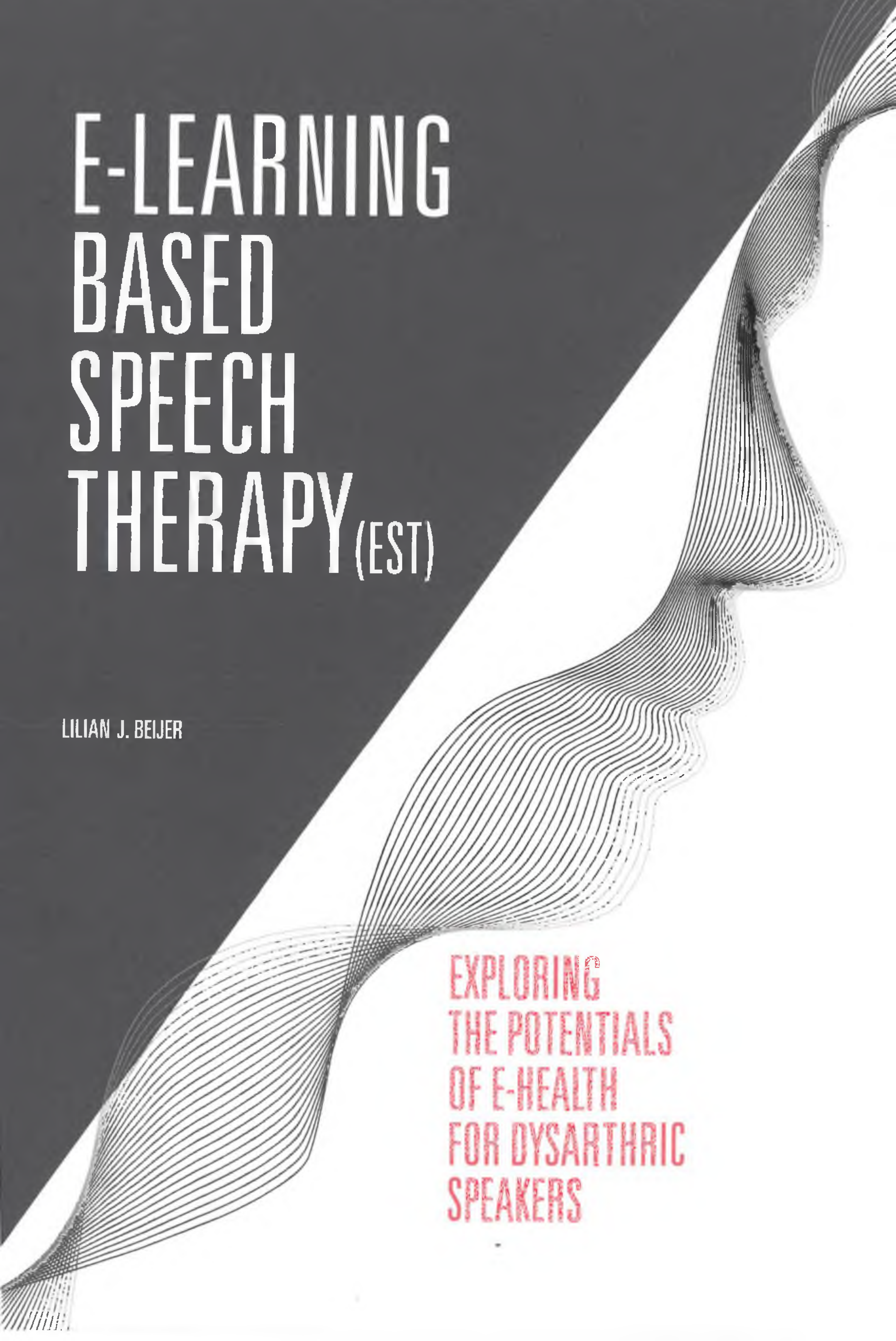
E-LEARNING BASED SPEECH THERAPY (EST)

LILIAN J. BEIJER



EXPLORING THE POTENTIALS OF E-HEALTH FOR DYSARTHRIC SPEAKERS

E-LEARNING BASED SPEECH THERAPY(EST)

LILIAN J. BEIJER

EXPLORING THE POTENTIALS OF E-HEALTH FOR DYSARTHRIC SPEAKERS

The publication of this thesis was financially supported by:
Sint Maartenskliniek
Stichting OrthoResearch Nijmegen
POM NIjmegen BV

The studies reported in this dissertation were supported by the Department of Research Development and Education of the Sint Maartenskliniek in Nijmegen, and were conducted in collaboration with the Department of Linguistics of the Radboud University and the Department of Rehabilitation of the University Medical Centre Nijmegen.

