektiv	Kompanie
Kompliment	kompliziert	Konsonant	Konstrukteur	Konsulat

Figure 4.2: A screenshot of the 'List Processing' module

List Processing

The ‘List Processing’ interface (see Figure 4.2) features a table, which contains a list of selected words, and a number of functions for processing this list. First, a list can be sorted according to the criteria employed in the ’Material Generation’ module (number of syllables, complexity, word stress, frequency rank), as well as by alphabetic or random order. Upon sorting, for the sake of clarity, the resulting groups of items are separated by two empty rows.

Next, a user can select these items which are to be included in the final list, by means of toggling on and off individual cells in the table, in combination with ’include all’ and ’exclude all’ buttons. When selection is complete, a list can either be saved for later processing, or printed to a PDF file.[7] Saving a list prompts the user to enter a filename where the data is saved.[8] When printing the list to a PDF file, the user can specify certain formatting variables (font size, number of columns, vertical space between cells).

Usability study

In order to evaluate the application and collect user feedback, we have conducted a usability study with speech clinicians in Germany. In total, 19 clinicians have participated, who had various degrees of experience with the treatment of AOS. The task of the participating clinicians was to simulate two cases of specific patients (either current patients, or from their past clinical experience), for whom a customized list of treatment targets must be configured.

Clinicians were instructed to include the following steps in the process of using the application: (1) generating material, (2) saving the list, (3) opening the list, (4) sorting the list, (5) printing the list to a PDF file. Once clinicians completed their experimentation with the program, they answered a questionnaire designed to elicit usability information. Items were presented in a 4-point Likert-type scale, and included statements regarding the overall structure of the application, the specific modules, and the potential value of the application in clinical practice.

4.3 Results & Discussion

The results for each item on the usability questionnaire are presented in terms of the frequency distribution of possible answers. The accompanying comments of clinicians are summarized in order to elaborate on the results and provide insights into their experience with the application.

---

[7] We use the term *print to PDF* in order to indicate that the data is formatted and ready for actual printing on paper, and to distinguish this from saving the data for later processing.

[8] Data is saved in a format specific to the application.

(1) The structure of the application is clear.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	15.8%	84.2%

(2) The menus and the buttons in the application are clear.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	21.1%	78.9%

(3) Navigation through the application is easy.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	31.6%	68.4%

(4) Items were selected correctly, according to my specifications.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	15.8%	84.2%

(5) Working with the phonetic symbols for selecting initial consonants is easy.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	5.3%	94.7%

(6) The selection process is based on parameters which are relevant for treatment goals.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	21.1%	78.9%

(7) It was comfortable to process the selected word lists, to save and to print.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	26.3%	31.6%	42.1%

(8) It is important in my work to create customized treatment targets for my patients.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	5.3%	94.7%

(9) This application is potentially useful in my own clinical practice.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	5.3%	21.1%	73.6%

(10) I believe that speech clinicians, in general, would be interested in such an application.

Do not at all agree	Do not agree	Agree	Entirely agree
0%	0%	26.3%	73.7%

Concerning the overall structure of the program (Question 1), clinicians reported (84.2% full agreement) that the structure is very clear and it is intuitive to use the application. Some remarked that no explicit instructions are necessary. The convenience of buttons and menus (Question 2) has also been judged positively (78.9% full agreement), with a suggestion to extend the selection of word stress when words with different number of syllables are selected. This suggestion is reasonable, as it will involve the addition of two menu placeholders, which should not harm the simplicity of the interface.

Navigation through the application (Question 3) has been described as easy (68.4% full agreement), with a suggestion to allow the marking of whole rows and columns in the 'List processing' table. Since in this table words within a row or a column are not grouped according to any meaningful category, an advantageous solution allowing faster marking of words would be the dragging of a computer mouse, or a keyboard key combination.

Regarding the 'Material Selection' module, clinicians observed that items were selected correctly according to specifications (Question 4, with 84.2% full agreement), and that working with the IPA phonetic symbols presented no problem (Question 5, with 94.7% full agreement). Nevertheless, it was proposed to include an explanatory IPA chart, as a help feature for clinicians who may have more difficulties with these symbols. Clinicians confirmed (78.9% full agreement) the relevance of selection parameters for treatment goals (Question 6). Additionally, a number of extensions were proposed for future versions:

- Include possibilities for selection of word-medial and word-final phonemes.
- A possibility to select words according to their syntactic word class.
- A possibility to determine all phonemes in the selected words.
- A way to include or exclude phonemes and phoneme classes separately for each word on the list.
- A way to differentiate between intra- and intersyllabic clusters.
- Including a special class of ‘every-day-relevant’ words for selection.
- A possibility to create ‘minimal-pairs’ (words contrasting in one phoneme).

The experience of clinicians with the ‘List Processing’ module, although generally positive, has revealed a number of shortcomings (Question 7). Upon opening a previously saved list, it is currently not possible to add new words to it, which would be desirable in certain cases. Additionally, going back to the ’Material Selection’ module would reset the list currently being processed, so that the selection would have to start from scratch, which may be disadvantageous. This calls for a more flexible mechanism of updating word selection while preserving the intermediate results.

Several clinicians found the ’Print PDF’ function rather confusing, as some have expected the list to be literally printed out. It was suggested to label the button with ’Save as PDF’. Furthermore, it was proposed to offer additional formats for saving the lists (Word or Excel), in order to allow further editing of the material, as well as extending the editing options within the program. Finally, a number of clinicians suggested that a *Preview* function will be beneficial, which allows to review the formatted material before saving it. Concerning the importance of creating customized treatment targets for their patients, 94.7% of the clinicians answered with full agreement (Question 8). When asked about the potential value of the application in their own practice (Question 9), 73.6% answered with full agreement, equally so regarding the interest in the application by German speech clinicians in general (Question 10).

In their open remarks, clinicians expressed highly positive views on this line of development. For example, it was stated that the application created a much faster and easier work-flow with creating speech material than the otherwise ’time expensive’ procedures. It was stressed that clinicians tend to use standard, ’non-individual’ speech materials, and the application would greatly benefit the customization of treatment targets. Finally, it was mentioned that the program may be useful for working with other speech and language disorders than AOS.

Alongside the evaluation of the current prototype, clinicians have produced a number of proposals for functional extensions in future versions. For example, it was suggested to facilitate the creation of customized treatment exercises and worksheets, where selected lexical items are integrated into dynamic templates (such as ‘missing word’ exercises). One proposal concerned the linking of visual stimuli to the lexical items, as clinicians, otherwise, spend a considerable amount of time in the search for appropriate materials.

4.4 Conclusions

Highlighting a range of evidence concerning the linguistic units implicated in apraxic speech, we have indicated that treatment methodologies of AOS can benefit from a systematic manipulation of speech targets based on the structural make-up of these units. This proposition formed the rationale for providing speech clinicians with the means for compiling customized treatment materials for their patients. In order to realize this goal, a new computer application has been developed specifically for speech clinicians, and evaluated through a usability study.

Encouraged by the results of the study, and equipped with user-defined specifications for additional functionality, the development of a widely-used clinical application lays ahead as a promising undertaking. Future development should provide an advanced module for material generation, offering more elaborated and flexible mechanisms for specifying targets. The module will address detailed configurations of syllabic parameters (i.e., syllable internal structure, syllable frequency), the clustering of items by phonological similarity (e.g. minimal pairs), and linkage of items with visual materials. A challenge in the design of this module is in increasing the complexity of material generation, while preserving the intuitive nature of navigation through the interface.

Looking further ahead, a logical extension is the creation of an additional module allowing clinicians to construct whole treatment programs from the speech targets they have previously selected. Within this module, it will be possible to specify the different ways in which speech items are grouped into exercises, and how exercises are scheduled between treatment sessions. Principles of training structure design have been elaborated in Maas et al. (2008). According to the authors, training structure is specified by practice amount, practice distribution, practice variability, practice schedule, attentional focus, and target complexity. These principles could serve as a starting point for creating a module for treatment program construction. For example, according to the choice of practice schedule (blocked vs. random), groups of different speech targets can either be grouped in successive blocks, distributed along the treatment sessions, or randomly intermixed throughout all treatment sessions.

Finally, the composed treatment programs can, in turn, serve as input to another module in which training itself takes place, and where feedback and evaluation of performance are given and stored. The precise design of the training module depends largely on the chosen training methodology, and can vary from the bare display of speech targets on the screen, to providing real-time audio and visual cues (Mauszycki and Wambaugh, 2008; Brendel and Ziegler, 2008), or feedback on certain aspects of speech production. In any case, speech targets previously selected by the clinician will be seamlessly inserted into the displayed exercises, freeing him/her from keeping track of a pile of printed pages.

A workflow in which the selection of speech targets, composition of treatment programs, and the execution of training exercises are all contained within the same application, could facilitate a more consistent and systematic delivery of treatment. In addition, utilizing systematic treatment procedures in the context of clinical research studies could produce more reliable measures, since speech materials and modes of presentation are standardized within the application.

It is worthwhile to note that the application described in this chapter includes rather language-specific elements. While the concept of designing an intuitive computer interface which reflects composition parameters based on theoretical and clinical knowledge is universal, the specific implementation presented here reflects the particularities of German, such as the selection among typically occurring metrical patterns of words. Therefore, in order to implement such an application for other languages, it would be necessary to adapt the annotations of language-specific materials for clinically relevant parameters, as well as the design of the corresponding selection criteria.

In summary, the work described in this chapter demonstrates how computer applications can facilitate the generation of treatment materials, and allow clinicians to fine-tune the generation process according to current theoretical advancements in the field of AOS.

CHAPTER 5

Developing methods for training timing skills in speech of children with motor speech disorders

> We don't stop playing because we grow old;
> we grow old because we stop playing.
>
> *George Bernard Shaw*

5.1 Introduction

This chapter addresses the stage in MSD treatment delivery in which patients learn and practice speech motor skills. Particularly, the appropriate conditions for creating a computer game for training timing accuracy in speech of children with motor speech disorders are investigated. This chapter does not focus on the aspects of treatment effects as the result of training. Rather, our goal is to examine the necessary elements for constructing a training application, in terms of exercise methodology, the required technology, and usability factors.

Timing deficit in children with speech-language disorders

Speech motor skills entail an intricate coordination and timing of motor events throughout the act of speech. Therefore, accurate timing skills in speech can be seen as one of the elements comprising speech motor skills. Recent studies have argued for the presence of a central timing deficit in children with speech and language disorders. For example, reports have shown that children with dyslexia have problems with musical rhythmic tasks, and performance on these tasks correlated with reading abilities (Wolff, 2002; Goswami, 2010). For children with specific language impairment (SLI), slow processing rates in perceptual and motor tasks have been associated with language deficits (Schul et al., 2004; Bishop, 2002). There is some, albeit limited, evidence of motor timing deficiencies of children who stutter (Westphal, 1933; Howell et al., 1997). Olander et al. (2010) reported that a subgroup of young stuttering children exhibited a non-speech motor timing deficit, after assessing between-hands coordination and rhythmic motor timing in 17 children who stutter and 13 controls.

Particularly, a problem with timing accuracy in speech is thought to be implicated in childhood apraxia of speech (CAS). Peter and Stoel-Gammon (2008) investigated timing accuracy of children diagnosed with CAS, by means of non-word imitation, clapped rhythm imitation, and paced repetitive tapping tasks. Their results showed less timing accuracy in their CAS participants, as well as stronger associations among measures, than in typically developing controls. The authors interpret the results as suggesting a central timing deficit, expressed in both verbal and limb modalities.

Assessing timing accuracy in speech

A common way to evaluate timing accuracy is employing rhythmic patterning tasks. In the domain of speech, non-word repetition tasks are used, in which sequences of syllables in various rhythmical structures are presented, and participants are asked to imitate the pattern as accurately as possible. Often, these tasks involve stress patterning, with sequences of stressed and unstressed syllables varying in length and metrical complexity (Tjaden, 2000; Shriberg et al., 2003; Peter and Stoel-Gammon, 2008). However, stress patterns may not be optimally suitable for evaluating timing accuracy per se, since the phonetic realization of stress patterns involves an interaction between duration, amplitude and pitch modulations, making the timing aspect rather implicit (Shriberg et al., 2003; Kim and Beutnagel, 2011). Stressed syllables, besides having longer duration, tend to have higher pitch levels and loudness than unstressed syllables (Silipo and Greenberg, 1999; Kochanski et al., 2005). In order to target timing accuracy more explicitly, tasks which isolate the temporal component of syllable production may be more appropriate. Explicit timing tasks can facilitate assessment and training procedures. The evaluation of performance in stress patterning tasks is based on the perceptual assessment of the therapist or teacher (Koike and Asp, 1981), which does not allow for autonomous modes of training. For evaluating the accuracy of events within explicit timing tasks, automatic procedures are more likely to be implemented.

Practicing timing skills in speech

At this point, an important dissociation must be made – the skill addressed in this chapter, referred to as ‘timing skills in speech’ is not equivalent to the notions of ‘speech timing’, which usually refers to the *natural* temporal organization of spontaneous speech. Consequently, the proposed training methodology does not target to model and exercise the correct timing patterns of natural speech. Rather, the methodology aims at developing the sensitivity to timing in a broader sense, and targets the skill of timing speech gestures in an arbitrary temporal sequence. By placing demands on the learner to voluntary regulate the timing of producing speech segments, the learning process aims at enabling a higher degree of temporal precision in speech production, with the idea that this skill will benefit spontaneous speech.

In order to practice timing skills in speech, abstracted from imitating the timing patterns of natural speech, the training methodology can be grounded on the most basic framework of temporal organization – the musical rhythm. Although rhythmic skills may seem rather remote from language processing, strong links have been drawn between the two. For example, researchers have shown that children with SLI are impaired in rhythmic entrainment tasks such as tapping to a metronome (Thomson and Goswami, 2008). The authors consider the ability to process low-level aspects of auditory temporal structure to have consequences for rhythmic behaviors that are integral both for the sense of musicality as well as for language development. In this view, musical training may have the potential to remediate language disorders which are related to deficits in basic auditory processing. Thanks to the explicit rhythmic structure in music, interventions which involve rhythmic coordination, such as playing an instrument in time with a song, singing along, or synchronizing limb movements with production of syllables, may enhance the skill of rhythmic organization, and have a positive effect on the development of language (Corriveau and Goswami, 2009; Goswami, 2010).

Music therapy is a developing clinical field which explicitly utilizes musical rhythmic exercises within speech-language interventions. Research has demonstrated that rhythmic auditory stimulation arouses the motor system, prompts it into a state of readiness, and facilitates more efficient, coordinated muscle movements (Thaut et al., 2001). Music-based speech protocols have been used for the rehabilitation of patients with non-fluent aphasia (Kim and Tomaino, 2008). Speech timing cues have been provided as external rate control to enhance verbal intelligibility in dysarthric speakers (Pilon et al., 1998). For both speech and singing, the careful manipulation of rhythm to match and enhance the patients' expectancy was an important factor in achieving improved word retrieval, prosody, and articulation (Wan et al., 2010). In an intervention developed to facilitate speech output in non-verbal children with autism, the therapist introduces the target words or phrases by intoning (singing) the words on two pitches, while simultaneously tapping the drums, to facilitate bimanual sound-motor mapping. The child is led from listening, to unison production, to partially-supported production and eventually to repetition (Wan et al., 2011). Finally, an effective treatment program for pre-school children who stutter has been based on learning to produce speech with the isochronous timing of syllables (Trajkovski et al., 2011).

Musical rhythmicity is also used in the area of language teaching. For example, a method called 'Jazz Chants' has been developed to teach spoken American English to foreign students (Graham and Rosenthal, 1993). In exercises (chants) each stressed word is pronounced with an extra emphasis (often together with physical activities such as clapping or jumping) and with an equal time interval (i.e., isochronism). Chants are normally performed with music that progresses regularly in 4/4 time (based on traditional Jazz progressions) where each beat corresponds to each stressed word. Exercises are based on sets of two bars (i.e., motive), which consists of eight beats in 4/4 time, or their multiples. So-called 'null stressed' words are inserted in a chant (e.g., "Black, yellow, brown. NULL. Jack fell down. NULL"), for which words are not actually pronounced but can be expressed with physical activities such as a clap (Nagata et al., 2011).

In summary, considering the converging reports from the areas of clinical linguistic, music therapy and language learning, we assume that musical rhythmicity is an appropriate framework for developing the sensitivity to timing, and can be utilized for practicing timing skills in the context of speech production. Therefore, an exercise in the proposed methodology requires the reproduction of a sequence of syllables arranged in a specific rhythmic pattern. A pattern is defined by a sequence of intervals between subsequent syllables, whose durations are multiples of the shortest interval. This definition is similar to the musical notation, so that intervals in an exercise can be described as various combinations of $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, and whole notes. As with a musical score, the tempo of performance can be scaled by defining an absolute duration for the shortest note, and multiplying the duration of all other notes accordingly.

This task is somewhat similar to a form of jazz-singing known as 'scat-singing' (see 5.1 for an example). In vocal jazz, scat singing is vocal improvisation with nonsense words and syllables or without words at all. Scat singing gives singers the ability to sing improvised melodies and rhythms, to create the equivalent of an instrumental solo using their voice.

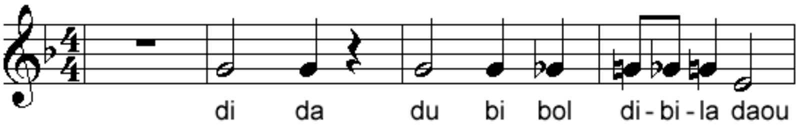



Figure 5.1: A riff from Louis Armstrong's 1926 Heebie Jeebies.

Existing systems for the training of timing skills

An application for training rhythmic skills in the context of musical performance has been developed and tested in Sadakata et al. (2008). The authors describe a real-time visual feedback system, which represents changes in loudness and timing of drumming gestures as parameters of an abstract visual image, transforming in size and shape. Among computer games for speech therapy, some include elements of rhythm training (Vicsi et al., 2001). However, most often these games only target isochronous rhythms, and do not focus on training accurate timing in speech through repetition of various rhythmic patterns.

Recently, a new genre of commercial video games has emerged, in which timing skills are explicitly targeted. In these games, players must press buttons with precise timing, following a specific sequence dictated by a visually presented score. The game evaluates the precision of input gestures in relation to the score, and provides feedback on the performance. Two eminent examples are 'Dance Dance Revolution', where players stand on a 'dance platform' and hit colored arrows with their feet in reaction to visual cues, and 'Guitar Hero', where players use a guitar-shaped game controller to match scrolling on-screen notes to colored buttons on the controller in time to the music (see Figure ??).

Commercial rhythm video games include many of the elements for forming an ideal platform for training timing accuracy with children, including a systematic variation in rhythmic patterning, a coherent exercise and feedback presentation scheme, as well as elaborate visual environments for boosting motivation. However, although these games often feature novel input devices, such as special joysticks, gamepads, and touch sensors, no *voice-based* games have so far been conceived for utilization in the speech training domain. Therefore, we aim to address this gap, by considering how the principles of rhythm games design can be combined with audio input and acoustic analysis, in order to form a basis for a computer system for training timing skills in speech.

5.2 A computer game for training timing in speech

In order to experiment with the proposed methodology, a game for practicing timing skills in speech – *Logoped Ski*, has been implemented. The theme of the game is a downhill slalom ski competition (see Figure 5.2). The goal of the player is to produce a sequence of syllables with precise timing, so that each syllable production corresponds with taking a well-timed turn at a flag. The player uses speech input to control the movement of a slalom skier, so that each uttered syllable causes the skier to change direction on his way down the piste. Rhythmic patterning is realized by positioning the flags in such a manner that their relative distances faithfully represent the rhythmic pattern to be practiced – the spatial distance between flag *i* and flag *i+1* corresponds to the duration of interval *i* in the exercise. Such spatial correspondence provides the basis for predictive feedforward control, which has been shown to support the motor initiation of rhythmic productions (Zatorre et al., 2007).