Cover design by Gaetano Ambrosino

Printed by Drukkerij Quickprint Nijmegen

ISBN 978-90-819826-0-3

E-LEARNING BASED SPEECH THERAPY (EST)

EXPLORING THE POTENTIALS OF E-HEALTH FOR DYSARTHRIC SPEAKERS

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op woensdag 21 november 2012
om 15.30 uur precies

door

Lilian Jeannette Beijer
geboren 23 maart 1965
te Heerlen

Promotoren: Prof. dr. A.C.M. Rietveld
Prof. dr. A.C.H. Geurts

Manuscriptcommissie: Prof. dr. ir. P.W.M. Desain (voorzitter)
Prof. dr. J.S. Rietman (Universiteit Twente)
Dr. R. L. Palmer (Sheffield University, UK)

Voor Huib,
die ons onvoorstelbaar veel liefde heeft gegeven.

TABLE OF CONTENTS

CHAPTER 1

General Introduction

The studies reported in this thesis explore the potentials of a web application for speech training in neurological patients with dysarthric speech: E-learning based Speech Therapy (EST). The introduction starts with the description of current developments in healthcare for neurological patients and the potential role of telehealth in this context. Particular attention will be paid to telerehabilitation, more specifically in the field of communication. EST will be introduced along with a rationale for the infrastructure and the content of the speech training. The EST research program will be presented, followed by an outline of the studies included in this thesis.

1.1 Rehabilitation

The World Health Organization describes rehabilitation for people with disabilities as a process aimed at enabling them to reach and maintain their optimal physical, sensory, intellectual, psychological and social functional levels. Rehabilitation provides disabled people with the tools they need to attain independence and self-determination (WHO, 2012). The field of rehabilitation currently faces a growing group of patients with acquired neurological diseases and disabilities due to ageing of the population and improved medical care (Brennan, Georgeadis, & Baron, 2002). As a consequence, an increasing number of chronic neurological patients suffer from motor, cognitive or communicative disabilities during a substantial part of their lives. Obviously, the activity limitations associated with these disabilities negatively affect the patients' quality of life. Healthcare providers therefore aim at diminishing these limitations, making efforts to enhance patients' independence and to optimize their participation in society. Being less dependent on the environment not only increases the quality of neurological patients' life, but also diminishes the need for expensive health care resources.

The International Classification of Functioning Disability and Health (ICF), as introduced by the World Health Organization (WHO, 2001), uses the 'performance' and 'capacity' qualifiers to describe a patient's functional status. 'Performance' is related to what a patient actually does in his own natural environment. In the context of enhancing participation, adapting neurological patients' environment (e.g. social or physical) is a pragmatic solution to improve their performance. However - to a certain extent - it keeps patients dependent on their environment. On the other hand, 'capacity' refers to a patient's ability to execute a task in a standardized environment, without the influence of environmental factors. It is a useful concept to indicate the highest probable level of functioning that a patient can reach in a particular functional domain. Therefore, improving patients' capacities genuinely increases their independence and may considerably contribute to enhanced participation, regardless of their einvironment. In motor and communicative rehabilitation, positive effects of intensive training on functional capacities are increasingly gaining interest (Bhogal, Teasell, & Speechley, 2003; Kwakkel, 2006; Rijntjes et al., 2009). The apparent benefits of intensification are in line with the fundamental principle of rehabilitation as expressed by Winters (2002): 'remodelling the tissues and systems of cells — ranging from connective soft tissue to muscle to neurocircuitry — on the basis of a sequence of inputs'. Unfortunately, there are two key obstacles to overcome before intensified training in neurological patients can be accomplished. First, intensive training puts considerable demands on therapeutic and financial resources. A second obstacle concerns the current ageing of our population. This demographic shift brings about a healthcare capacity problem, caused by an increasing shortage of clinical personnel (i.e. therapists) for the growing number of patients.

Given the indications for the beneficial effects of intensive training on functional capacities in neurological patients, telerehabilitation might provide a solution to the barrier caused by the financial and the healthcare capacity problem as described above.

1.2 Telerehabilitation

Telerehabilitation refers to the delivery of medical rehabilitation services and the support of independent living, using telehealth technologies (Rosen, 1999). The delivery of rehabilitation services over communication networks and the internet may contribute to the solutions of problems caused by the distorted 'need-and-availability-balance' in patients with acquired neurological diseases. Inspired by the increasing power of information technology, the spread of broadband internet access, and wireless connectivity over distances, clinicians are challenged to develop methods and tools for remote delivery of rehabilitation services and for supporting independent living in chronic neurological patients (Rosen, 2004). Telerehabilitation has developed from the parent disciplines of telemedicine and telehealth. Encompassing telemedicine, telehealth refers to the use of information and communication technology 'to support long-distance clinical health care, patient and professional health-related education, public health, and health administration' (Schmeler, Schein, & McCue, 2009). Telerehabilitation also provides a clear link to the field of telecare, which refers to the remote or enhanced delivery of health and social services to people in their own home by means of telecommunications and computerized systems (Hill, 2010).

Telerehabililtation is not intended to replace traditional rehabilitation services, but to enhance them. Telerehabilitation is not about 'technology' but about using technology to enhance and optimize rehabilitation services and patient outcomes. (Hill, 2010). It may be useful to facilitate the rehabilitation process in terms of functional diagnosis, treatment, and treatment evaluation. However, the implementation of telerehabilitation involves adaptations that will take some time. That is, communicative patterns between healthcare providers and their clients allow active partnership between these groups. This dimension of telerehabilitation should be encouraged since it facilitates patient empowerment, which is by now widely recognized to be effective in the management of chronic illness. That is, web applications might lower patients' thresholds to communicate their needs and give their feedback on existing therapies (Monteagudo Pena & Moreno Gil, 2007). Telerehabilitation applications thus might enhance provision of personalized information, education, training and support, and contribute to qualified patient centered healthcare. Particularly chronic neurological patients, who tend to experience numerous limitations in daily life, are in need of tailor-made support to optimize their participation in our dynamic society. Thus, telerehabilitation devices allow physicians and therapists to support patients with chronic conditions in the long term, and might even contribute to a core preventive service. As an example in Nijmegen, the Netherlands, the Radboud University, the University Medical Centre and the Sint Maartenskliniek currently collaborate in the 'ComPoli project' to implement teleconsultation for neurological patients with communicative disabilities. In the ComPoli speech and language technologies are applied to facilitate digital communication of these patients with healthcare professionals, who may otherwise be excluded from the possibility to play an active role in their rehabilitation process (Revalidatie Nederland, 2012).

Apart from patient empowerment, telerehabilitation has the potential of providing economic benefits. That is, the possibility to reduce face-to-face consultations allows rehabilitation professionals to enlarge the number of patients they are able to support and hence, to cope with a growing group of chronic clients. From the perspective of neurological patients, who tend to be less mobile and easily fatigued, telerehabilitation allows to save time, money and energy involved with visiting healthcare professionals.

An additional benefit of telerehabilitation is the lower threshold for the intensification of training, caused by patients' reduced dependence on face-to-face contact with therapists and hence, enhanced accessibility of rehabilitation services. Patients are able to practice communicative, cognitive or motor skills in their own environment at any time and as frequently as they choose. This possibility to practice skills independently from a therapist might compensate for the reducing length of inpatient stay in rehabilitation facilities (Rosen, 1999; Borghans, Heijink, Kool, Lagoe, & Westert, 2008). That is, criteria for discharging rehabilitation patients are mainly based on patients' abilities to stay safely in their home environment, either or not supported by informal caregivers such as spouses or family members. Hence, a considerable number of patients return to their home environment while motor, communicative and cognitive abilities would still benefit from extensive practice. Some patients profit from extended rehabilitation in day care facilities for some weeks, but eventually this will also often be ended before an optimal level of capacity is reached. Telerehabilitation provides the possibility of prolonged intensive training in the home situation after discharge, which has proven to be effective for motor and language capacities in neurological patients (Kwakkel, Wagenaar, Twisk, Lankhorst, & Koetsier, 1999; Bhogal et al., 2003; Rijntjes et al., 2009). It should be noticed that telerehabilitation also allows neurological patients in the chronic stage of their disease to maintain their skills over time as a result of repetitive practice. This benefit will gain importance with the on-going ageing of our population along with a growing group of chronic neurological patients.

Despite its obvious relevance, telerehabilitation does not belong to the top ten area of telehealth applications (Tulu, Chatterjee, & Maheshwari, 2007). This might be due to the fact that neurological rehabilitation is a relatively young discipline in healthcare. Nevertheless, the benefits mentioned for patients and professionals is inspiring to enhance this alternative way of healthcare delivery. In the field of communicative rehabilitation of patients with acquired neurological disability, the results of initial investigations into the potentials of telehealth applications are indeed very promising. This will be described in the next section.