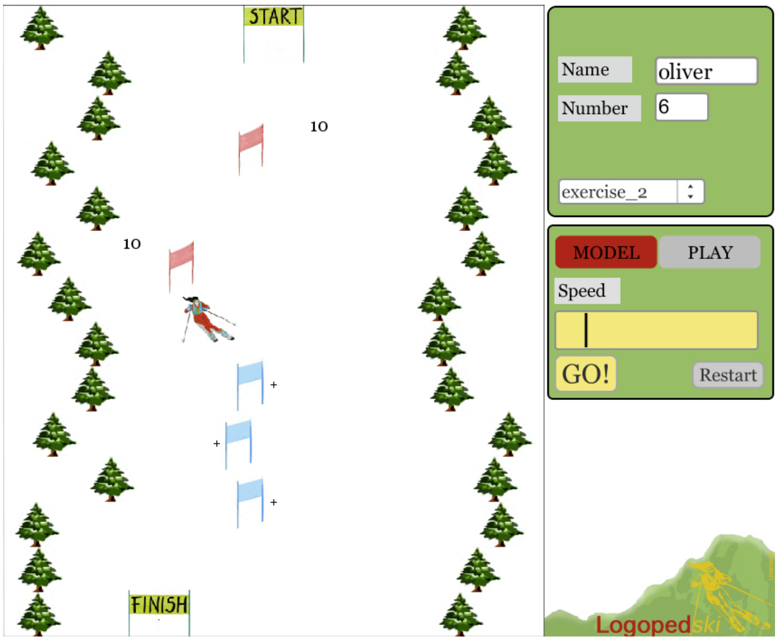

Name oliver
Number 6
exercise_2

Figure 5.2: A screenshot from the game.

For the sake of visual clarity, the exact points of expected syllables are marked with a cross, and the flags change colors (from blue to red) once they are passed by the skier.[1] The game follows the *call & response* format, in which the correct rhythmic pattern is first demonstrated, and the player is required to repeat

---

[1] Depending on feedback mode, flags change color upon an accurate interval production.

the pattern (Blaine and Perkis, 2000). A clinician selects an exercise from a menu of available exercises, and sets it in a *MODEL* or *PLAY* mode. First, an exercise is modeled by means of animating the skier to perform an optimal slalom ride, synchronized with an audio playback of that exercise.[2] Modeling the exercise can be repeated several times, until the player is comfortable with it. A clinician provides instructions as to which speech targets should be uttered in the exercise, and prompts the player to perform the exercise by setting the game in *PLAY* mode. The skier then sets off towards the first flag, but the timely taking of turns now relies on the acoustic input from the player.

Evaluation and feedback

In order to evaluate the timing accuracy produced by the player in relation to a given exercise, vowel onsets (corresponding to syllable nuclei) are detected and the durations of inter-onset-intervals (IOI) are calculated. The duration of the detected IOI is compared with the expected duration for the corresponding interval in the exercise, and their absolute difference is obtained. The timing accuracy is then expressed as the percentage of deviation from the expected interval duration, and is calculated as follows:

$$Accuracy = 100 - \left(\frac{|\text{produced IOI duration} - \text{expected IOI duration}|}{\text{expected IOI duration}} \times 100\right)$$

The percentage of timing accuracy is scaled to a 0 to 10 range, and displayed as the score for that interval, next to the corresponding flag. A score of 10 indicates a near-perfect timing performance (90% to 100% accurate). The average of all interval accuracy scores produces an overall score for the exercise, which is displayed next to the FINISH flag upon completion of the exercise. Displaying the scores for timing accuracy is provided not only for knowledge of result, which is an important ingredient in motor learning, but serves to motivate players to achieve higher scores in subsequent exercises.

Detecting the timing of syllables

In speech rhythm literature it has been well established that listeners base their perception of the rhythmic structure of speech on the so-called perceptual centers (p-centers) (Morton et al., 1976; Marcus, 1981). The best approximation to p-center location has been found to be the vowel onset, which also coincides with the point of maximal variation in the speech signal (Janker, 1996). In the case of our exercises, the task of detecting the timing of produced syllables is

---

[2] Played back with MIDI notes – an industry specification for encoding, synchronizing, and transmitting the musical performance and control data of electronic musical instruments.

expressed as detecting (in real-time) the onset of the vowel which corresponds to the nucleus of each syllable.

A widely used detection function is based on the assumption that, within the task of non-word sequence production, acoustic energy peaks correspond with syllable nuclei (Kochanski and Orphanidou, 2008). Wu et al. (1997) has demonstrated an accurate detection of vowel onsets in about 85-90% of the cases, using relatively simple signal processing techniques. de Jong and Wempe (2007) have developed a PRAAT script which automatically detects syllable nuclei using intensity and voicedness. However the script operates off-line, taking advantage of signal averages across the whole utterance, and therefore may not be suitable for real-time processing.

In our implementation, the function aims at detecting the rising of intensity above a certain threshold (Collins, 2005).[3] Upon a threshold crossing, the function assumes an inter-syllable energy break of at least 250 *ms*, so that no additional onset triggers are issued within this time period. This mechanism provides a low-pass filter for threshold crossings, and helps to prevent spurious onset triggers. The interface of the computer game includes manual settings for the noise-floor and the onset threshold levels.[4] The task of the clinician is to set the appropriate levels, in accordance with the acoustic environment where training takes place, and the voice of a specific child. The interface includes a visual indicator for assisting clinicians with correct threshold setting.

Adaptability

When a player produces an inaccurate interval, the game dynamically adapts the motion of the skier in order to sustain the synchrony between the next expected interval duration and the visual representation of that interval. For example, if a syllable is produced later than expected, the trajectory of the skier towards the next flag will be adjusted, so that the skier will arrive at the next flag on time for the next expected interval. This way, the player is not 'punished' for imprecise productions, and can immediately synchronize back to the correct rhythmic pattern after going 'out of sync'. Furthermore, a clinician can set the appropriate speed for a specific child, so that both the rhythmic pattern and the motion of the skier are scaled accordingly.

---

[3] Treating intensity on a logarithmic scale provides a better approximation for the perceptual quality of changes in voice intensity.

[4] A mechanism for canceling the ambient noise present in the room. The input below the noise-floor level is ignored, which can help to avoid irrelevant fluctuations of the signal.

5.3 Usability study

Participants

A usability study has been conducted in the School for the Deaf and Hard of Hearing in Ljubljana, Slovenia. Speech clinicians of the school provide speech therapy to the hard of hearing children, as well as to ambulant patients with a wide variety of speech and language disorders. Six clinicians and 26 children took part in the usability study.[5] Clinicians were asked to select a group of children with MSD or related speech disorders, who would, according to the clinician's experience, benefit from the training of timing skills in speech. The selection resulted in a group of 26 children, with 7 girls and 19 boys, aged 4-10 (M = 6;6 years; SD = 1;6). Among the group of participants, 7 children were diagnosed with stuttering, 7 with specific language impairment (SLI), 2 were hard of hearing, 2 with suspected childhood apraxia of speech (CAS), and the rest with motor speech disorders of unknown origin.

Procedure

During the study period of 6 weeks, clinicians used the computer game for practicing timing skills in speech with their patients. Each child completed 3 practice sessions with the computer game on separate occasions. Each practice session lasted for 15 minutes, in which clinicians presented exercises to the child, and provided verbal feedback on the performance. Each exercise consisted of a sequence of $\frac{1}{8}$, $\frac{1}{4}$, or $\frac{1}{2}$ note intervals, adding up to one whole note (8 ð $\frac{1}{8}$ intervals). The game contained 10 different exercises, covering a variety of rhythmical patterns. Clinicians were instructed to present the exercises in ascending order of complexity, as well as choose appropriate speech targets to be produced, in accordance with the child's abilities.

Evaluation

At the end of the study period, clinicians were presented with a questionnaire which included 10 items aiming at evaluating various usability aspects of conducting timing skills training with the computer game. The questionnaire was adapted from Öster (2006), where it was used for evaluating the usability of several computer games for speech language therapy (questions are listen in 5.3. For each item, clinicians chose the most appropriate answer on a 4-point Likert-type scale, and elaborated their comments on that question, where appropriate.[6]

---

[5] Initially, eight clinicians engaged in the study, but two of them dropped out.

[6] For items 9 & 10, dealing with appropriate age groups and clinical populations, multiple answer selection was possible.

Results & discussion

We present the results for the usability questionnaire and refer to observations collected from clinicians in order to elaborate on each item.

(1) How much previous experience with computers is required for using the computer game?

A lot	Some	Not much	None
0%	33.3%	50%	16.7%

Concerning previous computer knowledge required for using the game, 50% of the clinicians found that no previous knowledge is needed, while 33.3% found that some knowledge is warranted. Interestingly, 2 out of 8 clinicians stopped shortly after engaging in the study, as a result of being uncomfortable with a computer-based system. These were the eldest clinicians in the team, and their lack of comfort with the computer is an important factor for consideration. When computer-based methods are introduced into clinical settings, it might be worthwhile dedicating special attention to the group of clinicians who are less confident with technology, by providing additional instruction and guidance.

(2) How easy was it (for the clinician) to work with the computer game?

Difficult	Not so easy	Easy	Very easy
0%	16.7%	50%	33.3%

As for the game being easy to use by clinicians as a training tool, 50% judged the game to be easy, and 16.7% as very easy. However, some clinicians reported difficulties with manual settings for the noise-floor level and the onset detection threshold. This difficulty is understandable, as these settings involve rather technical concepts, which may not be intuitive to most speech clinicians. Future design should be concerned with mechanisms for automatic threshold level settings, in order to minimize the demands put on clinicians for this task.

(3) How well did children understand the feedback provided by the game?

Badly	Not so well	Well	Very well
0%	50%	50%	0%

The question whether children understood the visual feedback provided by the game remains inconclusive, with 50% positive replies. Clinicians report that while children understood that the skier turns in reaction to them producing speech, as well as the changing colors of flags, some did not understand the meaning of the numbers presented on the screen. For younger children, presenting the score in a numerical form might be too complex, and a more simplified visualization scheme is needed. One solution would be including an option for a binary feedback on performance, indicating whether a produced interval was accurate or not.

(4) In how far were children motivated to practice with the computer game?

Unmotivated	Not so motivated	Motivated	Very motivated
0%	16.7%	66.7%	16.7%

Concerning the motivation to practice with the game, 66.7% of clinicians report that children were motivated to practice in this way, 16.7% that they were very motivated, and another 16.7% replied with 'not so motivated'. A clear trend reported by the clinicians was that children were very motivated to practice during the first few exercises, but their enthusiasm faded after a few trials. This is not surprising considering the rather static nature of the current game.

(5) How was the interaction between the clinician and children during training?

Bad	Not so good	Good	Very good
0%	16.7%	33.3%	50%

The interaction between the child and the clinician during practice has been judged by 50% of clinicians as very good, and by 33.3% as good. Clinicians reported that children were willing to cooperate with them in the process of playing the game, and were ready to listen to instructions and guidelines while playing. In that regard, the computer game seems to have been successful in activating children and giving them a sense of control, while in the same time, keeping attention to the clinician intact. One clinician reported that some children were more occupied with the game than listening to her instructions.

(6) How effective is the computer game for practicing timing skills in speech?

Uneffective	Not so effective	Effective	Very effective
0%	50%	50%	0%

For the question whether the computer game is an effective way to practice timing skills in speech, only 50% of clinicians replied positively. Others have remarked that the game contained too many different rhythms, with little difference between them. The lack of graphical variation has also been reported as diminishing the effectiveness of training, due to its influence on the motivation to practice.

(7) How reliable was the game, in terms of providing correct and consistent feedback?

Not reliable	Not so reliable	Reliable	Very reliable
0%	33.3%	50%	16.7%

The reliability of the computer game in terms of providing consistent feedback was evaluated by 16.7% of the clinicians as very reliable, and by 50% as reliable. However, the fact that 33.3% of clinicians rated the feedback as 'not so reliable' points at a significant shortcoming of the system. Clinicians reported that consistent feedback depended on threshold settings, as well as on the speech targets being produced. We address patterns of system's errors, and propose appropriate solutions in 5.3.

(8) How well does the computer game target timing skills in speech?

Badly	Not so well	Well	Very well
0%	0%	83.3%	16.7%

As for how well the exercises target timing skills in speech per se, 83.3% considered the computer game to tap well into these skills. Clinicians remarked that many children with speech and language disorders can benefit from a regular training of timing skills, and the computer game can be a valuable addition to their practice, as long as reliable feedback can be provided.

(9) For which age group is training with the computer game most appropriate?

6-8 years	4-6 years
37.5%	62.5%

As for the age group for whom the computer game is an appropriate tool for training, 62.5% of the clinicians referred to 4-6 years old children, while 37.5% to 6-8 year olds. This observation is in agreement with previous studies on usability of computer games for speech therapy (Sandlund et al., 2009b), suggesting that younger children could best appreciate this form of training. In order to better address the group of older children, more sophisticated game scenarios must be employed.

(10) Which clinical population can benefit from training with the game?

MSD	Stuttering	SLI	Hearing impaired
66.6%	16.6%	50%	50%

Concerning the clinical populations for whom the training of timing skills in speech may be beneficial, 66.6% of the clinicians referred to motor speech disorders, 50% to children with hearing impairments, and 50% to children with specific language impairment. Another 16% believed the training is beneficial for children who stutter. The fact that clinicians considered the training method relevant for children with hearing impairments is encouraging, as it indicates a possible wider application of the method.

Creating custom exercises

From the results of the usability study, it became apparent that in order to insure that appropriate rhythmic patterns are used for training, a modular way of creating individual exercises is needed. Therefore, we have built an interface which allows the compilation of custom rhythmic sequences. The interface, illustrated in Figure 5.3, is based on a graphical representation of interval durations associated with each syllable in the exercise. First, a clinician selects the number of different syllables comprising the exercise. Subsequently, a corresponding number of interval duration boxes appears, and the clinician then selects a speech target for each box. Currently, the selection is made from a pre-compiled list of frequent syllables. However, the selection can also be linked to a more elaborate procedure, such as described in detail in chapter 4.

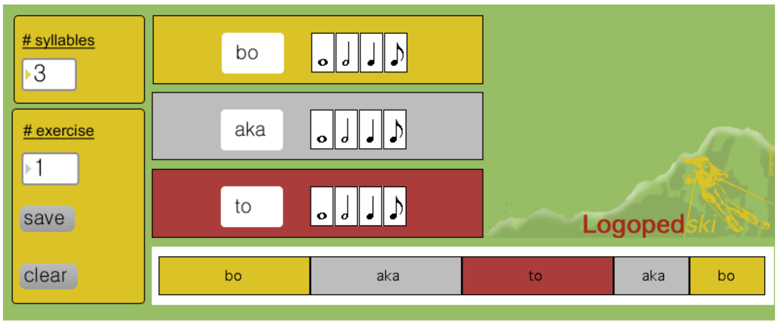


Figure 5.3: An interface allowing the compilation of custom rhythmic sequences

Next, the task of the clinician is to compile an exercise by concatenating interval durations associated with any of the selected speech targets, by pressing the corresponding duration symbols. In this illustration, each exercise is equivalent to two whole notes (or 16 ð$\frac{1}{8}$ notes). As the clinician selects a sequence of durations for a certain exercise, a schematic representation of the exercise is updated. When the exercise is compiled, it can be saved to a file, and the process repeated as long as necessary. The created exercises are automatically added to the menu of exercises in the computer game for training.

Vowel onset detection errors

Following the observations of clinicians suggesting that the game was not always reliable in terms of consistent feedback, we have analyzed the most typical errors which occurred in the process of vowel onset detection. This examination has revealed a number of acoustic event types, which compromise the accuracy of feedback provided in the game. We shortly discuss these types, and illustrate them with waveform examples taken from recordings made during the study.

(1) In a case when some of the syllables in a sequence are produced with too low amplitude, the system will detect no vowel onsets (for example, see the last three syllables in the left illustration of Figure 5.4). Too low a signal can result either from an inappropriate production of the player, or from an unfit distance of the player to the microphone.

(2) When the player produces speech too close to the microphone, the input signal is too high, causing clipping, for which no onset detection is possible (for example, see the flat 'chunks' in the right illustration of Figure 5.4).

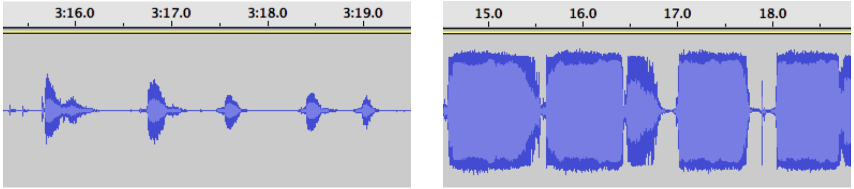


Figure 5.4: A waveform of a syllable sequence production, showing syllables produced with too little amplitude (left), and the clipped signal resulting from speaking too close to the microphone (right).

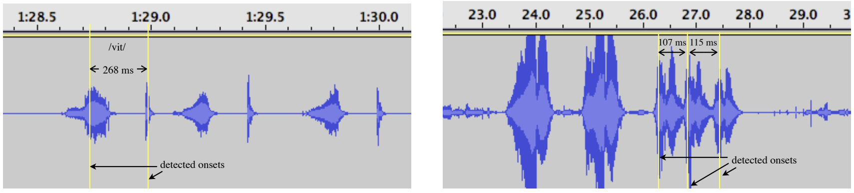


Figure 5.5: A waveform showing the double energy peak and the low energy interval for the syllable /vit/, causing a false double detection (left), and three syllables corresponding to $\frac{1}{8}$ note durations produced in succession(right).

(3) When the speech target involves a CVC syllable, the acoustic realization may produce a false double detection since the signal may contain two energy peaks, corresponding to each produced consonant. If the duration of the low energy interval between the two peaks is longer than the threshold of duration assumed for inter-syllable energy breaks (in our case 250 ms), the function will produce a false detection upon the second energy peak of the syllable (for example, see the double energy peaks for each production of the syllable /vit/ in the left illustration of Figure 5.5).

(4) When syllables are produced in succession (for example, when short intervals or a fast tempo are implicated in an exercise), the acoustic signal may not contain the breaks in energy necessary for a correct detection. If the duration of the interval between the detected peaks is shorter than the threshold assumed inter-syllable energy breaks, produced syllables will be undetected (see the right illustration of Figure 5.5).

In order to deal with errors arising from an inappropriate distance of the player to the microphone, the system could continuously monitor the signal input level, and alert once it is too low or too high, prompting the clinician to adjust the distance to the microphone. For dealing with problems arising from certain syllable structures (such as CVC), the detection function could rely on knowledge of the expected exercise. If the target syllable and the expected rhythmic pattern are known in advance of the exercise, the function could assign probabilities to onset events, considering the real-time pattern in relation to the expected pattern. For example, if a CVC syllable is expected in an exercise, and two proximate energy peaks occur which do not match the expected rhythmic pattern of the exercise, the function could prevent the false onset detection upon the second energy peak. Finally, more sophisticated signal processing can be applied to refine the detection mechanism. Currently, only signal intensity plays a role in vowel onset detection. However, the detection algorithm could take into account the spectral composition of the signal, and discard the peaks which are not voiced, as have been proposed by de Jong and Wempe (2007).

5.4 Conclusions

The purpose of this chapter was to address questions regarding the development of a computer game for the training of timing skills in speech of children with speech disorders. Having outlined previous work on training speech motor skills in the framework of musical rhythmic tasks, we have proposed a learning task which involves the production of syllable sequences arranged in a certain musical rhythmic pattern. In order to test the feasibility of providing an engaging practicing environment for children, we have built a computer game which implements the proposed exercises through visual guidance and feedback. The outcome of the usability study with the game has equipped us with a deeper understanding of the challenges and difficulties involved in developing this method of training, but also of its potentials. In spite of outlining a number of critiques on the usability of the current computer game, clinicians have expressed a positive attitude towards this line of development.

The main limitations at this point seem to be the lack of robustness in detecting syllable nuclei, and the lack of sufficient graphical variation during game-play. Addressing these shortcomings, we have proposed new lines of development on the way to an improved design of a training system for timing skills in speech. The development of custom games for training and therapy is time consuming and expensive. Moreover, it is most likely that such games will not stand up to the quality and entertainment standards of commercial games developed by large computer game producers. Therefore, future attempts of developing timing skills games should examine the integration of real-time speech input and analysis capabilities within existing video game platforms.

As mentioned in 5.1, modern rhythm games offer excellent mechanisms for training timing accuracy with children, including versatile and variable graphical environments. The missing element is the integration of audio input, and audio signal processing capabilities into an existing rhythm game platform. With open-source software architectures, such integration is feasible, and can offer considerable advantages. Relying on a gaming platform which has been tested and approved by a community of players, the integrated speech training system would more likely to provide a durable training opportunity.