1.3 Telehealth in speech-language rehabilitation

A variety of speech and language disorders, such as neurological communication disorders, fluency disorders, voice disorders, and developmental speech and language disorders have already been addressed by telecommunication-based and web-based telehealth applications (Mashima & Doarn, 2008). For the target group of patients with acquired neurological disability, the employment of various web-based applications has been explored in the last decade. Theodoros and her colleagues reported a preliminary validation of a web-based telehealth application for motor speech disorders in adults with acquired neurological disability (Theodoros, Russell, Hill, Cahill, & Clark, 2003). They found indications for the potentials of a real-time videoconferencing application for online assessment of the overall severity of motor speech disorders in neurological patients. Similar results were obtained for online assessment of aphasia, employing standardized assessments via a web-based videoconference system (Theodoros, Hill, Russell, Ward, & Wootton, 2008). Hill and her colleagues also explored the feasibility and effectiveness of a web-based telerehabilitation system for the assessment of motor speech disorders (Hill et al., 2006). Based on their pilot study, they confirmed the system's feasibility but suggested vital improvements in technology, study design and clinical assessment protocols. Implementation of these improvements in a re-designed study allowed confirmation of the validity and reliability of an initial web-based telerehabilitation system for the diagnosis and assessment of dysarthria (Hill et al., 2006). In addition, the structure of an alternative web-based assessment tool for dysarthria was described: the online Munich Intelligibility Profile (MVP) for dysarthric speakers with Parkinson's disease (Ziegler & Zierdt, 2008). MVP-online was found to be an efficient, reliable and valid method to assess speech intelligibility in dysarthria and hence seemed useful for standard clinical diagnosis. Given the online nature of these assessment tools, they could contribute to large scale studies of motor speech impairment and longitudinal studies into treatment effects.

As becomes apparent from the results of the above-mentioned studies, increasing evidence is currently being obtained indicating that telehealth is feasible, effective and appropriate for speech-language pathology services to a broad range of patients. The American Speech-Language-Hearing Association (ASHA) considers telehealth an appropriate model of service delivery for the profession of speech-language pathology (American Speech-Language Hearing Association, 2005). This professional association for therapist and researchers in the field of speech-language pathology has growing confidence that telehealth has the potential to overcome barriers of access to services caused by distance, unavailability of specialists and impaired mobility.

Although preliminary research supports the alleged benefits of telehealth, Brennan and Barker (2008) stress the importance of taking into account human factors for successful employment of telerehabilitation in general. They notice that telerehabilitation technology in the patients' home environment should be adjusted to patients' impairments, aiming at maximizing well targeted abilities and minimizing the influence of untreatable disabilities. In addition, the applied technology should ensure the use of devices that are simple and reliable to operate and that have a high level of fault tolerance. Finally, the physical training environment should be adjusted to the patients' motor abilities and minimize distractions in the training space. It is vital to meet these conditions since a web-based device (in the absence of a therapist) should compensate for the benefits of face-to-face sessions, where therapists support patients by encouraging and providing feedback on their speech efforts (Fourie, 2009). Web-based training devices should, therefore, challenge patients and keep them eager to practice. To this end, both the prerequisites for user accessibility and the minimal user conditions should be established along with the content of training programs.

As a prelude to the aims and outline of this thesis in section 1.5, the infrastructure and the rationale for web-based speech training will be presented in the next section.

1.4 E-learning based Speech Therapy (EST)

1.4.1 Infrastructure and training procedure

E-learning based Speech Therapy (EST) is a web-based speech training device that allows neurological patients with dysarthric speech to practice their speech in the home environment (Beijer et al., 2010). In fact, it is a web-based version of a pronunciation training method developed for Dutch students of English (Gussenhoven & Broeders, 1997). The EST infrastructure is presented below.

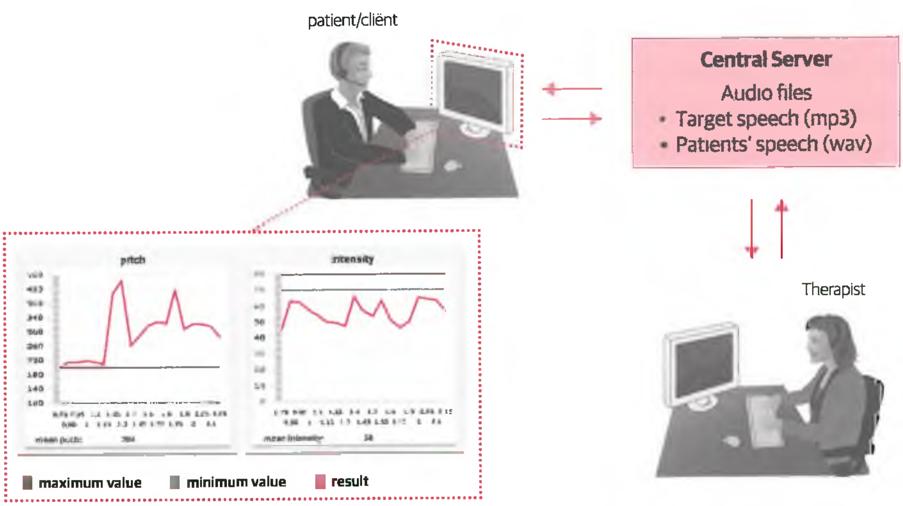


Figure 1.1

E-learning based Speech Therapy (EST)

Therapists and patients have access to a central server which initially only contains audio files of target speech. From this target speech, therapists can compile a tailor-made speech training program for their individual clients. The training procedure for neurological patients is based on the imitation of an audio example of target speech and, subsequently, the auditory comparison of the target and the patients' own speech realization.

The auditory presentation of target speech is based on research results indicating that the use of audio examples in speech training yields better results than speech practice by only reading targets aloud (Palmer, Enderby, & Cunningham, 2004). This is in line with evidence for the positive effects of phonemic perception training on the speech production and phonological awareness (Rvachew, 2004). These findings support the theoretical framework of speech motor control in the so-called Directions Into Velocities of Articulators (DIVA)-model, which assigns a vital role to auditory feedback in the speech act (Guenther, 2006) (section 1.4.2.3). The role of an acoustic target in speech training is also in agreement with the alleged benefit of using an external focus in motor skills practice, as proposed in the Schema Theory of general motor control (Maas et al., 2008). The concepts of the Schema Theory and the DIVA model of (speech) motor control will be elaborated in section 1.4.2, where the rationale of EST will be addressed.

Auditory discrimination skills play a vital role in speech training. In the EST training procedure, augmented visual feedback of the speech dimensions 'intensity' and 'pitch' is provided in addition to patients' independent auditory feedback. Patients receive a graphic display (Figure 2) of these speech dimensions once their speech has been uploaded to the server. The reason for augmented feedback is based on ample evidence for diminished auditory discrimination abilities in neurological patients (Beijer, Rietveld & van Stiphout, 2011).

Apart from the therapeutical benefits of EST, this web-based speech training application allows the upload and storage of audio files to a central server, thus facilitating data storage and longitudinal evaluation of treatment efficacy. In addition, the data base of dysarthric speech is vital for research in the field of speech technology such as automatic recognition of pathological speech and automatic error detection. Research outcomes addressing speech technological issues are necessary for the further improvement of web-based speech training systems.

1.4.2 Rationale of EST

1.4.2.1 Potentials of EST for chronic neurological patients

EST has the potential to provide the speech training conditions which perfectly fit in the fundamental principle of rehabilitation as expressed by Winters (Winters, 2002): 'remodelling the tissues and systems of cells - ranging from connective soft tissue to muscle to neurocircuity - on the basis of a sequence of inputs'. Thus, EST aims at optimizing improvement of speech intelligibility in patients with dysarthric speech due to acquired neurological diseases (e.g. stroke or Parkinson's disease (PD)). These patients frequently suffer from reduced mobility and increased physical fatigue, which are serious impediments to visit a speech therapist over a long period. Hence, telerehabilitation provides easily accessible speech therapy for these patients who, according to clinical experience, tend to suffer severely from their reduced communication abilities. EST allows patients to practice their speech intensively in a preferred environment at any time, and to extend speech training after discharge from an inpatient or an outpatient rehabilitation program.

In stroke patients with nonprogressive dysarthria, this possibility might prevent the frequently observed relapse curve due to 'sudden therapy stop' after clinical rehabilitation. (Rijntjes et al., 2009). EST may, in addition to face-to-face sessions, intensify speech training during inpatient rehabilitation in the subacute phase of a neurological disease such as stroke. The left side of Figure 1.2 shows how intensified speech rehabilitation (upper part) may contribute to the (speech) rehabilitation process during inpatient stay, since it enhances therapy outcomes (lower part). As depicted on the right side of Figure 1.2, EST also allows patients to repeat intensive speech training once face-to-face sessions have ceased in the chronic phase of the disease, thus enabling them to improve or at least maintain their speech quality in the long term. Thus, EST has the potential to contribute to the overall optimization of the rehabilitation plan stroke patients with dysarthric speech.

Apart from stroke patients, also chronic neurological patients with progressive diseases such as PD may benefit from EST once face-to-face sessions have ceased. The progressive nature of their disease urges them to practice speech in order to maintain a minimal speech quality as a prerequisite to participate in society. Given their deteriorating physical condition over time, the easy access of EST is considered a key advantage.