During the project, questions concerning the reliance on musical rhythms as training methodology have been raised. In particular, it is worthwhile to consider the practicing of timing patterns which model the natural speech timing of certain phrases, with the idea of learning speech timing directly. An example of a training method where the timing pattern of a phrase (in terms of the sequence of syllable onsets) is visually presented is the Metrical Pacing Technique designed for the treatment of apraxia of speech, described in Brendel and Ziegler (2008). However, even if a corpus of phrase timing models is obtained, it is not clear how these models can be used to provide feedback on the quality of speech timing production – children may produce a phrase in various ways, which are perceptually valid in terms of natural speech timing. An automatic evaluation of whether a 'deviation' from the timing pattern of the model falls within or outside of an accepted speech timing target is difficult, at best. Moreover, it is not possible to scale the model pattern in terms of rate/tempo because of the non-linear characteristics of natural speech timing.

Approaching the timing skill training methodology from the other end – the practicing of purely musical rhythmicity outside the speech domain (for example, through learning to play a musical instrument) – has not, so far, been acknowledged by speech clinicians as a prominent method in speech-language therapy. From this perspective, the methodology proposed in this chapter can be seen as a compromise – it aims at developing the sensitivity to timing in a broad sense, while remaining within the domain of speech production. By placing specific demands on the person who practices to regulate the timing of speech production, the ability to achieve higher degree of precision in speech movements is refined.

The clinical efficacy of training timing skills in speech with a computer game can only be evaluated when the proposed training method is sufficiently developed, in terms of feedback reliability and gameplay usability issues addressed in this chapter. Future efforts should tackle the technical challenges with robust vowel onset detection, as well as finding creative solutions for housing the training in an engaging game environment. We believe that this line of development is worthwhile pursuing further.

CHAPTER 6

Development of adaptive Altered Auditory Feedback procedures for adults who stutter

> He who does not understand your silence
> will probably not understand your words.
>
> *Elbert Hubbard, 1911*

6.1 Introduction

In this chapter, an improvement to the current methodologies of altered auditory feedback (AAF) for people who stutter is proposed. An attempt is made to examine *adaptive* feedback procedures, which are based on real-time analysis of the speech signal, and the dynamic activation of auditory feedback. Specifically, we propose to deal with the difficulty of initiating speech at phrase onsets by *selectively* providing a feedback signal only in the silent intervals between phrases. In order to evaluate the proposed AAF procedure, adults who stutter produced spontaneous speech under experimental feedback conditions.

Adaptive auditory feedback procedures

Altered auditory feedback (AAF) is known to reduce stuttering by 50-80% in some people who stutter (Lincoln et al., 2006). AAF can be defined as a manipulation of one's speech signal, in which the altered signal is fed back to the speaker 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 (for background information on the theoretical and methodological aspects of AAF procedures, please see Chapter 2.4).

One obvious limitation of AAF procedures is that they provide no signal at the moment when phrases are initiated. This is due to the fact that current AAF methods operate on the basis of a static feedback function. In other words, the feedback signal is given continuously, at the same output level, with the same delay/shift settings, regardless of any property of the audio signal produced by the speaker. One way to address these limitations is to design *adaptive* rather than static feedback functions, with the idea of *selectively* targeting those regions of speech which are at higher risk of being dysfluent (Howell, 2004). In his extensive overview of altered auditory feedback methodology, Howell suggests that AAF signals should be targeted only on, or around problematic locations. He further proposes that procedures which restrict the exposure to AAF signals while maintaining higher levels of fluency may be advantageous.

In general, an adaptive feedback procedure operates by analysing some property of the speech signal in real-time, and dynamically modifying the output audio signal in relation to that property. The analysis process thus aims to detect regions in the speech stream which correspond to certain 'states' of the speaker. In the case of stuttering, problematic regions include phrase onsets, regions in which a block is suspected, or regions in which speech rate is too fast (Wingate and Howell, 2002). Once a target region is detected, an external stimulus signal is then activated to assist speakers with maintaining their flow of speech. Once the system detects that the speaker shifted to another 'state', which is not considered vulnerable for dysfluencies, the external signal can be attenuated, or withdrawn altogether. It is important to note that the described framework does not involve attempting to detect *dysfluent events* per se, but rather aims at predicting problematic areas in a speech stream.

Targeting speech rate within a scheme of adaptive feedback is not a new idea. As a classic example, Howell mentions the 'Hector aid', a device pioneered by British engineers more than 30 years ago. The device measured speech rate using an audio input, and a vibrator switched on if speech rate was outside an acceptable range, in order to signal the speaker to slow down. The rationale for targeting speech rate modulation comes from evidence that dysfluencies are more likely to occur in the fast-rate areas of speech (Smith and Kleinow, 2000). Although a detailed report about the 'Hector aid' is unavailable, the idea behind this device is a good exemplification of the approach advocated in this chapter.

Providing an auditory bridge

The adaptive feedback procedure explored in this study addresses the limitation of current AAF procedures with regard to phrase initiation. A variety of experiments demonstrated the effect of priming fluent speech initiation by means of external auditory cues, with the help of metronome beats, AAF signals, choral speech, and other auditory cues (Azrin et al., 1968; Kuniszyk-Jó kowiak et al., 1997; Howell and El-Yaniv, 1987; Natke et al., 2001; Alm, 2004). Therefore, we assume that an auditory cue present at the moment of phrase initiation has the potential of *priming* a speaker who stutters to initiate a phrase fluently. Speaking of a 'phrase', we refer to a continuous interval of speech, bounded by a silent pause with a duration of 200 milliseconds, or more (Fant et al., 2003). This definition may only partially correspond to the one of the intonation phrase, since the relation between intonation phrase boundaries and silent pauses is not entirely agreed upon (Bulyko and Ostendorf, 2001; Yoon et al., 2007; Wagner, 2010). For example, a study by Yoon et al. (2007) revealed that less than 50% of phrase boundaries in English are followed by a silent interval, suggesting that silent pauses alone may not be systematically used as a cue to phrase boundary presence, and that listeners rely on additional acoustic cues, such as pre-boundary lengthening (Lin and Fon, 2009).

In order to activate an audio cue just before the onset of an upcoming phrase, a temporal reference point is needed, which can be detected in the speech stream, allowing to predict that a phrase onset is about to take place. The detection of a silence interval following the termination of the previous phrase is a way to anticipate the timing of the subsequent phrase initiation. Therefore, in our adaptive procedure, a signal is activated only during the silent intervals between phrases in order to prime the initiation of subsequent phrases. We propose that this procedure would provide a kind of an 'auditory bridge' between phrases, assisting speakers to smoothly transit from one phrase to the next. In order to create a feedback procedure with a minimal 'auditory intervention', the audio signal is withdrawn once an utterance has been initiated. However, an upper limit for the duration of the auditory cue must be determined, such that it approximates the longest typically occurring silent pause in spontaneous speech. In a large-scale study, Campione and Véronis (2002) have analyzed the distribution of pause durations in spontaneous speech for English, German, French, Italian and Spanish. Their results revealed the duration distribution pattern is a combination of three categories of pauses: brief, medium and long, with only 8% of pauses being shorter than 200 ms, and 2.8% of pauses being longer than 2000 ms. The class of long pauses has been shown to be centered around 1500 ms. Therefore, in order to account for the longest expected pauses, the 'auditory bridge' will be provided until a new phrase is initiated, lasting up to 1500 ms.

We hypothesise that the adaptive feedback procedure described above will enhance the initiation of phrases in spontaneous speech to a larger degree than a standard DAF procedure. Specifically, we predict that the proportion of dysfluent phrase onsets during speech with the adaptive feedback procedure will be decreased in relation to speaking without auditory feedback, and that the decrease will be greater than the one achieved with DAF. Based on previous findings that in spontaneous speech, stuttering is more likely to occur at the onset of utterances (Quarrington, 1965; Jayaram, 1984; Koopmans et al., 1991; Saltuklaroglu et al., 2009), we predict that a decreased proportion of dysfluent phrase onsets will lead to an overall improvement in fluency, comparable with an improvement achieved with a standard DAF procedure. Furthermore, we predict that the adaptive feedback procedure will be perceived as more comfortable than a standard DAF procedure, since it involves a less intrusive presentation of external audio signals.

6.2 Methods

Participants

The study included 12 adults who stutter (M = 32;3 years, S.D. = 11;5). Eleven were male and one was female. Participants were recruited by announcements placed in clinics for stuttering therapy throughout The Netherlands. All participants reported stuttering onset in childhood, none had severe hearing loss, and all were native Dutch speakers. Ten participants had no previous experience with AAF, and two had a short experience over 3 years prior to the study.

Procedure

The experiments took place in a quite room. Participants were seated approximately 1.5 meters in front of the investigator and a video camera. Participants were asked to produce a monologue under four auditory feedback conditions (Figure 6.1 provides a schematic illustration of the experimental conditions):

(a) *no-feedback condition* – participants were fitted with headphones, but did not receive any auditory signal while speaking, and could hear their own voice.

(b) *a standard DAF condition* – the audio signal coming into the microphone was fed back to the participants with a delay of 100 ms. This delay duration was chosen as an average, standard setting used in AAF research, as well as in AAF devices (Lincoln et al., 2006).

(c) *an adaptive feedback condition* – an audio signal was delivered only during the detected silent intervals. Upon silence detection, an audio buffer containing the most recent 1500 ms is played back until a new phrase onset is detected, with maximum duration of 1500 ms. A speech signal was chosen as the external auditory cue in order to remain in the same auditory domain as the standard DAF procedure, so that comparisons could be made. The choice of the buffer duration is outlined in refbridge.

(d) *a combination of conditions (b) and (c)* – since part of dysfluencies may occur at mid-phrase positions as well (Saltuklaroglu et al., 2009), the adaptive feedback condition (c) was supplemented with an additional delayed signal (100 *ms*) during detected speech episodes. During detected speech episodes, the audio signal was fed back to the participants with a delay of 100 *ms*. Upon a detected silent segment, an audio buffer containing the most recent 1500 ms is played back until a new phrase onset is detected, with maximum duration of 1500 ms.

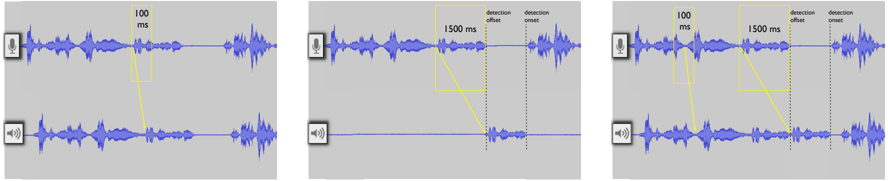


Figure 6.1: schematic representations of conditions (b) - left, (c) - middle, and (d) - right. The top signal track represents the speech output of the participant, while the bottom track represents the audio signal received by the participant through headphones.

The order of presenting the various conditions was randomized across participants. For each condition, the investigator presented a new topic for discussion, asking the participant for his/her opinion on the topic, and eventually signaled the end of discussion once enough speech material was collected (roughly 3 minutes per condition). The same set of questions and topics was used for all participants. A monologue task was chosen since it is a variant of spontaneous speech, which realistically approximates everyday communication situations. Participants were asked to speak in their normal, natural manner, with a normal speech rate, and as far as possible, to avoid using any fluency inducing techniques. In order to prevent the investigator's voice being fed back to the participant, the investigator spoke only when it was necessary to re-engage the participant in the monologue.

After each condition, the investigator asked the participant to reflect on the experience with the current auditory feedback procedure, as well as to rate the relative level of comfort with the audio signal. A 10-point visual analogue scale was presented on a computer screen, and the participant indicated a score on that scale. This break served to avoid possible carry-over effects between conditions, but also provided valuable insights into the way participants experienced their speech in the given feedback mode. At the end of the experiment, participants were asked to give their general impressions and personal comments on the methods and ideas explored in this study.

Apparatus

All experiment sessions were recorded through a video camera (Panasonic S150). The audio was digitized in real-time by the ZOOM4H microphone and an external audio card. The DAF function was implemented within the MAX/MSP environment (www.cycling74.com). In order to implement the silence/onset detection required for the adaptive feedback procedure, the following real-time signal processing scheme was employed:

1. The raw linear amplitude was converted to amplitude in decibels, so that 0dB = 1 (full amplitude). [1] Treating amplitude on a logarithmic scale provides a better approximation for the perceptual magnitude of changes in intensity.

2. Input level in decibels was then expanded by considering a noise-floor threshold (set manually by the investigator to offset the ambient noise inside the experiment room). The expanded level was calculated as a percentage of the above-threshold range.[2]

---

[1] Amplitude expressed in decibels, with: $dB = 20 \times \log 10(A)$

[2] Noise floor canceling is an important step in audio processing, as it can help avoid irrelevant fluctuations of the signal, and make the detection function more robust.

3. Next, the expanded signal was compared against a silence/onset threshold, which was set at 15% of the full dB range (and manually adjusted if necessary). Upon a threshold crossing in either direction (from ‘speech’ to ‘silence’, or the other way around) the system waited for 200 milliseconds before issuing a silence/onset trigger[3]. If, however, a crossing occurs in the opposite direction during these 200 milliseconds, the timer is reset, and no trigger is issued. This mechanism provides a low-pass filter for the detection function and prevents spurious silence/onset triggers.

Data analysis

For each speech sample, the fraction of dysfluent phrase onsets from the total number of phrase onsets was calculated. Phrase onsets were defined as moments of speech initiation after an interval of silence longer than 400 *ms* (Swerts and Geluykens, 1993). A phrase onset was defined as dysfluent if a stuttering-like dysfluency has occurred on, or before the first stressed syllable (Natke et al., 2002). Since the adaptive feedback procedure targets to enhance the initiation of phrases, the proportion of dysfluent phrase onsets reflects how well speakers deal with initiating phrases throughout their speech.

Further, in order to evaluate the overall fluency levels across the experimental conditions, the percentage of discontinuous speech time (PDST) was calculated. In order to arrive at PDST scores, a two-step analysis of the original raw audio samples was performed. First, all segments from the sample are removed which do no belong to the speech of the participant. This includes speech segments of the investigator, pauses longer than 2 seconds, coughing, or any other irrelevant noise. The resulting sample is referred to as ’the cleaned original sample’. Next, all stuttering-like dysfluencies are removed from the cleaned sample. Stuttering-like dysfluencies include part word repetitions, monosyllabic-word repetitions, prolongations, and blocks (the presence of dysfluencies, including phrase-initial blocks were discerned from the video recordings). The result of this procedure is a quasi-fluent speech sample. PDST is then calculated as:

$$PDST = 100\ \eth\ (1 \quad \frac{\text{duration of the quasi-fluent speech sample}}{\text{duration of the cleaned original sample}})$$

PDST scores for 3 participants revealed near zero levels in all conditions, reflecting no stuttering dysfluencies in their speech, and were excluded from statistical analysis.

---

[3] This duration threshold was based on findings that a silence interval (a pause) marking the intonation phrase boundary usually has a duration of 200 milliseconds, or more (Campione and Véronis, 2002; Fant et al., 2003).

6.3 Results

Group results

Statistical analysis (Friedman test) revealed a significant difference between the four experimental conditions for proportions of dysfluent phrase onsets (^{2} = 12.6; d.f. = 3; p = .006). Conditions (b), (c), and (d) all reduced the proportions significantly in relation to the no-feedback condition (Wilcoxon signed ranks test, Z = 2.42; p = .015, Z = 2.549; p = .011, Z = 2.312; p = .021 respectively). There were further no significant differences between conditions (b), (c) and (d) [4].

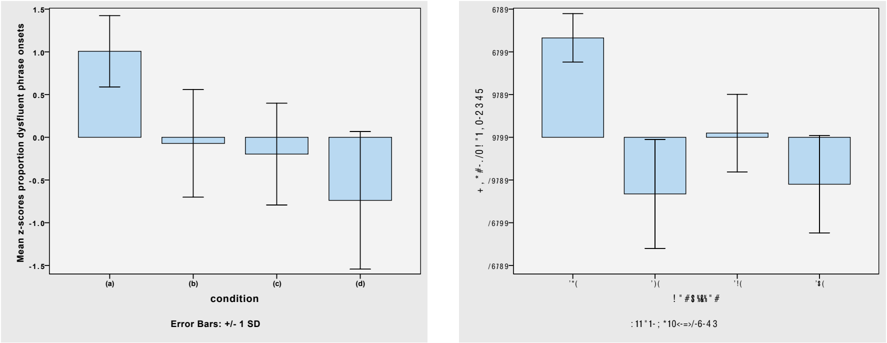


Figure 6.2: Group means for proportions of dysfluent onsets across conditions (left), and for PDST measures across experimental conditions (right), z-scores normalized.

Group results for PDST measures show a similar pattern (Figure 6.2). Also with this variable, there was a significant effect of experimental condition (^{2} = 17.4; d.f. = 3; p = .001), and a significant reduction of PDST scores in conditions (b),(c), and (d) in relation to the no-feedback condition (Wilcoxon signed ranks test, Z = 2.666; p = .008, Z = 2.429; p = .015, Z = 2.666; p = .008 respectively). There were no significant differences between conditions (b), (c) and (d). Comfortability scores revealed no clear preference between conditions (b), (c), and (d).

---

[4] In order to display group means and individual trends in a meaningful way, the raw scores are normalized, so that each individual's trend across experimental conditions can be overlaid in the same numerical space. To achieve this, standard normal deviates (z-scores) were calculated for the raw scores for each individual speaker. The z transformation produces a distribution of scores with a mean of 0, and SD of 1, so that each score represents the distance in SD from the mean. After this transformation, a comparison can be drawn between the dysfluency trend of different speakers.

Individual results

Since group results in experiments with people who stutter show only a part of the picture, we present individual results for proportions of dysfluent onsets and for PDST measures (both as raw scores, as well as after z-scores normalization).

Although most participants show a reduction in conditions (b), (c), & (d) relative to the no-feedback condition (a), there are different trends of dysfluency across these conditions (Figure 6.3). For example, participants 8 and 12 have better reduced their dysfluencies on phrase onsets in condition (c) than in condition (b), and even more so in condition (d). The same time, participants 1 and 11 achieve a strong reduction in condition (c), but show more dysfluent onsets in condition (d). Participants 3 and 5 exhibit less reduction in condition (c) than in both conditions (b) & (d).

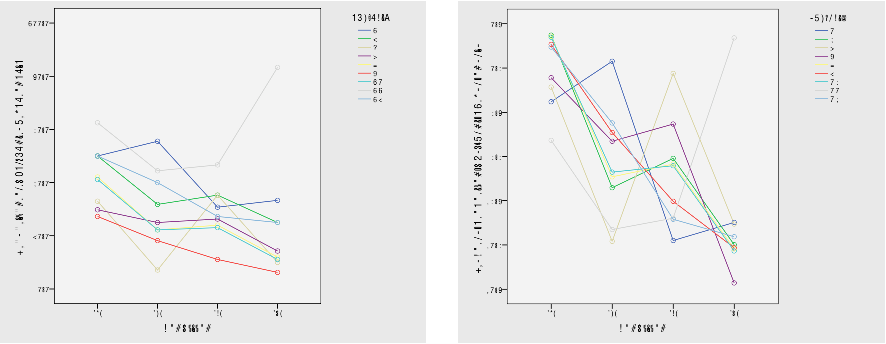


Figure 6.3: Individual results for proportions of dysfluent onset across conditions, presented as raw scores (left) & as z-scores normalized (right)

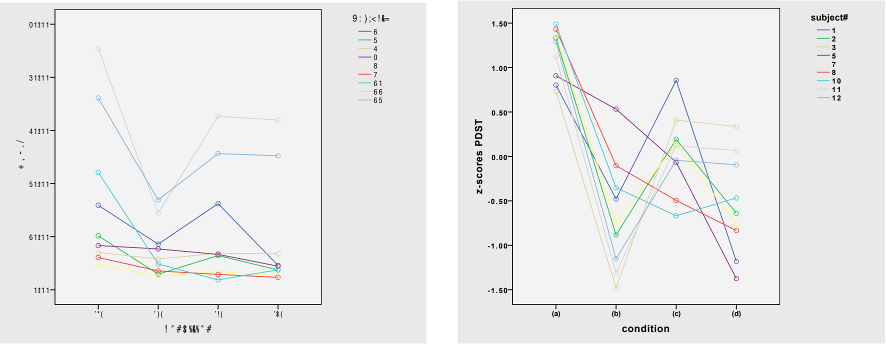


Figure 6.4: Individual PDST results across conditions, presented as raw scores (left) & as z-scores normalized (right).