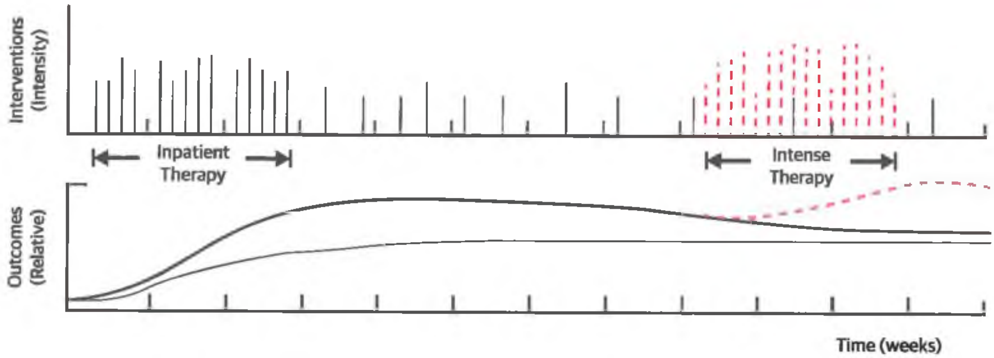


Figure 1.2

Optimizing the rehabilitation plan for neurological patients with dysarthric speech (after Winters & Winters, 2004). Therapeutic interventions are represented as vertical lines in the top trace. EST could contribute to the frequency (i.e. intensity) of therapeutic interventions, which positively affects recovery in the short term and in the long term (thick line and dashed line) compared to spontaneous recovery (thin line).

Apart from EST, a growing number of speech training software for dysarthric speakers – providing the above-mentioned benefits – is currently emerging (Parker, Cunningham, Enderby, Hawley & Green, 2006; Palmer, Enderby & Hawley, 2007). EST however provides a speech training program specifically for the Dutch language, and unlike most computerized speech training programs, regards a web-based application of speech training.

1.4.2.2 Content of web-based speech therapy

A matter of discussion concerns the content of speech therapy that is assumed to play a role in the potential effects of web-based speech training (Fourie, 2009). Apart from its particular relevance in web-based training per se, the content of traditional face-to-face speech therapy has been a major issue in clinical outcome studies of speech-language therapy for years now. Unfortunately, no scientific evidence was found for the efficacy of speech and language therapies in dysarthric speakers after nonprogressive brain damage such as stroke, traumatic brain injury or encephalitis (Sellars, Hughes, & Langhorne, 2005). The major reason was the lack of unconfounded randomized controlled trials, implying an urgent need for good quality research in this area. Hence, for now we will have to rely on general principles of motor (speech) learning. As far as dysarthric speakers with Parkinson's disease (PD) are concerned, in the Netherlands evidence based guidelines have been developed for diagnostic and treatment procedures (Kalf, de Swart, Zwarts, Munneke, & Bloem, 2008). These guidelines provide speech and language pathologists (SLPs) with recommendations for clinical practice. The Pitch Limiting Voice Treatment (PLVT) (Swart de, Willemse, Maassen, & Horstink, 2003), an adapted version of the Lee Silverman Voice Therapy (LSVT) (Ramig et al., 2001), is strongly recommended in these guidelines. The PLVT aims at maximizing respiratory and phonatory functions, inducing increased vocal adduction, laryngeal muscle activity and synergy, laryngeal and supralaryngeal articulatory movements and vocal tract configuration. At the same time the PLVT aims at setting vocal pitch at an adequate level to prevent a strained or stressed voicing.

Telepractice - without the possibility of face-to-face therapeutical tuning with neurological patients (Fourie, 2009) - necessitates to explicitly consider the format of speech training programs. This requires a thorough understanding of speech motor learning. That is, rather than relying on intuitive therapeutical approaches (as is generally the case in traditional face-to-face training), dimensions of telepractice such as type, amount and distribution of speech exercises should be based on theoretical frameworks of speech production. Available scientific evidence indicating that behavioral treatment promotes reorganization and plasticity of the brain (Schmidt & Lee, 2005) also challenges professionals in the area of speech pathology to gain more insight into the underlying principles of treatment of motor speech disorders. That is, understanding how the speech motor system reorganizes itself after neural lesions may result in more effective treatments for motor speech disorders in different underlying neurological diseases.

1.4.2.3 Principles of motor learning

As a prelude to the content of EST for the target groups of patients with stroke and PD in this thesis, two models of speech production and their implications for behavioral intervention in motor speech disorders are discussed below. Since EST was based on a pronunciation method for Dutch students of English, which emphasizes feedback on speech production (section 1.4.1.), these methods were selected on the basis of the key role they assign to feedback on motor speech production.

General Motor Planning and parametrization.