The normalized individual PDST score in Figure 6.4 reveal that all participants improved their fluency level in conditions (b), (c), & (d). Two trends can be distinguished in the group. Participants 5, 7, and 10 have achieved more fluency in condition (c) than in condition (b), with even better or roughly the same level of fluency in condition (d). All other participants, however, show a pattern in which fluency levels are worse in condition (c) than in condition (b), while the level in condition (d) is either similar or better than in (c).

Sub-group results

An examination of the PDST scores under condition (a) in the left graph in Figure 6.4 reveals two rather distinct groups, with 5 participants having PDST scores lower than 10%, while the other 4 participants having PDST scores of 16% and higher. We term the first group as the 'milder' stuttering severity group, and the second as the 'more severe' group. Next, in order to compare the trends of both groups, we present z-scores normalized results for the two sub-groups.

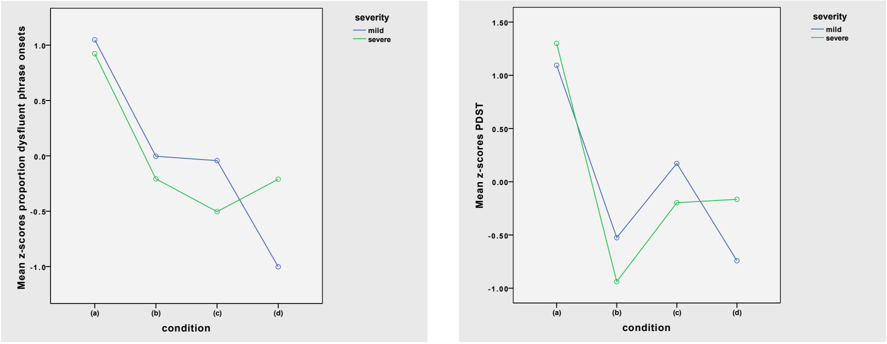


Figure 6.5: Sub-group results for proportions of dysfluent phrase onsets across conditions, z-scores normalized

While in conditions (b) and (c) the more severe group seems to benefit from the feedback signal slightly more than the milder group, in condition (d) this trend is reversed (left graph in Figure 6.5). The milder group shows more reduction in dysfluent phrase onsets in condition (d) than the more severe group. A similar pattern can be seen in the sub-group results for PDST measures (right graph in Figure 6.5). Also here, the group of participants with a more severe dysfluency seem to improve their fluency levels to a greater extent than the other group only in conditions (b) and (c), but not in condition (d).

Participants' perceptions

We report a number of remarks expressed by participants in relation to the experimental conditions:

- Most participants agreed that condition (b), and DAF procedures in general, may be effective in reducing dysfluencies, but they do not help with the initiation of phrases, which remains an important concern.
- One participant remarked that the time when audio is fed back to the speaker in condition (c) can provide a 'resting' interval before initiating the next phrase, and therefore can help with maintaining fluency.
- Participants noted that it is difficult to predict the timing of feedback activation in condition (c), which may compromise its effectiveness to stabilize the speech production process.
- A number of participants noted that condition (d) is effective in helping to initiate phrases. Others, however, regarded the 1500 ms delay as being too slow, and not assisting with fluency since it is not possible to monitor one-self in real-time.

6.4 Discussion

Group trends

The group results for the proportion of dysfluent phrase onsets reveal that the adaptive feedback condition (c) significantly reduced the proportion of dysfluent onsets in relation to the baseline condition. However, the results did not support our hypothesis for a significant advantage of the adaptive feedback condition (c) over the standard DAF condition (b). Judging from Figure 6.2, there is a slight tendency for larger improvement in phrase initiation under condition (d). We interpret the finding that the adaptive feedback condition (c) did not provide significant advantage over the other experimental conditions in relation to the fact that the audio feedback in condition (c) was withdrawn once a phrase onset had been detected. If the audio cue was removed before speakers had the chance to produce an articulatory movement sufficient for resulting in a fluent syllable, the feedback procedure would not offer an efficient support for initiating phrases. If that is the case, then activating the standard DAF signal once phrase onset has been detected, as occurs in condition (d), would provide a continuation for the audio cue, and allow speakers to complete a fluent phrase initiation.

The group pattern for overall fluency results, measured by PDST scores, reveals a similar pattern, by which the adaptive feedback condition (c) significantly reduced PDST scores in relation to the baseline condition, but not significantly better than other experimental conditions. In fact, Figure 6.2 reveals a tendency for conditions (b) and (d) to reduce PDST scores somewhat more than condition (c). These results seem to suggest that while speakers are able to utilize the DAF signal in mid-phrases, the additional audio cue during silences does not further enhance fluency. This 'ceiling' effect could be due to providing too much audio information during speech, when both mid-phrase and silences are filled with audio feedback. It seems that beyond some amount of auditory information, additional signals may be detrimental instead of beneficial. Alternatively, the finding that the additional audio cue during silences did not provide significant advantage may reflect the inadequate nature of the auditory cue used in this study.

Although comfortability scores revealed no clear preference between conditions (b), (c), and (d), participants tended to rate condition (b) as (slightly) more comfortable than condition (c). The reason could lie in the fact that most participants were familiar with a standard DAF procedure (either by experience or with the idea of it). As some participants have remarked, anticipating the nature of the audio signal in condition (b), could be making it more acceptable and less annoying than the less predictable audio signals in other conditions.

Individual variability

Individual data reveal high variability in participants' responses to different feedback conditions, suggesting that different AAF procedures may have differential effects on various types of dysfluent speech. This idea is consistent with the notion that biological subtypes of stuttering may exist, and that different subtypes respond differently to AAF exposure. A similar idea was proposed in Foundas et al. (2004), who reported that DAF had a stronger fluency-inducing effect in the subgroup of stuttering individuals with rightward planum temporale asymmetry (i.e. right planum temporale larger than left planum temporale). In order to account for the high individual variability in response to AAF methods, it has been suggested that individual differences are influenced by the degree of auditory sensitivity (Howell et al., 2006), or by idiosyncratic abnormalities in the organization of auditory and motor related brain areas of people who stutter (Salmelin et al., 1998; Fox et al., 2000; Watkins et al., 2008). Altogether, research findings point towards the existence of multi-causality in the etiology of stuttering, and therefore, of subtypes among individuals who stutter (Yairi, 2007).

An interesting trend was discovered when the group of participants was partitioned into sub-groups according to stuttering severity. Results for both the proportion of dysfluent phrase onsets and PDST scores, as can be seen in Figure 6.5, suggest that speakers with a higher level of dysfluency at baseline (no-feedback) condition have achieved larger relative dysfluency reductions compared to speakers with a milder PDST level of dysfluency at baseline. This pattern is consistent with earlier reports (Van Borsel et al., 2007). For example, in Antipova et al. (2008), the authors reveal that the stuttering frequency of those subjects with more severe stuttering was reduced more than stuttering frequency of those with mild stuttering. The experience with our participants suggests that individuals with a milder dysfluency may not benefit as much from AAF procedures, since their speech motor system is relatively stable, and their level of confidence towards speech is relatively high. On the other hand, individuals with a more severe dysfluency, may be equipped with more fragile or sensitive speech motor systems (Van Lieshout et al., 2004), which could 'gain' more stability or reassurance from external cues.

Limitations of the current study

Individuals produce spontaneous speech in extremely variable manners, making experimental control more difficult for these tasks. It is, therefore, worthwhile to illustrate a number of potential sources of ambiguity related to the sampling of dysfluencies in spontaneous speech. First, speakers may be unconsciously using fluency techniques. Although we have explicitly instructed participants to not use any, at least two of them reported to have been using 'slow speech' and breathing techniques during the experimental tasks. Having internalized and automated fluency strategies to a certain degree, speakers may not be able to 'switch them off' upon command.

Second, when individuals who stutter produce a monologue on an engaging topic, they may place themselves in a 'performative' mode, in which their speech pattern deviates from their habitual way of speech, leading to enhanced fluency. Two of our participants have associated 'talking enthusiastically' with enhanced fluency, while noting that this situation may not represent their 'normal' speech performance. This observation is in line with an account of fluency enhancement outlined in Alm (2005), which proposes that switching one's speech manner from habitual to modified (as in imitation, acting, speaking with unnatural rhythm etc.) may enhance fluency through the de-automatization of speech motor control. According to this view, fluency is achieved by switching speech production from the supposedly dysfunctional medial system (the basal ganglia & supplementary motor area) responsible for automatized speech movements, to the lateral system (lateral premotor cortex & cerebellum) which coordinates de-automatized, and externally cued speech production (Alm, 2004).

Furthermore, dysfluencies seem to be strongly modulated by the presence of 'difficult' phonemes in the stream of speech, individual to each person. Anticipating those 'fearful sounds' is an integral part of the psychological experience of most people who stutter. Consequently, these speakers develop a variety of idiosyncratic strategies (which are distinct from 'standard' fluency techniques) for coping with the problematic sounds, such as reformulating phrases, substituting synonyms, and others. The influence of these 'online speech manipulations' on fluency may interact with the effects of external audio cues in an unpredictable way.

6.5 Conclusions and future directions

The adaptive feedback procedure used in this study resulted in strong reductions of dysfluencies at phrase onset positions. The results suggest that even higher reduction can be achieved if the audio feedback cue persists beyond the moment of phrase onset. Although the results indicate that an audio cue provided during silent segments may assist some speakers to initiate speech, they do not support the hypothesis that on a group level, this form of auditory cues result in a significantly larger fluency enhancement than the classic DAF procedure. The results for overall fluency levels reveal that an audio cue provided during silent segments leads to fluency improvement comparable with that of a standard DAF procedure. Investigation of individual results revealed a large degree of variability, confirming earlier evidence that persons who stutter react differently to AAF procedures.

At the outset of the study, it was assumed that, at least for some speakers who stutter, an adaptive presentation of feedback signals will appear as more comfortable than the continuously presented feedback signal in standard AAF procedures. However, considering the reactions of the study participants, it now seems that playing back a buffer of 1500 *ms* was not an optimal choice for an audio cue during silences. Hearing a segment this long seems to create a distracting effect in the process of speech production, and can be experienced as uncomfortable for some speakers. The distracting effect may be attributed to the interference with higher-level linguistic processes of utterance planning, which would normally take place during the silent pauses between phrases.

Within the framework of adaptive feedback procedures, audio cues can be considered which do not involve the straightforward playback of one's speech. Instead, an auditory cue can be designed which preserves the characteristics of the speaker's voice, but is devoid of linguistic content. Auditory cues can be further abstracted from the speaker's voice, and contain pre-recorded audio samples which can vary in the degree of linguistic content, from single syllables or single vowels, to non-linguistic cues, such as percussive sounds, or even synthetic tones and beats. There is, in fact, experimental evidence for fluency enhancing effects of non-linguistic auditory cues (Azrin et al., 1968; Martin and Haroldson, 1979; Howell and Archer, 1984), which could be used within an adaptive scheme. This approach would especially be relevant from a theoretic perspective which attributes fluency enhancing effects of auditory cues to influencing the timekeeping processes in speech production, thus not requiring linguistic content (Howell, 2004). However, following the insights from our study, it can be suggested that arriving at effective and comfortable feedback configurations can be best done on an individual basis.

Our experience with studying the effects of AAF procedures with people who stutter brings into question the group experimental design. The extent of individual variability (especially in spontaneous speech tasks) makes it difficult to draw any solid conclusions on the group level. To be able to potentially observe group effects, a considerably large sample size would be needed, a demand which may slow down research activities. Another constraining factor is the standardization of AAF settings when studies focus on the group level. It has become a common understanding that individuals benefit most from customized AAF settings, and by standardizing these settings during studies, researchers may be overlooking the potential value of AAF procedures. Instead, research could focus on single-case studies where investigators strive to optimize AAF settings for each individual, and examine how performance with these settings changes over time and communicational contexts. This approach is echoed in the conclusions drawn by Lincoln et al. (2010), who propose that more useful information will be gained from carefully designed case studies that establish optimal AAF settings for individuals who stutter.

We believe that the current investigation provides a good starting point for further experimentations with variations of adaptive feedback procedures. There is little doubt that results can be improved, both in terms of fluency enhancement, and levels of comfortability with the procedure. Although we have outlined a number of potential caveats in conducting AAF experiments with spontaneous speech, the motivation for this approach must be put forward. The social validity of applying AAF procedures to realistic communication situations is of great importance (Lincoln et al., 2010), and demands a better understanding of fluency modulation in the context of spontaneous speech. Collecting participants' impressions about their experience with the experimental conditions has been instrumental for a better understanding of the limitations and potentials of the proposed AAF procedures. Future attempts should take into account the aspect of individual variability, customization vs. standardization of AAF settings, and remain attentive to the experience of persons who stutter.

Despite the vast amount of research into stuttering during the last 40 years, there is still no consensus among researchers and clinicians (often decorated by heated debates) about the nature and etiology of the disorder, or the goals and methods of clinical intervention (Ratner, 2010). The widespread availability of affordable mobile technologies is likely to keep driving the development of a new generation of stuttering management devices. It is likely that future devices will be based on adaptive feedback procedures, with robust signal analysis capabilities, and elaborate customization options. In spite of the complexity involved, we strongly believe that it will be worthwhile to further explore adaptive feedback procedures, and devise methods for AAF procedures to be 'tailored' to individuals' speakers needs.

CHAPTER 7

General discussion and conclusions

The word of man is the most durable of all material.

Arthur Schopenhauer

7.1 Main conclusions and insights

In this chapter we summarize the main findings and conclusions, in regard to the different phases of treatment delivery addressed in this thesis, and propose directions for further development. Next, the rationale for a comprehensive framework for technological innovation in SLT are presented, and a functional model of that framework is outlined.

Summary of contributions

The work presented in this thesis contributes to the field of speech-language therapy through the demonstration of specific methodological refinements applied to three central phases of treatment delivery. First, a methodological shift has been proposed for the preparation of treatment programs – from the traditional usage of rather static inventories of speech items – towards providing clinicians with the means for compiling individually customized treatment targets and exercises. Chapters 3 and 4 presented two complementary steps on the way to realizing this refinement – the obtaining of linguistic materials needed for generating parametrized speech targets, and the design of an appropriate computer-interface for clinicians, allowing to perform the generation. The results of Chapter 4 confirm the assumption that speech clinicians, in general, recognize the added value of being able to compile customized treatment targets for their patients, as well as the efficiency and flexibility offered by the computer application designed for that purpose.

Next, a methodology has been proposed for practicing timing skills in speech with a computer game. Although voice-based computer games in speech-language therapy are not new, and several systems for practicing timing skills in the motor domain exist, the prospect of training timing skills with speech input has not yet been examined. The investigation presented in Chapter 5 contributes to the understanding of the elements necessary for constructing a training application for timing skills in the speech domain.

Finally, an improvement of the current methodologies of altered auditory feedback (AAF) for people who stutter is proposed and tested. Chapter 6 examined the utility of *adaptive* feedback procedures, which are based on real-time analysis of the speech signal and the dynamic activation of auditory cues, as oppose to the static feedback function utilized in current AAF methods. The results of this work provide valuable insights for future attempts of taking forward the methods for enhancing the speech fluency of individuals who stutter. Next, we discuss possible future directions in the work presented in this thesis.

Generation of speech treatment materials

In order to enable a systematic selection of treatment targets, databases of linguistic material are necessary, which are annotated with clinically relevant parameters. In Chapter 3, spoken and canonical syllabic inventories of Dutch were obtained by means of automatic syllabification of spoken language data, and the consistency of both syllabic inventories was examined, providing insights on their applicability as speech training materials. The obtained syllabic inventories may, in turn, be used to address ongoing psycholinguistic inquires. For example, by examining the way spoken syllable frequencies interact with the frequencies of the sub-syllabic parts (demisyllables, onsets, and rhymes), and with the frequencies of bi-syllabic patterns, new insights can be sought about the internal organization of the mental syllabary, and the role of syllables in speech production. Enabling clinical researchers to perform an elaborate parameterization of linguistic materials in their studies may facilitate the investigation of finer-grain phenomena in pathological speech production. Enabling clinicians to include and structure these materials in their treatment programs may facilitate the delivery of treatment based on current theoretical knowledge.

In Chapter 4, a computer application has been described, designed to provide a flexible, efficient, and intuitive workflow. A usability study with the application has echoed highly positive responses, and clinicians regarded the program as a useful, time saving tool for the creation of treatment programs. A natural evolution in this direction will be linking the various functions involved in the preparation of speech treatment programs into one coherent computer application, which will facilitate a structured workflow, hinged on best-practice evidence from the clinical literature. Besides promoting a systematic delivery of treatment, a computer application can be designed to access different databases of linguistic materials, which can be easily extended and updated (for example, a web-based implementation would ensure such flexibility). Finally, it is likely that all the different procedures involved in preparing treatment programs for MSD patients will eventually be integrated into one coherent computer application for the management of treatment protocols. Furthermore, mobile devices may gradually replace paper-based protocols used for presenting speech targets to patients during practice, as well as for the entering of results by the clinician.

Computer games for the training of speech motor skills

In Chapter 5, the use of a computer game for practicing speech timing skills of children with MSD was examined through a usability study. The results of the investigation presented in this thesis suggest that interactive games for practicing speech timing skills can promote a positive experience of speech motor training and enhance motivation for practice, at least in the short term. The interaction with the game added to the training situation, and was appreciated by children. Naturally, children have high demands on quality and entertainment values of the computer games they play, and these demands remain valid also in speech training situations. Consequently, there is a significant risk that custom-made games will not be able to compete with entertainment-oriented commercial games. One possible way to address this challenge is by *retrofitting* existing games, so that they can be interfaced with audio input. For open-source computer games, audio processing can be directly integrated within the game's code, producing an audio-based variant of that game. For example, for supporting the training of speech timing skills discussed in this thesis, the following two games could be considered (See Figure 7.1). In both cases, since progress within the game is determined by the timing of input gestures, syllable detection events could be mapped to the dynamics of the game.

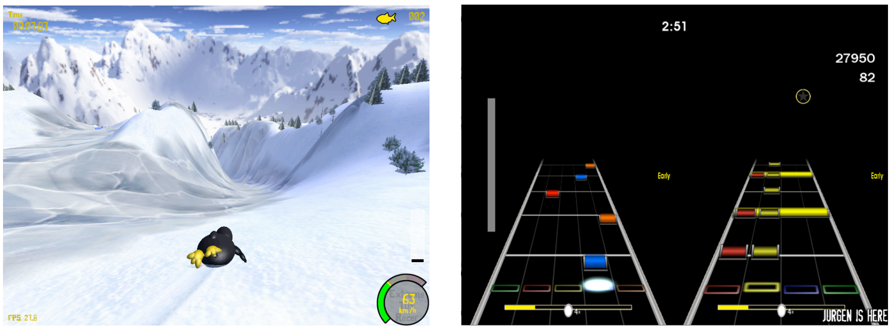


Figure 7.1: Tux Racer (left) is an open-source downhill racing game, where the player controls Tux (the Linux mascot) as he slides down a course of snow, collecting herring. Frets on Fire (right) is is an open-source music video game that imitates the commercial game Guitar Hero. The game is normally played by using a keyboard, pressing the fret and pick buttons.

The immediate advantage of this approach is in opening up a large variety of potentially effective synergies between available computer games, and the desired training effect throughout the gaming activity. The large selection of different games and interaction modes should provide opportunities to choose and match games according to the specific exercise goals and skills to be trained. Furthermore, utilizing popular computer and video games for therapy purposes is likely to decrease the risk that gaming will be perceived as tedious, compared to custom-made games (Sandlund et al., 2009a).

However, a number of potential caveats in this direction must me mentioned. First, available games may not be able to provide as specialized training as would games designed particularly for a certain practice methodology. Second, the mapping between acoustic parameters to action commands native to a certain game may not be intuitive, and designers may need to re-map these parameters for different games.

Demonstrating clinical effectiveness of computer game training methods in speech therapy with children involves an obvious difficulty. Exercises performed with a computer game only constitute a fraction of the whole range of treatment activities, regardless of the disorder being treated. Therefore, it is rarely possible to isolate the effect of this form of training on speech performance, from the effects of other treatment activities, the deferral of which is not realistic from a clinical perspective (Onslow et al., 1994).

Refinement of fluency enhancing procedures

Altered Auditory Feedback (AAF) devices grow smaller and more inconspicuous with the years. However, the underlying feedback procedures remain unchanged, characterized by a significant limitation of not taking into account the 'state' of the speaker at any given moment. A possible refinement examined in this thesis is the design of *adaptive* AAF procedures. Future efforts in advancing AAF methodologies should take advantage of techniques and theories from other related fields. For example, the 'synchronous reading' task (a situation when two people read the same text simultaneously – a fluency enhancing condition known as 'choral speech') has been thoroughly investigated with fluent speakers, by Cummins (2009). Integrating knowledge from similar experiments is likely to promote better theoretical understanding of the underlying processes of fluency modulation in individuals who stutter.

Although most participants in our study were aware of the existence of AAF devices, none had a prolonged positive experience with such a device. This situation might indicate that current AAF methodologies may not be optimally developed. As long as AAF methodologies will not include intelligent procedures to customize and adapt settings to individuals, it is not likely that their potential benefits to people who stutter will be fully realized. The design of adaptive feedback procedures involves a large space of possible configurations in terms of input signal analysis, selection of auditory cues for output, and the mapping function between input and output. Since the adaptive procedure examined in this work has lead to a fluency enhancement comparable with current AAF methods, we believe that this line of development will contribute to future efforts to innovate fluency-enhancing techniques. Support for this notion comes directly from participants in our study, the majority of whom have expressed positive attitudes towards these efforts. We quote one of the participants:

> "Ik denk wel dat er op een gegeven moment manieren komen waarop je elke vorm van spraak zo kan sturen dat hij vloeiend gaat lopen, dat er wel middelen voor te maken moeten zijn."
>
> "I believe that at some point methods will be available with which every form of speech could be directed towards fluent speech, that tools for this purpose will be developed."

7.2 A durable development framework

Challenges in technological innovation for SLT

A central question concerning innovation efforts in the field of communication disorders is - who, and within which framework, will initiate and carry out this kind of work. In principle, this can happen in two possible ways. First, clinicians may identify a potential for improving a certain methodology they use, and seek partners to provide the solution. Second, a group of researchers who develop a certain technology may wish to apply their techniques to a real-world problem, and seek clinical partners to identify a speech-language treatment methodology as an appropriate application domain.

In reality, although both workflows exist, their cumulative productivity does not, as it seems, reach a stable inertia of innovation in the field. A number of reasons can be put forward. First, due to the fact that SLT education curricula usually do not include experience with technological topics, the majority of clinicians do not promote innovative ideas involving technological concepts. As a consequence, partly due to lack of awareness of technological possibilities, there is, generally, little pro-active development initiation from the side of communication disorders specialists.

On the other hand, students and researchers at technical departments seldom choose to apply their skills to the field of communication disorders. Surely, there are exceptions, often driven by a personal experience, or interest. The resulting situation is that innovation efforts in the field occur rather sporadically, with few dedicated research and development structures to advance these efforts. This stands in stark contrast to other fields of assistive technology, for which ongoing research activities are organized within special interest groups, laboratories and specialized centers (Pino et al., 2010; Bates et al., 2007).

Towards a dedicated framework

The experience gained with development projects described in this thesis converges to the conclusion that a dedicated framework is needed in order to promote a durable process of innovation in the field of communication disorders. Such a framework would consist of ongoing collaborative research and development activities, involving university departments, and eventually industrial partners. Recently, more researchers recognize the need in an overarching framework for advancing innovation in the field of SLT. Cucchiarini et al. (2008) provide examples of successful previous efforts to organize a durable framework for the development of language technologies in the Netherlands, and propose mechanisms for the establishment of a comprehensive stimulation program for utilizing these technologies for the support of speech language therapy (Ruiter et al., 2010), supported by government, industry and academia.

In order to further contribute to these efforts, we propose a functional model of a dedicated framework for innovation in the field of SLT. The eventual aim of such framework is to enable a perpetual generation of innovative solutions, from which therapists can pool resources and capabilities for advancing novel technology supported therapies. The purpose of the model is to define the participating actors, and outline the main activities, interactions, and workflow within the framework.

A model of innovation in the field of SLT

First, the main actors participating in the framework are identified as follows:

SLT department A university department where speech language therapists are educated and trained. Ideally, the department is oriented towards research, and maintains ongoing collaborations with clinical institutions.

Partner departments A number of university departments where the realization of proposed solutions takes place. These will be computer-science, engineering, or other technical departments, as well as groups dealing with graphical design, interaction design, film and new media.

GISLT A team which consists of experts from both SLT and partner departments, whose roles are to lead the coordination between the involved disciplines. The team member's role is twofold: they act as supervisors to student projects from their respective department, and they are responsible for the selection, analysis and design of the initiated projects. In order to refer to this team in a consistent manner, we shall term it – *Group for Innovation in Speech Language Therapy* (GISLT).

Next, the workflow is described, which outlines the basic interactions between the participating actors, and their respective roles in the framework (see Figure 7.2). The model consists of the following steps:

Innovation in SLT course. At the heart of the model is the educational course for speech language therapy students, which aims at generating innovative concepts and solutions for real-life problems and identified improvement potentials in this clinical field. A short description of the course, it's goals and methods is given in Appendix B.

Project selection. The course for SLT generates a number of concepts and project proposals for innovation in the field. These proposals are evaluated for their potential impact and feasibility, so that a limited number of projects are selected for realization. For each selected project, the requirements are further specified and the design is elaborated in preparation for the development phase. These activities are carried out by the GISLT.

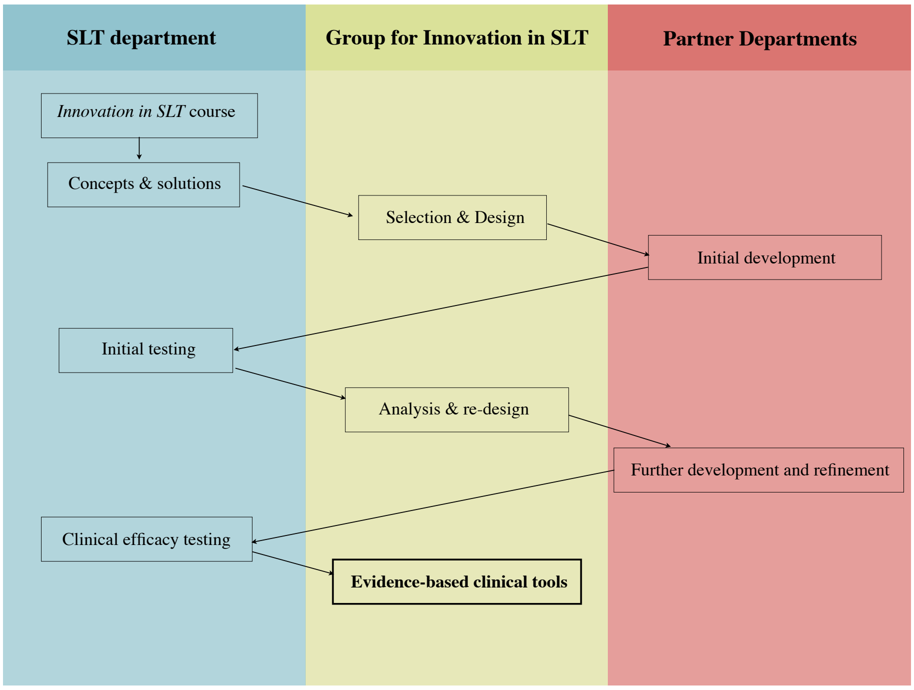


Figure 7.2: An outline of the innovation in SLT model

Initial development. Selected projects are defined and offered to students as a graduation/thesis project, within an appropriate university department (which commits to an ongoing collaboration with the GISLT), where the required expertise, in relation to the defined problem, exists . Members of the GISLT are appointed as co-supervisors for these projects. The goal of this phase is producing a working prototype of the proposed solution, with sufficient documentation to allow subsequent phases of development.

Initial testing. The developed prototype is than evaluated in an initial usability study. The study is likewise defined as a graduation project within the SLT department, co-supervised by the GISLT. The study aims at gathering performance and usability data from real clinical experiences with the proposed solution by the target user group.

Analysis and re-design. The data collected from initial testing is analyzed and the feasibility of the proposed solution is re-evaluated by the GISLT. In case of positive outcomes, further refinements to the design are made, based on the collected data, and a further development phase is defined.

Further development. The refined specifications form the basis for another student project, possibly within the same department where the initial development took place. The goal of this phase is producing a more mature realization of the solution, eventually including the human-computer interface for allowing a long-term deployment by non-experts.

Efficacy testing. The potential clinical impact of the proposed solution is evaluated through an efficacy study with the developed application. Defined as a group project within the SLT department, the study is carried out in collaboration with clinical partners, and aims at examining evidence for the effects of utilizing the proposed solution in clinical procedures.

Evaluation and deployment. In the case that efficacy testing produces positive outcomes, the GISLT takes further steps towards integrating the developed solutions into the field. If further refinements are indicated, an additional cycle of development may be defined. Eventually, mature products can be distributed within the clinical field, or commercialized through the university's spin-off facilities.

The proposed model entails a number of potential advantages in relation to the current, sporadic way of advancing technological innovation for SLT:

» The initiation and the drive for creating innovative solutions comes from the clinical field (through the participation of SLT students in the practical part of the SLT course).

» The framework may promote the application of a wealth of technological knowledge resources, residing in technical departments, to real-life problems, as well as promoting awareness of assistive technologies among researchers in these departments.

» The framework may facilitate the commitment of speech clinicians to participate in long-term clinical studies.

» Since most of the activities throughout the model are performed within graduate student projects, innovation efforts are not heavily constrained by market size. The commercialization of eventual products is an option, but not a precondition for the research and development agenda.

7.3 Final remarks

The work presented in this thesis aimed to operate as the missing link between the needs of the clinical field of speech-language therapy, and a pool of existing techniques from technological domains (indications for that missing link originated in communication with different speech clinicians, who confirmed that efforts are needed to bridge this gap). A successful aspect of that role can be seen in demonstrating that a variety of clinical procedures can be supported through technological innovation, once attention is given to identified problems. However, the work process has also revealed significant structural challenges in the efforts to coordinate technological development in the field. The interpretation of these challenges suggested a need for a comprehensive, durable framework, and has led to the proposal of a model for technological innovation for SLT, described in 7.2.

Particularly, the model extends the workflow carried out within this thesis (as outlined in 1.3) in two important ways. First, the identification of problems and needs, as well as the generation of innovative concepts is placed within the field of SLT itself. This strategy ensures that the Analysis Phase is performed with the strongest involvement of clinicians, so that most relevant problems in the field are addressed. Second, established collaborations within the model should ensure that multiple iterations of development and testing can be carried out, on the way to mature clinical tools.

Finally, with all due respect to the potentials of technological innovation for supporting the procedures of speech-language therapy, it is good to redraw the realistic proportions of their role. In spite of the fact that technology-based methods have been the focus of this thesis, we must remind that in the very central role in the delivery of treatment always stands the speech therapist, and the role of technological support is complementary. After all, to be able to assist people with communication disorders, a kind of sympathetic sensitivity is required, which is uniquely human.

APPENDIX A

List of syllable types

Table A.1: The top 100 syllable types and their frequencies for the intersection and difference sets of the spoken (SPO) and canonical (CAN) syllabaries. Frequency expressed as percentage of all syllable tokens in the set.

SPO ∩ CAN			SPO \ CAN		CAN \ SPO	
Type	Freq. SPO	Freq. CAN	Type	Frequency	Type	Frequency
[də]	4.15	3.89	**[dɑd]**	0.33	**[gɔːt]**	0.020
[tə]	2.52	1.89	**[ɪg]**	0.32	**[wʌuw]**	0.016
[ja]	2.45	2.51	**[əd]**	0.26	**[bʌuwt]**	0.005
[xə]	1.86	1.60	**[og]**	0.17	**[tənts]**	0.003
[ɛn]	1.66	1.90	**[mɛd]**	0.11	**[ɟuw]**	0.003
[jə]	1.36	1.49	**[fel]**	0.11	**[prɔx]**	0.003
[ɪn]	1.23	1.27	**[wɑd]**	0.10	**[pɑː]**	0.003
[dɑt]	1.19	2.07	**[hɛb]**	0.10	**[sxʌuwt]**	0.002
[di]	1.11	1.26	**[əŋ]**	0.10	**[kɔrbz]**	0.002
[ət]	1.08	1.16	**[nɔɣ]**	0.09	**[tʌuw]**	0.002
[ən]	1.08	1.47	**[ləg]**	0.06	**[fɑjls]**	0.002
[ɪk]	1.07	1.51	**[sɛx]**	0.06	**[nʌuwt]**	0.001
[bə]	1.03	0.96	**[tɑd]**	0.06	**[dʒems]**	0.001
[sə]	0.95	0.50	**[wɑnd]**	0.06	**[mʏːrs]**	0.001
[o]	0.94	0.83	**[xwon]**	0.05	**[grɛː]**	0.001
[lə]	0.94	0.94	**[wed]**	0.05	**[tʏrf]**	0.001

[kə]	0.91	0.90	**[ɑlz]**	0.05	**[plɛ̃]**	0.001
[ɪs]	0.83	1.19	**[tɔɣ]**	0.05	**[pɑːs]**	0.001
[ɣə]	0.82	1.15	**[əz]**	0.05	**[trek]**	0.001
[mə]	0.82	0.56	**[ɑz]**	0.05	**[dʒɔr]**	0.001
[rə]	0.77	0.72	**[ɛb]**	0.04	**[tiɟ]**	0.001
[fɑn]	0.77	0.01	**[xən]**	0.04	**[trɛit]**	0.001
[nə]	0.72	0.52	**[ləf]**	0.04	**[rɑɟtʃ]**	0.001
[ɔp]	0.71	0.81	**[trʏx]**	0.04	**[mwa]**	0.001
[ne]	0.67	0.68	**[əg]**	0.04	**[dʒɛ]**	0.001
[ok]	0.62	0.76	**[dəd]**	0.04	**[pɔzd]**	0.001
[wɛl]	0.60	0.66	**[əl]**	0.03	**[ʒɔ]**	0.001
[fə]	0.60	0.07	**[dɪd]**	0.03	**[sxrør]**	0.001
[və]	0.60	0.51	**[tət]**	0.03	**[tmɪs]**	0.001
[ma]	0.57	0.26	**[hef]**	0.03	**[ɣno]**	0.001
[mar]	0.57	0.99	**[xad]**	0.03	**[dʒo]**	0.001
[wə]	0.55	0.53	**[tyk]**	0.03	**[lant]**	0.001
[nit]	0.54	0.85	**[nja]**	0.03	**[kʏnzd]**	0.001
[dər]	0.52	0.92	**[wɔrd]**	0.02	**[hɑrk]**	0.001
[dɑn]	0.51	0.85	**[fɑm]**	0.02	**[bʏlt]**	0.001
[ɔm]	0.50	0.48	**[dɑz]**	0.02	**[blowt]**	0.001
[əm]	0.48	0.25	**[əf]**	0.02	**[dɪxtst]**	0.001
[ɑ]	0.48	0.37	**[ləv]**	0.02	**[sɔp]**	0.001
[na]	0.48	0.41	**[xon]**	0.02	**[pɛiŋ]**	0.001
[nʌu]	0.47	0.57	**[hət]**	0.02	**[zwart]**	0.001
[zə]	0.43	0.74	**[tɔd]**	0.02	**[lɛj]**	0.001
[ni]	0.42	0.17	**[fɪnd]**	0.02	**[lɑjns]**	0.001
[an]	0.42	0.43	**[rɑl]**	0.02	**[ʒɛf]**	0.001
[lək]	0.42	0.55	**[tɪt]**	0.02	**[nɑk]**	0.001
[ɑl]	0.41	0.41	**[fru]**	0.02	**[pɑːst]**	0.001
[hɛ]	0.40	0.26	**[rəkt]**	0.02	**[namz]**	0.001
[i]	0.39	0.34	**[xəs]**	0.01	**[tfər]**	0.001
[wɑt]	0.39	0.47	**[əp]**	0.01	**[rɪŋz]**	0.001
[bɛi]	0.38	0.42	**[ɛpt]**	0.01	**[rɑnts]**	0.001
[e]	0.38	0.31	**[ʏs]**	0.01	**[frək]**	<0.001
[vɑn]	0.38	1.31	**[ləz]**	0.01	**[vrʌuwt]**	<0.001
[mɛt]	0.37	0.51	**[kɛig]**	0.01	**[dʒɛm]**	<0.001
[so]	0.37	0.04	**[eft]**	0.01	**[phi]**	<0.001
[for]	0.34	0.01	**[ɛix]**	0.01	**[løz]**	<0.001
[pə]	0.33	0.25	**[wəs]**	0.01	**[vens]**	<0.001
[wɑs]	0.33	0.40	**[bɛŋ]**	0.01	**[sxɑlk]**	<0.001
[œyt]	0.33	0.34	**[nɛd]**	0.01	**[vlar]**	<0.001
[tər]	0.33	0.40	**[xər]**	0.01	**[sats]**	<0.001
[ɛ]	0.32	0.10	**[kɔmd]**	0.01	**[wʏːrlt]**	<0.001
[ɔf]	0.31	0.39	**[ɔnz]**	0.01	**[mʌuw]**	<0.001
[ŋə]	0.30	0.29	**[hefd]**	0.01	**[hedn]**	<0.001
[dʏs]	0.30	0.47	**[ers]**	0.01	**[tɑzd]**	<0.001
[hel]	0.30	0.27	**[ɛiɣ]**	0.01	**[vɔrst]**	<0.001
[ɪz]	0.29	0.00	**[təd]**	0.01	**[nɪrt]**	<0.001

[**ti**]	0.29	0.12	[**stad**]	0.01	[**srœym**]	<0.001
[**nɔx**]	0.29	0.44	[**wət**]	0.01	[**jɑrs**]	<0.001
[**dar**]	0.27	0.49	[**rhal**]	0.01	[**nʌux**]	<0.001
[**he**]	0.27	0.24	[**tyg**]	0.01	[**dʒons**]	<0.001
[**ɔn**]	0.26	0.22	[**xeŋ**]	0.01	[**rɛɣdz**]	<0.001
[**me**]	0.26	0.25	[**zɪd**]	0.01	[**varts**]	<0.001
[**vər**]	0.26	0.87	[**mɑnd**]	0.01	[**xrʌuw**]	<0.001
[**te**]	0.26	0.24	[**əv**]	0.01	[**wrɪx**]	<0.001
[**de**]	0.25	0.23	[**psis**]	0.01	[**kɑrl**]	<0.001
[**si**]	0.25	0.18	[**kʏnd**]	0.01	[**tɔrs**]	<0.001
[**da**]	0.24	0.09	[**siz**]	0.01	[**ŋəx**]	<0.001
[**ɑn**]	0.24	0.20	[**mɛnd**]	0.01	[**dʌuwn**]	<0.001
[**tɑt**]	0.24	0.00	[**rəɣ**]	0.01	[**dbə**]	<0.001
[**hɛi**]	0.24	0.24	[**kənd**]	0.01	[**tlɛi**]	<0.001
[**dɑ**]	0.24	0.01	[**rəp**]	0.01	[**jɑrx**]	<0.001
[**we**]	0.23	0.15	[**ʝɑ**]	0.01	[**vuɣ**]	<0.001
[**kɑn**]	0.23	0.27	[**sord**]	0.01	[**sɑlts**]	<0.001
[**ko**]	0.23	0.21	[**kom**]	0.01	[**dʒərs**]	<0.001
[**le**]	0.22	0.21	[**fɔnd**]	0.01	[**bujnx**]	<0.001
[**war**]	0.22	0.28	[**ntyk**]	0.01	[**kʌux**]	<0.001
[**fər**]	0.21	0.03	[**dɛŋt**]	0.01	[**wɛm**]	<0.001
[**zo**]	0.21	0.57	[**rəg**]	0.01	[**sirs**]	<0.001
[**wa**]	0.21	0.14	[**əb**]	0.01	[**mɑjl**]	<0.001
[**li**]	0.21	0.20	[**tuŋ**]	0.01	[**flɑ̃n**]	<0.001
[**zɛin**]	0.20	0.44	[**sɛiŋ**]	0.01	[**xris**]	<0.001
[**twe**]	0.20	0.17	[**dud**]	0.01	[**dinzd**]	<0.001
[**a**]	0.20	0.19	[**ʏz**]	0.01	[**mot**]	<0.001
[**en**]	0.20	0.06	[**twa**]	0.01	[**kɑlm**]	<0.001
[**y**]	0.20	0.16	[**løg**]	0.01	[**vɔjs**]	<0.001
[**ny**]	0.20	0.22	[**zɛiŋ**]	0.01	[**dʒɔp**]	<0.001
[**tɔx**]	0.20	0.24	[**fɪts**]	0.01	[**glɑːs**]	<0.001
[**xut**]	0.20	0.22	[**xev**]	0.01	[**rɛw**]	<0.001
[**wer**]	0.20	0.21	[**mɪns**]	0.01	[**vrɔm**]	<0.001
[**mɑ**]	0.19	0.06	[**tɪx**]	0.01	[**zwɑlm**]	<0.001
[**wɑnt**]	0.19	0.27	[**vət**]	0.01	[**vardz**]	<0.001
[**təx**]	0.19	0.20	[**sɛlfs**]	0.01	[**skɔpf**]	<0.001

Table A.2: The top 100 bi-syllabic types and their frequencies for the intersection and difference sets of the spoken (SPO) and canonical (CAN) syllabaries. Frequency expressed as percentage of all syllable tokens in the set.