A prominent theory of general motor control is the so-called Schema Theory (Schmidt et al., 2005). The key concepts of this theoretical framework evolved from studies addressing motor learning in general, but are probably applicable to motor (re)learning principles of speech as well. The Schema Theory assumes that production of rapid discrete movements involves units of action (motor programs) that are retrieved from memory and then adapted to a particular situation. According to this theory motor programs are generalized, in that they capture the invariant aspects of the movement. To select the optimal activation pathways to the musculature and to control the body in a wide range of situations, the motor system must know the relations between 1) the initial conditions, 2) the generated motor commands, 3) the sensory consequences of these motor commands and 4) the outcome of the movement. Relations between these types of information are encoded in memory representations, called schemas. The recall schema computes the appropriate parameters that are required to produce a movement. The computation of these parameters is based on information delivered by the motor system. The sensory consequences that will occur if the movement goal is reached is predicted by the recognition schema, thus allowing the system to evaluate movements by comparing the actual with the expected sensory consequences. The Schema Theory would consider speech production a result of a General Motor Plan (GMP) and a process of parametrization that encompasses the coordination of all speech production systems. In the control of speech, a 'GMP' might correspond to the motor commands associated with a phoneme, a syllable, a word or perhaps even a frequently produced phrase, whereas speech rate and degree of clarity - resulting from the strength, the direction and the speed of speech articulators - might be considered 'parameters'. However, it is not altogether clear which aspects of speech production should be considered a 'GMP' and which aspects a 'parameter'.

DIVA model.

A theoretical framework specifically developed for motor speech control is represented by the Directions Into Velocities of Articulators (DIVA) Model (Guenther, 2006). In agreement with the Schema Theory, the DIVA model supports the interaction between perception and speech production. According to this model of speech motor control, the production of speech sounds requires the integration of diverse information sources in order to generate intricate patterning of muscle activation. To this end, a feedforward control system works in concert with a feedback control system. The feedforward system involves the premotor cortex (motor plan), the primary motor cortex (articulatory velocity and position) and the cerebellum (coordination and smoothing of movement). It works in concert with systems for somatosensory (i.e. orosensory) feedback (how the sound 'feels' in the oral cavity) and auditory feedback (how the speech segment sounds) . The control systems thus involve both sensory and motor cortical areas. New speech sounds are learned by first storing the auditory target of a sound, then using the auditory feedback system to control production of the sound in early repetitions. Repeated production of the sound leads to tuning of feedforward commands which eventually substitute the feedback-based control signals.

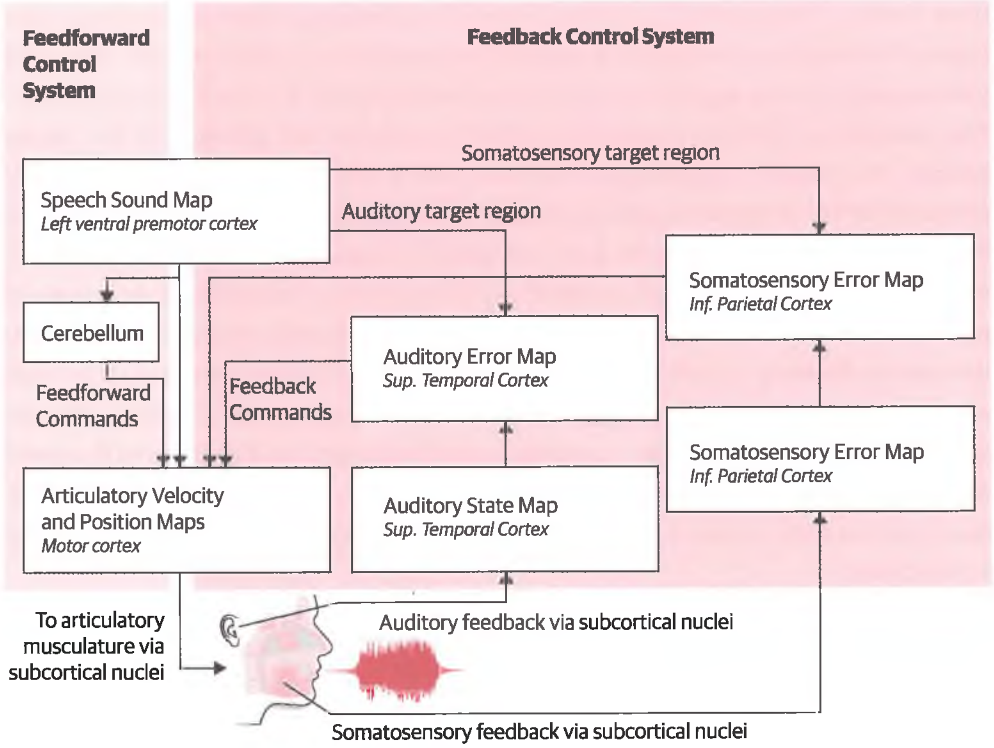


Figure 1.3

A schematic depiction of the Directions Into Velocity of Articulators (DIVA) Model. Projections to and from the cerebellum are simplified for clarity (after Guenther, 2006.)