SPO ∩ CAN			SPO \ CAN		CAN \ SPO	
Type	Freq. SPO	Freq. CAN	Type	Freq.	Type	Freq.
[ɪn-də]	0.436	0.306	**[fɑn-də]**	0.164	**[ə-dɑt]**	0.078
[hɛ-bə]	0.298	0.304	**[ɔn-də]**	0.072	**[vɑn-ə]**	0.070
[o-vər]	0.259	0.431	**[o-fər]**	0.060	**[dɑn-ə]**	0.070
[də-rə]	0.249	0.214	**[ɔf-so]**	0.057	**[ə-ɪk]**	0.054
[he-lə]	0.227	0.232	**[mə-sxin]**	0.051	**[dʏs-ə]**	0.054
[ja-ja]	0.217	0.219	**[o-fə]**	0.050	**[ə-ən]**	0.054
[kʏ-nə]	0.208	0.145	**[ɔb-də]**	0.042	**[ə-ɛn]**	0.034
[o-və]	0.201	0.016	**[fɑn-ət]**	0.041	**[ɪk-vɪnt]**	0.032
[ɑ-lə]	0.199	0.188	**[dɑt-sə]**	0.040	**[ə-nʌu]**	0.030
[bet-jə]	0.194	0.142	**[for-də]**	0.039	**[lat-stə]**	0.029
[ko-mə]	0.177	0.135	**[ɪk-ɛp]**	0.037	**[ə-wɑt]**	0.028
[ɑn-də]	0.175	0.124	**[ɛi-xə]**	0.037	**[ə-mɛt]**	0.026
[e-və]	0.169	0.188	**[ty-lək]**	0.036	**[dɑt-vɪnt]**	0.023
[wɔr-də]	0.168	0.131	**[ɪz-ət]**	0.035	**[ə-dɑn]**	0.023
[mɛn-sə]	0.167	0.144	**[ɪg-bɛn]**	0.032	**[dɑt-zɛin]**	0.023
[ɛn-də]	0.165	0.098	**[wɛ-rək]**	0.032	**[ə-ɑls]**	0.023
[vɑn-də]	0.153	0.294	**[fɑn-di]**	0.031	**[vɛif-ɛn]**	0.022
[xə-won]	0.150	0.246	**[dɑd-ɪs]**	0.031	**[ə-ɪs]**	0.021
[lə-kə]	0.136	0.091	**[ɪz-ən]**	0.030	**[ɔf-ə]**	0.019
[ɔn-dər]	0.136	0.164	**[əl-kar]**	0.026	**[wɑnt-ə]**	0.019
[mu-tə]	0.136	0.116	**[fɑn-ən]**	0.026	**[ə-o]**	0.019
[o-ja]	0.136	0.109	**[dɑd-ɪz]**	0.024	**[vɪnt-ək]**	0.018
[də-xə]	0.134	0.089	**[ja-dɑ]**	0.024	**[ə-dar]**	0.017
[ɛn-dɑn]	0.129	0.185	**[ɛ-rəx]**	0.024	**[tin-mi]**	0.017
[dɑt-ɪs]	0.129	0.274	**[də-fə]**	0.024	**[ə-mar]**	0.016
[an-də]	0.128	0.102	**[dɑd-ɪk]**	0.023	**[dɑt-zʌuw]**	0.015
[ɛn-dɑt]	0.118	0.149	**[dɑt-ɪz]**	0.022	**[hɛt-ə]**	0.014
[ɔp-tə]	0.117	0.018	**[dɑd-ət]**	0.022	**[won-ə]**	0.014
[ɛi-ɣə]	0.113	0.187	**[ɛn-dɑd]**	0.022	**[tɛn-mɪn]**	0.014
[ɪn-ət]	0.111	0.074	**[ɪg-dɛŋk]**	0.022	**[ə-nit]**	0.014
[dɪ-ŋə]	0.111	0.081	**[ɪn-əd]**	0.022	**[ə-hu]**	0.013
[əm-bet]	0.111	0.000	**[wɛ-rə]**	0.022	**[nɔx-ə]**	0.013
[wa-rə]	0.110	0.093	**[rə-kə]**	0.022	**[ə-hɛt]**	0.013
[ne-də]	0.108	0.013	**[ɪg-bə]**	0.021	**[ə-tun]**	0.013
[də-də]	0.107	0.073	**[lats-tə]**	0.021	**[mɛt-zən]**	0.012
[kɛi-kə]	0.107	0.090	**[ɪs-tɑt]**	0.021	**[hɛt-vər]**	0.012
[bɛi-də]	0.107	0.075	**[ət-ɪz]**	0.020	**[ə-ɔf]**	0.011
[te-ɣə]	0.102	0.120	**[fo-də]**	0.020	**[ə-hel]**	0.011
[ma-kə]	0.101	0.083	**[əd-bə]**	0.019	**[ə-nar]**	0.011

[ɔp-xə]	0.100	0.070	[ja-ɪg]	0.019	[ə-an]	0.011
[ɔm-dɑt]	0.100	0.102	[pə-sis]	0.019	[ok-vor]	0.011
[o-ke]	0.099	0.000	[ja-dɑd]	0.019	[ə-dər]	0.011
[lə-mal]	0.099	0.244	[ni-mer]	0.019	[vɪnt-dɑt]	0.011
[de-zə]	0.096	0.098	[dɑt-tɑt]	0.018	[zɪɣ-zɛlf]	0.011
[ɛn-di]	0.095	0.082	[hɛb-ɪk]	0.018	[won-nit]	0.010
[xro-tə]	0.092	0.068	[og-nit]	0.018	[ə-ok]	0.010
[dɑt-jə]	0.092	0.102	[hɛb-ək]	0.018	[zɛxt-vɑn]	0.010
[la-tə]	0.091	0.069	[kʏ-jə]	0.018	[vər-ə]	0.010
[e-fə]	0.091	0.000	[ɪz-ok]	0.018	[mest-ɑl]	0.010
[də-bə]	0.087	0.066	[hɔn-dət]	0.017	[ə-war]	0.010
[ɣə-lək]	0.087	0.178	[aŋ-xə]	0.017	[its-vɑn]	0.010
[ɛn-ɪk]	0.085	0.087	[nə-ty]	0.017	[zʌuw-kʏ]	0.010
[ɑ-len]	0.085	0.093	[ɛn-ɪg]	0.017	[dɑn-zʌuw]	0.010
[er-stə]	0.082	0.064	[ɔm-dɑd]	0.017	[nar-ə]	0.009
[ət-ɪs]	0.081	0.168	[wed-ɪk]	0.017	[ɑl-ə]	0.009
[xə-dan]	0.081	0.057	[mɛt-ti]	0.017	[wɑt-dɑt]	0.009
[bɪ-nə]	0.080	0.059	[og-nɔx]	0.017	[zʌuw-ət]	0.009
[ɑx-tər]	0.078	0.089	[ɪk-ɛb]	0.017	[ɪk-vɔnt]	0.009
[mɛt-tə]	0.078	0.001	[we-ni]	0.016	[ə-e]	0.009
[hɑ-də]	0.077	0.070	[əŋ-ker]	0.016	[ə-dor]	0.009
[də-ɪn]	0.077	0.071	[ɑl-tət]	0.016	[rə-ə]	0.009
[ja-ɪk]	0.077	0.092	[mɛd-də]	0.015	[ɑls-ə]	0.009
[ja-dɑt]	0.077	0.140	[fɛ-dər]	0.015	[ə-dʏs]	0.009
[na-də]	0.075	0.019	[dɑt-sɛin]	0.015	[ɣə-ə]	0.008
[tʏ-sə]	0.075	0.063	[lə-mɑl]	0.015	[hɛp-ə]	0.008
[tə-bə]	0.073	0.038	[le-fə]	0.015	[mal-ə]	0.008
[ɑ-ləs]	0.073	0.066	[dɑd-dɑt]	0.015	[wits-vɑn]	0.008
[wɪ-lə]	0.073	0.050	[fɑn-dax]	0.015	[kə-ə]	0.008
[ɑf-xə]	0.072	0.051	[nɑ-ty]	0.015	[vɑn-mɛin]	0.008
[ɑx-tə]	0.072	0.005	[ɔŋ-xə]	0.015	[ə-wɛl]	0.008
[ən-ən]	0.071	0.059	[tə-wɛil]	0.015	[tɛit-ə]	0.008
[ɑl-tɛit]	0.070	0.108	[ət-fə]	0.015	[zon-ə]	0.007
[œyt-xə]	0.069	0.047	[ni-so]	0.015	[wer-ə]	0.007
[ɪk-hɛp]	0.069	0.137	[sɛx-mar]	0.015	[dərt-vɛif]	0.007
[ne-ne]	0.069	0.061	[ɪg-dɛŋg]	0.015	[vɑn-der]	0.007
[lo-pə]	0.069	0.055	[ɪŋ-xə]	0.014	[ə-hɛi]	0.007
[we-tə]	0.068	0.054	[ɑn-dəs]	0.014	[mar-əns]	0.007
[nʌu-ja]	0.068	0.095	[ɪk-hɛb]	0.014	[ək-zɛx]	0.007
[ɪn-ən]	0.067	0.068	[ɑl-tɛid]	0.014	[zʌuw-ək]	0.007
[hɛp-jə]	0.067	0.086	[də-for]	0.014	[den-hax]	0.007
[bœy-tə]	0.065	0.050	[ɣə-ləg]	0.014	[lək-vor]	0.007
[jʏ-li]	0.065	0.073	[fɛr-dər]	0.014	[zʌuw-zɛin]	0.007
[jə-wɛl]	0.064	0.040	[o-to]	0.014	[vɑn-den]	0.007
[tə-xə]	0.064	0.037	[sɛ-ləf]	0.014	[hɛt-ok]	0.007
[zɪ-tə]	0.063	0.081	[fɛif-təx]	0.013	[hel-lɑŋ]	0.007
[xə-west]	0.063	0.065	[ɑ-ləz]	0.013	[ə-twe]	0.007
[ət-xə]	0.062	0.040	[ɪk-fɪn]	0.013	[ət-zʌuw]	0.007

[**ɪn-di**]	0.062	0.046	[**mɛd-ən**]	0.013	[**ə-əm**]	0.007
[**ne-mə**]	0.062	0.045	[**ɪk-sɛx**]	0.013	[**dul-ə**]	0.007
[**na-mə**]	0.060	0.047	[**nə-tyr**]	0.013	[**vel-ə**]	0.006
[**bə-lɑŋ**]	0.060	0.047	[**mɔr-xə**]	0.013	[**ə-vɔl**]	0.006
[**wet-jə**]	0.059	0.047	[**hʌu-wə**]	0.013	[**ə-mut**]	0.006
[**i-dər**]	0.059	0.074	[**ɑ-mɑ**]	0.013	[**mɑ̃-dɑt**]	0.006
[**sa-mə**]	0.059	0.025	[**xə-fɑl**]	0.013	[**ək-vɪnt**]	0.006
[**nit-so**]	0.059	0.000	[**dɑd-də**]	0.013	[**ɔp-zən**]	0.006
[**ka-mər**]	0.059	0.046	[**ɑ-mɑl**]	0.013	[**ə-de**]	0.006
[**xə-le**]	0.058	0.043	[**bɛ-jə**]	0.012	[**ɛr-ə**]	0.006
[**bə-dul**]	0.058	0.059	[**sɛ-xə**]	0.012	[**mən-le**]	0.006
[**ɪn-tə**]	0.057	0.036	[**fɑn-fɑn**]	0.012	[**ɪk-vər**]	0.006
[**e-nə**]	0.056	0.032	[**xə-xe**]	0.012	[**ə-ək**]	0.006

APPENDIX B

The ‘Innovation in SLT’ course

B.1 Course description

At the center of the functional model described in 7.2 is an educational course, which aims at developing SLT students’ understanding of the issues associated with technological innovation in SLT. The course originated from the realization of the need to integrate the seeds of awareness about innovation in the field into the SLT education curriculum. It is a semester-long course for advanced speech language pathology students, with the following objectives in mind:

» The course will prepare students to identify and propose appropriate strategies for solving problems related to clinical procedures in SLT.

» Students will learn to investigate current methodologies applied in SLT practice, identify problems and needs, and propose solutions.

» Student will have gained familiarity with a range of technological solutions for typical problems in SLT practice, as well as an ability to source and develop new solutions, through practical experience within a specific topic of choice.

» As future clinicians, students will be encouraged to be proactive in their working environment, and consider the role of technological innovation in creative problem solving.

Course structure

The course involves two parts: theory and a practicum. During theory classes, students are introduced to a range of innovative applications in the field of SLT. Various communication disorders are reviewed, and state-of-the-art tools which address the associated specific needs are illustrated through rich multimedia materials. The principles and technologies behind these tools, as well as assistive applications in other clinical fields are discussed. The following topics are addressed:

- History of innovation in the field of speech language therapy.
- Application of biofeedback techniques for training various speech skills.
- Tools developed to assist people with hearing impairments.
- Tools and methodologies to which assist people who stutter.
- Applications of tele-rehabilitation in SLT.
- Augmented and Alternative Communication (AAC) systems.
- Special tools developed for working with children with autism.

During the practicum, students experience the process of creating an innovative concept themselves by going through a structured work-process. Students divide into groups and work together on a project towards a formulation of an innovative concept. Groups are supervised throughout their work process, with regular meetings, discussions and evaluation sessions. The practicum proceeds through the following phases:

Choose a context. Each group chooses a clinical context to focus on. This can be a certain clinical population, a treatment method, or any niche of SLT work which is of interest to the group.

Identify a need. Each group performs a period of field work in a clinical setting (e.g., hospitals, schools, speech clinics), in order to investigate the current state of a certain treatment/methodology/activity, or other aspects of the clinical situation, and identify a certain need/problem through observations, interviews and research.

Conceptualize a solution. Next, groups work on the proposal and formulation of an innovative concept which addresses the findings from the field-work. Different solutions are evaluated and discussed, by analyzing their relative feasibility, the potential benefits and problems.

Presentation & discussion. Finally, each group prepares an end-presentation for the class with an overview of their work process.

Experience

The *Innovation in SLT* course was given during the 2010 Spring semester to 4th year speech language pathology students at the Department of Speech Therapy and Special Education, Faculty of Education at the University of Ljubljana, in Slovenia. During the first 6 classes, theory material was given, accompanied by discussions on the current state of SLT practice in Slovenia. Further on, students divided into 6 groups based on common interests, and chose a clinical context for the practicum. Student groups focused on treatment methodologies in aphasia, stuttering, phonological disorders, as well as on aspects of AAC protocols. Appointments with clinicians were made in order to allow students to perform field-work in clinical settings, with the support of the faculty.

After the period of field-work, regular meetings were held to process the collected information and come up with ideas for improvement. Finally, the groups summarized their findings and prepared end presentations. The practicum resulted in a number of innovation proposals, including a computer game for training phonological discrimination skills, a computer game for training correct tempo of speech and fluent transitions between phonemes for children who stutter, improvements to the usability of PECS (Picture Exchange Communication System) and WIWIK (an AAC application used in Slovenia's rehabilitation centers), a system to assist aphasic patients to independently perform oral motor exercises, and a computer game to complement the VLAJA method used in therapy with adults who stutter.

Evaluation

In order to evaluate the potential benefits of the developed course for enhancing the SLT education curricula, we have collected questionnaire responses from SLT students, as well as from speech therapists working in Slovenia. Items were presented in a 5-point Likert-type scale, and included statements regarding both the specific merit of the *Innovation in SLT* course, as well as the importance of preparing SLT students to deal with new technological developments in their future work. In total, 53 SLT students and 39 speech therapists have completed the questionnaire. The outcomes of this evaluation are presented next as percentages of possible replies provided by the group of SLT students and speech clinicians.

(1) The goals of the 'Innovation in SLT' course are clear to me.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	2%	7%	11%	55%	23%
Clinicians	0%	8%	29%	45%	18%

(2) It is necessary to prepare SLT students to deal with technological developments in their future professional work.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	0%	0%	6%	20%	75%
Clinicians	0%	0%	0%	40%	60%

(3) Technological innovation poses a threat to the profession of SLT.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	23%	44%	32%	0%	0%
Clinicians	37%	40%	21%	0%	3%

(4) The theoretical part of the ‘Innovation in SLT’ course covers some of the essential topics in speech-language therapy practice.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	2%	2%	21%	68%	6%
Clinicians	0%	0%	31%	58%	10%

(5) Learning about assistive technologies in other clinical areas contributes valuable knowledge for SLT students.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	2%	2%	4%	55%	37%
Clinicians	0%	0%	10%	55%	34%

(6) The field-work which the students perform during the ‘Innovation in SLT’ course is a good way for students to identify potential areas for innovation.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	0%	4%	25%	57%	14%
Clinicians	0%	8%	31%	58%	10%

(7) Most speech clinicians will be willing to cooperate with students during their practicum field-work.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	4%	12%	49%	30%	6%
Clinicians	3%	10%	40%	36%	10%

(8) During the course, SLT students can generate valuable innovative concepts.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	2%	0%	10%	64%	24%
Clinicians	0%	0%	21%	42%	36%

(9) Th? 'Innovation in SLT' course can provide a valuable contribution to the education curriculum of SLT students.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	2%	0%	8%	54%	36%
Clinicians	0%	0%	10%	60%	29%

(10) The course provides knowledge which can later be used in the professional practice.

	Strongly disagree	Disagree	Partly agree	Agree	Entirely agree
Students	0%	2%	2%	66%	30%
Clinicians	0%	0%	8%	55%	36%

The results above suggest that the course addresses relevant aspects of speech-language therapy practice, that students are able to gain knowledge which can later be used in their professional work. Both students and clinicians agreed that the course can provide a valuable contribution to the education curriculum of SLT students. From the perspective of the model described in 7.2, the course seems as a promising incubator, where SLT students can generate valuable innovative concepts.

Bibliography

Adsett, C.R. and Y. Marchand. A Comparison of Data-Driven Automatic Syllabification Methods. In *String Processing and Information Retrieval: 16th International Symposium, SPIRE 2009 Saariselkä, Finland, August 25-27 Proceedings*, volume 5721, page 174. Springer-Verlag New York Inc, 2009.

Aichert, I. and W. Ziegler. Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88(1):148–159, 2004.

Aichert, I. and W. Ziegler. Learning a syllable from its parts: Cross-syllabic generalisation effects in patients with apraxia of speech. *Aphasiology*, 22(11): 1216–1229, 2008.

Aichert, I. and W. Ziegler. Influence of word stress in patients with apraxia of speech. In *Academy of Aphasia 47th Annual Meeting, Boston, MA*, 2009.

Aichert, I., C. Marquardt, and W. Ziegler. Frequenzen sublexikalischer Einheiten des Deutschen: CELEX-basierte Datenbanken. *Neurolinguistik*, 19: 55–81, 2005.

Alm, P.A. *On the causal mechanisms of stuttering*. Department of Clinical Neuroscience, Lund University,, 2005.

Alm, Per A. Stuttering and the basal ganglia circuits: a critical review of possible relations. *J Commun Disord*, 37(4):325–69, 2004.

Anderson, Julie D. Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *J Speech Lang Hear Res*, 50(1):229–47, Feb 2007.

Antipova, Elena A, Suzanne C Purdy, Marjorie Blakeley, and Shelley Williams. Effects of altered auditory feedback (AAF) on stuttering frequency during monologue speech production. *J Fluency Disord*, 33(4):274–90, Dec 2008.

Archibald, L. and S.E. Gathercole. Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49(5):970, 2006.

Archibald, L.M.D., S.E. Gathercole, and M.F. Joanisse. Multisyllabic nonwords: More than a string of syllables. *The Journal of the Acoustical Society of America*, 125:1712, 2009.

Armson, J. and A. Stuart. Effect of extended exposure to frequency-altered feedback on stuttering during reading and monologue. *Journal of Speech, Language, and Hearing Research*, 41(3):479, 1998.

Armson, J., M. Kiefte, J. Mason, and D. De Croos. The effect of SpeechEasy on stuttering frequency in laboratory conditions. *Journal of fluency disorders*, 31(2):137–152, 2006.

Azrin, N., RJ Jones, and B. Flye. A synchronization effect and its application to stuttering by a portable apparatus. *Journal of applied behavior analysis*, 1(4):283, 1968.

Baayen, RH, R. Piepenbrock, and L. Gulikers. CELEX2 Linguistic Data Consortium, 1996.

Bakker, K. Technical Support for Stuttering Treatment. *Current issues in stuttering research and practice*, page 205, 2006.

Bartlett, S. Discriminative approach to automatic syllabification. Master's thesis, DepartmentofComputing Science, University of Alberta., 2007.

Bartlett, S., G. Kondrak, and C. Cherry. Automatic syllabification with structured SVMs for letter-to-phoneme conversion. *Proceedings of ACL-08: HLT*, pages 568–576, 2008.

Bates, R., M. Donegan, HO Istance, JP Hansen, and K.J. Räihä. Introducing COGAIN: communication by gaze interaction. *Universal Access in the Information Society*, 6(2):159–166, 2007.

Bishop, D.V.M. Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics*, 114(1):56–63, 2002.

Black, Rolf, Joseph Reddington, Ehud Reiter, Nava Tintarev, and Annalu Waller. Using NLG and Sensors to Support Personal Narrative for Children with Complex Communication Needs. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 1–9, Los Angeles, California, June 2010. Association for Computational Linguistics.

Blaine, T. and T. Perkis. The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 165–173. ACM, 2000.

Block, Susan, Mark Onslow, Rachael Roberts, and Samantha White. Control of stuttering with EMG feedback. *International Journal of Speech-Language Pathology*, 6(2):100–106, 2004.

Booij, G. Cliticization as prosodic integration: the case of Dutch. *The Linguistic Review*, 13(3-4):219–242, 1996.

Brendel, B. and W. Ziegler. Effectiveness of metrical pacing in the treatment of apraxia of speech. *Aphasiology*, 22(1):77–102, 2008.

Brundage, S.B. Virtual reality augmentation for functional assessment and treatment of stuttering. *Topics in language disorders*, 27(3):254, 2007.

Buhmann, J., J. Caspers, V. van Heuven, H. Hoekstra, J.P. Martens, and M. Swerts. Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In *Proceedings of LREC 2002*, pages 779–785, 2002.

Bulyko, I. and M. Ostendorf. Joint prosody prediction and unit selection for concatenative speech synthesis. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 2, pages 781–784. IEEE, 2001.

Campione, E. and J. Véronis. A large-scale multilingual study of silent pause duration. In *Speech Prosody 2002, International Conference*, 2002.

Caruso, A.J. and E.A. Strand. Motor speech disorders in children: Definitions, background, and a theoretical framework. *Clinical management of motor speech disorders in children*, pages 1–27, 1999.

Cholin, J. The mental syllabary in speech production: An integration of different approaches and domains. *Aphasiology*, 22(11):1127–1141, 2008.

Cholin, J., N.O. Schiller, and W.J.M. Levelt. The preparation of syllables in speech production. *Journal of Memory and Language*, 50(1):47–61, 2004.

Cholin, J., W.J.M. Levelt, and N.O. Schiller. Effects of syllable frequency in speech production. *Cognition*, 99(2):205–235, 2006.

Collins, N. A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions. In *Proceedings of AES118 Convention*. Citeseer, 2005.

Cordova, D.I. and M.R. Lepper. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88(4):715, 1996.

Corriveau, K.H. and U. Goswami. Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, 45(1): 119–130, 2009.

Cucchiarini, C., D. Lembrechts, and H. Strik. HLT and communicative disabilities: The need for co-operation between government, industry and academia. *Proceedings of LangTech-2008, Rome*, pages 125–128, 2008.

Cummins, F. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1):16–28, 2009.

Daelemans, W. and A. van den Bosch. Generalization performance of backpropagation learning on a syllabification task. In *Proceedings of the 3rd Twente Workshop on Language Technology*, pages 27–38. Citeseer, 1992.

de Jong, N.H. and T. Wempe. Automatic measurement of speech rate in spoken Dutch. *ACLC Working Papers*, 2(2):49–58, 2007.

Dell, G.S. A spreading-activation theory of retrieval in sentence production. *Psychological review*, 93(3):283, 1986.

Dell, G.S. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and language*, 27(2):124–142, 1988.

Donker, A. and P. Reitsma. Usability testing with young children. In *Proceedings of the 2004 conference on Interaction design and children: building a community*, pages 43–48. ACM, 2004.

Duffy, J.R. *Motor speech disorders: Substrates, differential diagnosis, and management*. Elsevier Mosby St Louis, MO, 2005.

Edwards, J., M.E. Beckman, and B. Munson. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2):421, 2004.

Fant, G., A. Kruckenberg, and J. Barbosa-Ferreira. Individual variations in pausing. A study of read speech. In *Proceedings of Fonetik*, pages 193–196. Citeseer, 2003.

Foundas, A L, A M Bollich, J Feldman, D M Corey, M Hurley, L C Lemen, and K M Heilman. Aberrant auditory processing and atypical planum temporale in developmental stuttering. *Neurology*, 63(9):1640–6, Nov 2004.

Fox, P.T., R.J. Ingham, J.C. Ingham, F. Zamarripa, J.H. Xiong, and J.L. Lancaster. Brain correlates of stuttering and syllable production. *Brain*, 123 (10):1985, 2000.

Goldiamond, I. Stuttering and fluency as manipulatable operant response classes. *Research in behavior modification*, pages 106–156, 1965.

Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, 2010.

Gould, J.D. and C. Lewis. Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311, 1985.

Graham, C. and M.S. Rosenthal. *Grammarchants: More jazz chants*. Oxford University Press, 1993.

Graybiel, AM and E. Saka. The basal ganglia and the control of action. *The cognitive neurosciences*, 3:495–510, 2004.

Green, CS and D. Bavelier. Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, 23(4):692, 2008.

Greenberg, S. On the origins of speech intelligibility in the real world. In *Robust Speech Recognition for Unknown Communication Channels*, 1997.

Greenberg, S. Speaking in shorthand-A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 29(2-4):159–176, 1999.

Greenberg, S. Pronunciation variation is key to understanding spoken language. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, pages 219–222, 2003.

Greenberg, S., H. Carvey, L. Hitchcock, and S. Chang. Beyond the phoneme: A juncture-accent model of spoken language. In *Proceedings of the second international conference on human language technology research*, pages 36–43. Morgan Kaufmann Publishers Inc., 2002.

Greenberg, S., H. Carvey, L. Hitchcock, and S. Chang. The phonetic patterning of spontaneous American English discourse. In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 2003.

Grosser, J., U. Natke, S. Langefeld, and K.T. Kalveram. Reduction in stuttering by delayed and frequency shifeted auditory feedback: effect of adaptation and sex differences. *J. Fluency Dis*, pages 252–252, 2001.

Guenther, F.H. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3): 594, 1995.

Guenther, F.H. Neural control of speech movements. *Phonetics and phonology in language comprehension and production: Differences and similarities*, pages 209–239, 2003.

Hersh, M.A. The design and evaluation of Assistive Technology products and devices Part 1: Design. *International Encyclopedia of Rehabilitation. Available online: http://cirrie. buffalo. edu/encyclopedia/en/article/309*, 2010.

Hourcade, J.P. Interaction design and children. *Foundations and Trends in Human-Computer Interaction*, 1(4):277–392, 2008.

Howell, P and A Archer. Susceptibility to the effects of delayed auditory feedback. *Percept Psychophys*, 36(3):296–302, Sep 1984.

Howell, P. and N. El-Yaniv. The effects of presenting a click in syllable-initial position on the speech of stutterers: Comparison with a metronome click. *Journal of fluency disorders*, 12(4):249–256, 1987.

Howell, P., N. El-Yaniv, and D.J. Powell. Factors affecting fluency in stutterers when speaking under altered auditory feedback. *Speech Motor Dynamics in Stuttering. New York: Springer Press.*, pages 361–369, 1987.

Howell, P., J. Au-Yeung, and L. Rustin. Clock and motor variances in lip-tracking: A comparison between children who stutter and those who do not. *Speech Production: Motor Control, Brain Research and Fluency Disorders.*, pages 573–578., 1997.

Howell, Peter. Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Res*, 1(1):31–46, Apr 2004.

Howell, Peter, Stephen Davis, and Sheila M Williams. Auditory abilities of speakers who persisted, or recovered, from stuttering. *J Fluency Disord*, 31 (4):257–70, 2006.

Hubbard, P. Evaluating CALL software. *Calling on CALL: From theory and research to new directions in foreign language teaching*, pages 313–318, 2006.

Ingham, R.J., R.A. Moglia, P. Frank, J.C. Ingham, and A.K. Cordes. Experimental investigation of the effects of frequency-altered auditory feedback on the speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 40(2):361, 1997.

Ingham, R.J., M. Kilgo, J.C. Ingham, R. Moglia, H. Belknap, and T. Sanchez. Evaluation of a stuttering treatment based on reduction of short phonation intervals. *Journal of Speech, Language, and Hearing Research*, 44(6):1229, 2001.

Janker, P.M. Evidence for the P-center syllable-nucleus-onset correspondence hypothesis. *ZAS Paper in Linguistics*, 7:94–124, 1996.

Jayaram, M. Distribution of stuttering in sentences: relationship to sentence length and clause position. *J Speech Hear Res*, 27(3):338–41, Sep 1984.

Kahn, H., T. Stannard, and J. Skinner. The use of words versus nonwords in the treatment of apraxia of speech: A case study. *ASHA Special Interest Division 2: Neurophysiology and Neurogenic Speech and Language Disorders*, 8(3):5–10, 1998.

Kalinowski, J. and T. Saltuklaroglu. Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. *Neuroscience & Biobehavioral Reviews*, 27(4):339–347, 2003.

Kalveram, K.T. Neurobiology of speaking and stuttering. In *Fluency Disorders: Theory, Research, Treatment and Self-help. Proceedings of the Third World Congress of Fluency Disorders*, pages 59–65, 2001.

Kent, R.D. Research on speech motor control and its disorders:: A review and prospective. *Journal of Communication disorders*, 33(5):391–428, 2000.

Kim, M. and C.M. Tomaino. Protocol evaluation for effective music therapy for persons with nonfluent aphasia. *Topics in stroke rehabilitation*, 15(6): 555–569, 2008.

Kim, Y.J. and M.C. Beutnagel. Automatic assessment of american english lexical stress using machine learning algorithms. In *Speech and Language Technology in Education*, 2011.

Kochanski, G. and C. Orphanidou. What marks the beat of speech? *The Journal of the Acoustical Society of America*, 123:2780, 2008.

Kochanski, G., E. Grabe, J. Coleman, and B. Rosner. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118:1038, 2005.

Koike, K.J.M. and C.W. Asp. Tennessee test of rhythm and intonation patterns. *Journal of Speech and Hearing Disorders*, 46(1):81, 1981.

Koopmans, M., I. Slis, and T. Rietveld. The influence of word position and word type on the incidence of stuttering. *Speech motor control and stuttering*, pages 333–340, 1991.

Kuniszyk-Jóźkowiak, W., E. Smołka, and B. Adamczyk. Effect of acoustical, visual and tactile reverberation on speech fluency of stutterers. *Folia phoniatrica et logopaedica*, 49(1):26–34, 1997.

Laganaro, M. Is there a syllable frequency effect in aphasia or in apraxia of speech or both? *Aphasiology*, 22(11):1191–1200, 2008.

Levelt, WJ. Models of word production. *Trends in Cognitive Sciences*, 3(6): 223–232, 1999.

Levelt, W.J.M. and L. Wheeldon. Do speakers have access to a mental syllabary? *Cognition*, 50(1-3):239–269, 1994.

Light, J. and P. Lindsay. Cognitive science and augmentative and alternative communication. *Augmentative and alternative communication*, 7(3):186–203, 1991.

Lin, H.Y. and J. Fon. Perception of temporal cues at discourse boundaries. In *Tenth Annual Conference of the International Speech Communication Association*, 2009.

Lincoln, Michelle and Clare Walker. A survey of Australian adult users of altered auditory feedback devices for stuttering: use patterns, perceived effectiveness and satisfaction. *Disability and rehabilitation*, 29(19):1510–7, Oct 2007.

Lincoln, Michelle, Ann Packman, and Mark Onslow. Altered auditory feedback and the treatment of stuttering: a review. *J Fluency Disord*, 31(2):71–89, 2006.

Lincoln, Michelle, Ann Packman, Mark Onslow, and Mark Jones. An experimental investigation of the effect of altered auditory feedback on the conversational speech of adults who stutter. *J Speech Lang Hear Res*, 53(5): 1122–31, Oct 2010.

Ludlow, C.L. and A. Braun. Research evaluating the use of neuropharmacological agents for treating stuttering: Possibilities and problems. *Journal of Fluency Disorders*, 1993.

Maas, Edwin, Donald A Robin, Shannon N Austermann Hula, Skott E Freedman, Gabriele Wulf, Kirrie J Ballard, and Richard A Schmidt. Principles of motor learning in treatment of motor speech disorders. *Am J Speech Lang Pathol*, 17(3):277–98, Aug 2008.

Marchand, Y., C.R. Adsett, and R.I. Damper. Automatic syllabification in English: A comparison of different algorithms. *Language and Speech*, 52(1): 1, 2009.

Marcus, S.M. Acoustic determinants of perceptual center (P-center) location. *Attention, Perception, & Psychophysics*, 30(3):247–256, 1981.

Martin, R. and S.K. Haroldson. Effects of five experimental treatments on stuttering. *Journal of Speech and Hearing Research*, 22(1):132, 1979.

Maryn, Y, M De Bodt, and P Van Cauwenberge. Effects of biofeedback in phonatory disorders and phonatory performance: a systematic literature review. *Appl Psychophysiol Biofeedback*, 31(1):65–83, Mar 2006.

Mauszycki, S.C. and J.L. Wambaugh. The effects of rate control treatment on consonant production accuracy in mild apraxia of speech. *Aphasiology*, 22 (7-8):906–920, 2008.

Max, L. Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. *Speech motor control in normal and disordered speech*, pages 357–387, 2004.

McNeil, M.R. *Clinical management of sensorimotor speech disorders*. Thieme Medical Pub, 2008.

McNeil, M.R., S.R. Pratt, T.R.D. Fossett, B. Maassen, R. Kent, H. Peters, P. van Lieshout, and W. Hulstijn. The differential diagnosis of apraxia of speech. *Speech motor control in normal and disordered speech*, pages 389–413, 2004.

McNeill, BC, GT Gillon, and B. Dodd. Phonological awareness and early reading development in childhood apraxia of speech (CAS). *International Journal of Language & Communication Disorders*, 44(2):175–192, 2009.

Mitterer, H. How are words reduced in spontaneous speech? *Experimental Linguistics ExLing 2008*, page 165, 2008.

Moriarty, B.C. and G.T. Gillon. Phonological awareness intervention for children with childhood apraxia of speech. *International Journal of Language & Communication Disorders*, 41(6):713–734, 2006.

Morton, J., S. Marcus, and C. Frankish. Perceptual centers (P-centers). *Psychological Review*, 83(5):405, 1976.

Nagata, R., T. Marueki, K. Funakoshi, T. Kitamura, and M. Nakano. A method for predicting stressed words in English Jazz Chants. In *Speech and Language Technology in Education*, 2011.

Namasivayam, A.K. and P. van Lieshout. Investigating speech motor practice and learning in people who stutter. *Journal of fluency disorders*, 33(1):32–51, 2008.

Natke, U. and K.T. Kalveram. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of speech, language, and hearing research*, 44(3):577, 2001.

Natke, U., J. Grosser, and K.T. Kalveram. Fluency, fundamental frequency, and speech rate under frequency-shifted auditory feedback in stuttering and nonstuttering persons. *Journal of fluency disorders*, 26(3):227–241, 2001.

Natke, U., J. Grosser, P. Sandrieser, and K.T. Kalveram. The duration component of the stress effect in stuttering. *Journal of fluency disorders*, 27(4): 305–318, 2002.

Odell, K., M.R. McNeil, J.C. Rosenbek, and L. Hunter. Perceptual characteristics of consonant production by apraxic speakers. *Journal of Speech and Hearing Disorders*, 55(2):345, 1990.

Odell, K.H. Considerations in target selection in apraxia of speech treatment. In *Seminars in Speech and Language*, pages 309–324, 2002.

Olander, L., A. Smith, and H.N. Zelaznik. Evidence that a motor timing deficit is a factor in the development of stuttering. *Journal of Speech, Language, and Hearing Research*, 53(4):876, 2010.

Onslow, M., C. Andrews, and M. Lincoln. A control/experimental trial of an operant treatment for early stuttering. *Journal of Speech and Hearing Research*, 37(6):1244, 1994.

Öster, A.M. Computer-based speech therapy using visual feedback with focus on children with profound hearing impairments. In *Doctoral Thesis, KTH School of Computer Science and Communication, Stockholm.*, 2006.

Paris, C.R., M.H. Thomas, R.D. Gilson, and J.P. Kincaid. Linguistic cues and memory for synthetic and natural speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(3):421–431, 2000.

Peter, Beate and Carol Stoel-Gammon. Central timing deficits in subtypes of primary speech disorders. *Clin Linguist Phon*, 22(3):171–98, Mar 2008.

Peters, H.F.M., W. Hulstijn, and P.H.H.M. Van Lieshout. Recent developments in speech motor research into stuttering. *Folia Phoniatrica et Logopaedica*, 52(1-3):103–119, 2000.

Peterson, C. Bringing ADDIE to life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia*, 12(3), 2003.

Pilon, M.A., K.W. McIntosh, and M.H. Thaut. Auditory vs visual speech timing cues as external rate control to enhance verbal intelligibility in mixed spastic ataxic dysarthric speakers: a pilot study. *Brain Injury*, 12(9):793–803, 1998.

Pino, A., G. Kouroupetroglou, H. Kacorri, A. Sarantidou, and D. Spiliotopoulos. An open source/freeware assistive technology software inventory. *Computers Helping People with Special Needs*, pages 178–185, 2010.

Pollard, R., J.B. Ellis, D. Finan, and P.R. Ramig. Effects of the SpeechEasy on objective and perceived aspects of stuttering: a 6-month, phase I clinical trial in naturalistic environments. *Journal of Speech, Language, and Hearing Research*, 52(2):516, 2009.

Quarrington, B. Stuttering as a function of the information value and sentence position of words. *Journal of abnormal psychology*, 70(3):221, 1965.

Ratner, N.B. Translating recent research into meaningful clinical practice. In *Seminars in speech and language*, pages 236–249. Thieme, 2010.

Riecker, A., B. Brendel, W. Ziegler, M. Erb, and H. Ackermann. The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and language*, 107(2):102–113, 2008.

Riegel, I. and S. Buedel. Hierarchically Organized Speech Materials for AOS Treatment. *Sprache, Stimme, Gehör*, 34(03):149–150, 2010.

Riley, G D and J C Ingham. Acoustic duration changes associated with two types of treatment for children who stutter. *J Speech Lang Hear Res*, 43(4): 965–78, Aug 2000.

Roark, Brian, Jacques de Villiers, Christopher Gibbons, and Melanie Fried-Oken. Scanning methods and language modeling for binary switch typing. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 28–36, Los Angeles, California, June 2010. Association for Computational Linguistics.

Roelofs, A. WEAVER++ and other computational models of lemma retrieval and word-form encoding. *Aspects of language production*, pages 71–114, 2000.

Roy, P. and S. Chiat. A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47(1):223, 2004.

Ruiter, Marina B., Toni C. M. Rietveld, Catia Cucchiarini, Emiel J. Krahmer, and Helmer Strik. Human language technology and communicative disabilities: requirements and possibilities for the future. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), 2010.

Ryan, R.M. and E.L. Deci. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.

Sadakata, M., D. Hoppe, A. Brandmeyer, R. Timmers, and P. Desain. Real-time visual feedback for learning to perform short rhythms with expressive variations in timing and loudness. *Journal of New Music Research*, 37(3): 207–220, 2008.

Sahlen, B., C. Reuterskiold-Wagner, U. Nettelbladt, and K. Radeborg. Non-word repetition In children with language impairment-pitfalls and possibilities. *International Journal of Language & Communication Disorders*, 34(3): 337–352, 1999.

Salmelin, R., A. Schnitzler, F. Schmitz, L. Jäncke, OW Witte, and H.J. Freund. Functional organization of the auditory cortex is different in stutterers and fluent speakers. *Neuroreport*, 9(10):2225, 1998.

Saltuklaroglu, T., J. Kalinowski, M. Robbins, S. Crawcour, and A. Bowers. Comparisons of stuttering frequency during and after speech initiation in unaltered feedback, altered auditory feedback and choral speech conditions. *International Journal of Language & Communication Disorders*, 44(6):1000–1017, 2009.

Sandlund, M., K. Hoshi, E.L. Waterworth, and C. Häger-Ross. A conceptual framework for design of interactive computer play in rehabilitation of children with sensorimotor disorders. *Physical Therapy Reviews*, 14(5):348–354, 2009a.

Sandlund, M., S. McDonough, and C. Häger-Ross. Interactive computer play in rehabilitation of children with sensorimotor disorders: a systematic review. *Developmental Medicine & Child Neurology*, 51(3):173–179, 2009b.

Schiller, N.O. *The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography*. Sociale Wetenschappen: Nijmegen: 1997., 1997.

Schiller, N.O., A.S. Meyer, R.H. Baayen, and W.J.M. Levelt. A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, 3 (1):8–28, 1996.

Schiller, N.O., A. Costa, and A. Colomé. Phonological encoding of single words: In search of the lost syllable. *C. Gussenhoven, & N. Warner, Papers in Laboratory Phonology*, 7:35–59, 2002.

Schmid, H., B. Möbius, and J. Weidenka . Tagging syllable boundaries with joint N-gram models. In *Proceedings of Interspeech*. Citeseer, 2007.

Schmidt, R.A. and T.D. Lee. *Motor control and learning: A behavioral emphasis*. Human Kinetics Publishers, 2005.

Schmidt, R.A. and C.A. Wrisberg. *Motor learning and performance: a situation-based learning approach*. Human Kinetics Publishers, 2008.

Schoor, A., I. Aichert, and W. Ziegler. A motor learning perspective on phonetic syllable kinships: How training effects transfer from learned to new syllables in severe apraxia of speech. *Aphasiology*, 2012.

Schul, Rina, Joan Stiles, Beverly Wulfeck, and Jeanne Townsend. How 'generalized' is the 'slowed processing' in SLI? The case of visuospatial attentional orienting. *Neuropsychologia*, 42(5):661–71, 2004.

Shaiman, S. and V.L. Gracco. Task-specific sensorimotor interactions in speech production. *Experimental brain research*, 146(4):411–418, 2002.

Sharma, Harsh Vardhan and Mark Hasegawa-Johnson. State-Transition Interpolation and MAP Adaptation for HMM-based Dysarthric Speech Recognition. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 72–79, Los Angeles, California, June 2010. Association for Computational Linguistics.

Shriberg, L.D., T.F. Campbell, H.B. Karlsson, R.L. Brown, J.L. Mcsweeny, and C.J. Nadler. A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. *Clinical linguistics & phonetics*, 17(7):549–574, 2003.

Silipo, R. and S. Greenberg. Automatic transcription of prosodic stress for spontaneous English discourse. In *Proc. of the XIVth International Congress of Phonetic Sciences (ICPhS)*, volume 3, page 2351, 1999.

Smith, A. and L. Goffman. Interaction of motor and language factors in the development of speech production. *Speech motor control in normal and disordered speech*, pages 225–252, 2004.

Smith, A. and J. Kleinow. Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. *Journal of Speech, Language, and Hearing Research*, 43(2):521, 2000.

Smith, A. and H.N. Zelaznik. Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1):22–33, 2004.

Staiger, A. and W. Ziegler. Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology*, 22(11):1201–1215, 2008.

Stuart, Andrew, Shixiong Xia, Yining Jiang, Tao Jiang, Joseph Kalinowski, and Michael P Rastatter. Self-contained in-the-ear device to deliver altered auditory feedback: applications for stuttering. *Ann Biomed Eng*, 31(2):233–7, Feb 2003.

Swerts, M. and R. Geluykens. The prosody of information units in spontaneous monologue. *Phonetica*, 50(3):189–196, 1993.

Tellis, G.M. *Efficacy of the Computer-Aided Fluency Establishment Trainer in the treatment of stuttering*. PhD thesis, California State University, 1996.

Thaut, M.H., K.W. McIntosh, G.C. McIntosh, and V. Hoemberg. Auditory rhythmicity enhances movement and speech motor control in patients with Parkinson's disease. *Functional Neurology*, 16(2):163–172, 2001.

Thomson, J.M. and U. Goswami. Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, 102(1):120–129, 2008.

Tian, J. Data-driven approaches for automatic detection of syllable boundaries. In *Proceedings of 8th ICSLP*. Citeseer, 2004.

Tjaden, K. Exploration of a treatment technique for prosodic disturbance following stroke. *clinical linguistics & phonetics*, 14(8):619–641, 2000.

Tourville, J.A. and F.H. Guenther. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7):952–981, 2011.

Trajkovski, N., C. Andrews, M. Onslow, S. O'Brian, A. Packman, and R. Menzies. A phase II trial of the Westmead Program: Syllable-timed speech treatment for pre-school children who stutter. *International Journal of Speech-Language Pathology*, 13(6):500–509, 2011.

Van Borsel, John, Gert Reunes, and Nathalie Van den Bergh. Delayed auditory feedback in the treatment of stuttering: clients as consumers. *Int J Lang Commun Disord*, 38(2):119–29, 2003.

Van Borsel, John, Sarah Sierens, and Monica Medeiros de Britto Pereira. Using delayed auditory feedback in the treatment of stuttering: evidence to consider. *Pro Fono*, 19(3):323–32, 2007.

Van der Merwe, A. A theoretical framework for the characterization of pathological speech sensorimotor control. *Clinical management of sensorimotor speech disorders*, pages 1–25, 1997.

Van Lieshout, P., W. Hulstijn, and H.F.M. Peters. Searching for the weak link in the speech production chain of people who stutter: a motor skill approach. *Speech motor control in normal and disordered speech*, pages 313–355, 2004.

Vaquero, C., O. Saz, E. Lleida, J.M. Marcos, C. Canalís, and C.P. de Educación. VOCALIZA: An application for computer-aided speech therapy in spanish language. *IV jornadas en tecnologías del habla, Zaragoza, España*, pages 321–326, 2006.

Vicsi, K., P. Roach, A.M. Öster, Z. Kacic, F. Csatári, A. Sfakianaki, R. Veronik, and G. Gordos. A multilingual, multimodal, speech training system, SPECO. In *Seventh European Conference on Speech Communication and Technology*, 2001.

Vitevitch, M.S. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4):735, 2002.

Vroomen, J. and B. Gelder. Speech segmentation in Dutch: No role for the syllable. In *Third International Conference on Spoken Language Processing*, 1994.

Wagner, A. Acoustic Cues for Automatic Determination of Phrasing. In *Speech Prosody 2010-Fifth International Conference*, 2010.

Wan, C.Y., T. Rüber, A. Hohmann, and G. Schlaug. The therapeutic effects of singing in neurological disorders. *Music perception*, 27(4):287, 2010.

Wan, C.Y., L. Bazen, R. Baars, A. Libenson, L. Zipse, J. Zuk, A. Norton, and G. Schlaug. Auditory-Motor Mapping Training as an intervention to facilitate speech output in non-verbal children with autism: a proof of concept study. *PloS one*, 6(9):e25505, 2011.

Watkins, Kate E, Stephen M Smith, Steve Davis, and Peter Howell. Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, 131(Pt 1):50–9, Jan 2008.

Werner, S., M. Wolff, M. Eichner, and R. Hoffmann. Modeling pronunciation variation for spontaneous speech synthesis. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 1, pages I–673. IEEE, 2004.

Westphal, G. An experimental study of certain motor abilities of stutterers. *Child Development*, 4(3):214–221, 1933.

Wingate, M.E. and P. Howell. Foundations of stuttering. *The Journal of the Acoustical Society of America*, 112:1229, 2002.

Wolff, P.H. Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, 15(1):179–206, 2002.

Wu, S.L., M.L. Shire, S. Greenberg, and N. Morgan. Integrating syllable boundary information into speech recognition. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 2, pages 987–990. IEEE, 1997.

Wulf, G. *Attention and motor skill learning*. Human Kinetics Publishers, 2007.

Yairi, E. Subtyping stuttering I: a review. *Journal of fluency disorders*, 32(3): 165–196, 2007.

Yaruss, JS. and K.J. Logan. Evaluating rate, accuracy, and fluency of young children's diadochokinetic productions: a preliminary investigation. *Journal of fluency disorders*, 27(1):65–86, 2002.

Yoon, T.J., J. Cole, and M. Hasegawa-Johnson. On the edge: Acoustic cues to layered prosodic domains. In *Proceedings of ICPhS*, 2007.

Zatorre, R.J., J.L. Chen, and V.B. Penhune. When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7):547–558, 2007.

Ziegler, P.D.D.W. Modelling the architecture of phonetic plans: Evidence from apraxia of speech. *Language and Cognitive Processes*, 24(5):631–661, 2009.

Ziegler, W. Apraxia of speech. *Handbook of clinical neurology*, 88:269–285, 2008.

Zimmerman, S., J. Kalinowski, A. Stuart, and M. Rastatter. Effect of altered auditory feedback on people who stutter during scripted telephone conversations. *Journal of Speech, Language, and Hearing Research*, 40(5):1130, 1997.

Samenvatting

Voor de behandeling van patiënten met communicatie- stoornissen maken spraak taal therapeuten gebruik van een variëteit aan methoden en technieken , waarvan een aanzienlijk deel is gebaseerd op technologische hulpmiddelen. Dit proefschrift onderzoekt een aantal manieren waarop de behandeling van een vorm van communicatie-stoornissen, namelijk *motorische spraakstoornissen* (Motor Speech Disorders - dysartrie, apraxie, en stotteren) kan worden ondersteund met behulp van bestaande technologieen. Hierbij wordt uitgegaan van twee elkaar aanvullende perspectieven: het eerste belicht de vraag hoe nieuwe computergestuurde toepassingen de huidige behandeltechnieken van motorische spraakstoornissen kunnen verfijnen. Het tweede, bredere perspectief van dit proefschrift gaat over de vraag welke ingrediënten noodzakelijk zijn voor een duurzaam en interdisciplinair proces van innovatie op het gebied van de behandeling van communicatie-stoornissen. Om de uitdagingen en ontwikkelingskansen van de technologie-gebaseerde innovatie te onderzoeken, richt het werk in dit proefschrift zich op de volgende drie belangrijke fasen in de behandeling van motorische spraakstoornissen:

Voorbereiding van behandelprogramma's. In de eerste stadia van de behandeling moeten logopedisten de behandeling ontwerpen en plannen op basis van de behoeften van individuele patiënten. Binnen een behandelprogramma moeten linguïstische items worden geselecteerd die het meest geschikt zijn voor de aanpak van de relevante aspecten van de spraakstoornis. De technische ondersteuning van logopedisten met de selectie van behandeldoelen bestaat uit twee delen - het creeren van relevante linguïstische materialen, en het verstrekken van middelen voor een effectieve generatie van spraakoefening items.

In hoofdstuk 3 van dit proefschrift gaan we in op de eerste stap, en laten we zien hoe technieken uit de computationele taalkunde kunnen worden aangewend voor het genereren van ondersteunende taalkundige materialen. We bespreken het belang van het inventariseren van de lettergrepen van een taal, en motiveren het voordeel van het gebruik van de gesproken woordvorm, in tegenstelling tot de kanonieke uitspraak van een woord. Vervolgens berekenen we de syllabische inventarisatie van de Nederlandse taal met behulp van een computer-algorithme voor de automatische indeling van woorden afkomstig uit de gesproken taal, afkomstig van de Corpus Gesproken Nederlands (CGN).

In Hoofdstuk 4 wordt beschreven hoe theoretische kennis op het gebied van Apraxie van de Spraak (AOS) kan worden toegepast bij de ontwikkeling van een nieuwe computer-applicatie, die het mogelijk maakt om op de patient afgestemde behandelingsmaterialen te genereren. Gemotiveerd door bestaand onderzoek, dat aantoont dat een systematische manipulatie van de syllabische parameters in spraakdoelstellingen voordelig is voor de behandeling van AOS, wordt een computer-applicatie ontwikkeld voor het genereren van klinisch relevante woordenlijsten. De computer-applicatie is ontworpen om logopedisten een flexibele, efficiënte en intuïtieve werkflow te bieden. Een onderzoek naar de bruikbaarheid van de computer-applicatie leverde zeer positieve reacties op. Logopedisten beschouwden het programma als een nuttig en tijdbesparend middel voor de creatie van behandelprogramma's.

Leren en oefenen van spraak-motorische vaardigheden. Tijdens het behandelingsproces wordt veel tijd besteed aan het oefenen van spraak-motorische vaardigheden. De uitdaging voor de logopedist is om de patiënten te betrekken bij het oefenproces door middel van een systematische demonstratie van de gewenste spraakdoelstellingen, het geven van adequate feedback, en het stimuleren van patiënten om een grote hoeveelheid oefeningen uit te voeren. Om deze uitdagingen het hoofd te bieden, onderzoeken we in hoofdstuk 5 de mogelijkheden van het gebruik van een computerspel om de oefening van spraaktiming vaardigheden van kinderen met motorische spraakstoornissen te ondersteunen. Hiervoor proberen wij te bepalen welke elementen nodig zijn voor het samenstellen van een training-systeem, wat betreft de methodologie van de oefening, de vereiste technologie, en de bruikbaarheidsfactoren. We implementeren spraak-timing oefeningen binnen een computerspel, en evalueren de toepasbaarheid van het spel als een aanvullend therapeutisch middel.

De resultaten suggereren dat interactieve spellen voor het oefenen van spraaktiming vaardigheden de positieve ervaring van spraak-motorische training kunnen bevorderen en de motivatie om te oefenen kunnen versterken, althans op de korte termijn. De interactie met het spel droeg bij aan de oefensituatie, en werd door kinderen en logopedisten zeer gewaardeerd. Nochtans werd het computerspel niet afwisselend genoeg bevonden om kinderen op de lange termijn te motiveren. Een mogelijke manier om deze uitdaging aan te gaan is door de *retro-fitting* van bestaande computerspellen, zodat ze gekoppeld kunnen worden aan auditieve input. Op deze manier kan het grote aanbod van commerciele computerspellen worden benut voor spraak-motorische trainingsdoeleinden.

Zelfstandige behering van communicatie. Na een behandelingsperiode kan het nodig zijn dat de patient de communicatie zelfstandig leert beheren. Een voorbeeld van een technologie ter ondersteuning van dit proces is het gebruik van Altered Auditory Feedback (AAF) apparaten door sommige volwassenen die stotteren, om de spreekvloeiendheid in hun dagelijkse communicatie te bevorderen. In hoofdstuk 6 stellen wij een verfijning van de huidige AAF procedures voor ter verbetering van spreekvloeiendheid. Daartoe wordt een *adaptieve* feedbackprocedure voorgesteld, die gebruikt maakt van digitale signaalverwerkingstechnieken om het spraaksignaal in real-time te analyseren, en de auditieve feedback dynamisch te kunnen activeren. De voorgestelde AAF procedure wordt geëvalueerd met volwassenen die stotteren, waarbij wordt gekeken naar de invloed op spraakvloeiendheid, maar ook naar de mate van comfort met de auditieve cues.

De in dit onderzoek gebruikte adaptieve feedback procedure leidde tot een sterke vermindering van onvloeiendheiden op zinsdeel-begin posities. De resultaten geven aan dat een audio-cue tijdens stille segmenten in bepaalde gevallen kan helpen om spraak vloeiend te hervatten. Bovendien suggereren de resultaten dat een nog hogere reductie kan worden bereikt als de audio-cue voortduurt voorbij het moment waar een uiting begint. Echter, wat betreft het algehele niveau van spraakvloeiendheid, leidde de adaptieve feedback procedure niet tot de grootste verbetering op groepsniveau. Uit een onderzoek van individuele resultaten bleek een grote mate van variabiliteit, hetgeen eerdere aanwijzingen bevestigt dat personen die stotteren verschillend reageren op AAF procedures.

De in dit proefschrift gepresenteerde voorstellen beogen de ontbrekende schakel te zijn tussen enerzijds de klinische vraag op het gebied van de spraak-taal therapie, en anderzijds het aanbod aan bestaande technieken uit technologische domeinen. Het succes van deze benadering wordt aangetoond door het feit dat verschillende klinische procedures met behulp van technologische innovatie ondersteund kunnen worden. Echter, het onderzoeksproces heeft ook belangrijke structurele uitdagingen bij het coördineren van technologische ontwikkelingen in het veld aan het licht gebracht. Onze ervaring heeft geleerd dat een alomvattend en duurzaam onderzoekskader nodig is om tot tastbare resultaten te komen, en resulteerde in een voorstel voor een nieuw model voor technologische innovatie op het gebied van spraak-taaltherapie.

Het uiteindelijke doel van een dergelijk kader is om een voortdurende productie van innovatieve oplossingen mogelijk te maken, waaruit therapeuten de middelen kunnen halen voor het avanceren van nieuwe technologie-ondersteunde therapieën. Het in dit proefschrift beschreven model definieert de deelnemende partijen, hun activiteiten, de informatie-uitwisseling en interactie tussen de partijen. Binnen het model wordt de identificatie van problemen en behoeften, alsmede het genereren van nieuwe concepten geplaatst binnen het gebied van spraak-taaltherapie zelf. Deze strategie zorgt ervoor dat de meest relevante problemen in het gebied worden aangepakt. Bovendien moeten bestaande samenwerkingsverbanden binnen het model ervoor zorgen dat meerdere ontwikkelings- en test-rondes kunnen worden uitgevoerd, op weg naar volwaardige klinische instrumenten.

Stellingen behorend bij het proefschrift

1. Enabling speech clinicians to structure linguistic materials in their treatment programs will facilitate the delivery of treatment based on current theoretical knowledge.
2. The frequency of various syllables in the Dutch language is exponentially distributed.
3. For the generation of monosyllabic speech treatment targets in Dutch, a syllabary derived from the canonical pronunciation of words can be seen as an appropriate material.
4. Interactive computer games for practicing speech timing skills can promote a positive experience of speech motor training and enhance motivation for practice.
5. Utilizing popular computer and video games for therapy purposes is likely to decrease the risk that gaming will be perceived as tedious, compared to custom-made games.
6. It is worthwhile to explore adaptive Altered Auditory Feedback (AAF) procedures, and devise methods to ‘tailor’ AAF procedures to the individual needs of people who stutter.
7. An audio cue provided during silent segments may assist some speakers who stutter to initiate speech fluently.
8. A dedicated framework is needed in order to promote a durable process of innovation in the field of communication disorders.
9. The role of technological support in the field of communication disorders is complementary, as speech therapists will always be central in the delivery of treatment.

10. Guard your roving thoughts with a jealous care, for speech is but the dialer of thoughts, and every fool can plainly read in your words what is the hour of your thoughts.