Both the Schema Theory and the DIVA model for speech motor control assign a key role to continuous information exchange between somatosensory systems (feedback) and motor production systems (feedforward control) for adequate speech production.

Motor speech disorders.
Frequently observed symptoms in dysarthric speech of neurological patients, such as reduced stress, monoloudness, monopitch, variable speech rate, imprecise consonants and distorted vowels are symptoms of impaired motor speech production (Darley, Aronson, & Brown, 1969). According to the previously discussed theoretical frameworks of motor speech control, these types of distorted speech result from an impaired somatosensory system, an impaired motor production system, or from a distorted collaboration between these systems. According to the Schema Theory, features in dysarthric speech are the result of a distorted communication between the recall schema (i.e. performance) and the recognition schema (i.e. perceptual identification). The parameter computation of the intended speech movements delivered by the recall schema and the resulting actual sensory outcomes of the speech act do not match with the expected sensory consequences predicted by the recognition schema. This is caused by altered relations between the motor speech commands, the sensory consequences of the motor act, and the movement outcomes as defined in the recall schema. Hence, the motor system is urged to modify the recall and the recognition schemas to establish new relations in order to achieve the movement goal. This occurs for example in stroke patients, where premorbid motor specifications will not produce the intended movement outcomes, nor will the actual sensory consequences match the sensory consequences predicted from the movement goal. Motor speech commands, sensory outcomes and movement outcomes need to be adjusted and, in time, will lead to adapted recognition schemas. Similar to the Schema theory, the DIVA model also assigns a key role to distorted sensorimotor integration in dysarthric speech. This model holds a distorted coordination between the auditory and orosensory feedback system on the one hand and the motor speech production on the other hand responsible for the impaired tuning of feedforward commands in dysarthric speech. As an example, the frequently observed failure in Parkinsons' patients to recognize that their speech is hypophonic (Ho, Bradshaw, & Iansek, 2000) would result from the mismatch between experienced vocal effort (feedforward) and perceived loudness (auditory feedback). As a consequence, no internal error signal is generated by

these speakers. The Schema Theory assumes the expected sensory consequences of the speech act, as predicted by the recognition schema, to be poorly calibrated with respect to some external reference of correctness. In the case of Parkinson's patients failing to perceive the hypophonic speech, this external reference might concern a certain minimal loudness level in order to be understood. It is noteworthy that the actual and expected sensory consequences may nonetheless match and hence, no error signal will occur that can be used to update the recall schema for future attempts. As it is, external references are assigned a more prominent role in the schema theory than in the DIVA model. This might be due to the fact that the principles of speech production according to the Schema Theory were originally derived from principles of general motor learning. Nevertheless, both the Schema Theory and the DIVA model assign a key role to an adequate coordination between motor planning, motor execution and sensory feedback in speech production. Neurological diseases that distort the continuous matching of efferent and afferent processes will affect speech production. Since the impaired ability to adequately perceive their own speech is believed to play a pivotal role, it is likely that augmented feedback will facilitate speech training in neurological patients.

Treatment of motor speech disorders.

The above-mentioned theoretical frameworks for normal and distorted speech production give direction to optimal speech training. Maas and his colleagues (2008) reviewed evidence from general motor training that apply to speech motor training. This resulted in an overview of considerations for good clinical practice, which are listed below.

- **Prepractice**

  Prepractice is meant to prepare the learner for the practice session and is dependent on the specific training program to be applied. Instruction are vital to understand the task at hand. The instruction should be concise and simple, especially in the case of co-existing language disorders.

- **Practice**

  The structure of practice encompasses the following aspects:

  - Amount of practice:

    Large numbers of practice items (i.e. intensive training) have been found to be beneficial in learning non-speech motor skills (Kwakkel, 2006; Rijntjes et al., 2009). From the perspective of the Schema Theory, intensive training would enhance the stability of recall and recognition schemas and may facilitate the formation of GMPs. Maas et al. (2008) reported no empirical evidence for the amount of practice in speech motor training. However, there is evidence indicating the importance of intensity of training in aphasic patients (Bhogal et al., 2003; Basso, 2005) as well as in Parkinson's patients with dysarthric speech (Sapir et al., 2003; Sapir, Spielman, Ramig, Story, & Fox, 2007; Mahler, Ramig, & Fox, 2009).

- Practice distribution:
  Practice distribution refers to how a fixed amount of practice is distributed over time. Although positive effects have been found for massed training of the LSVT (Sapir et al., 2007) and the PLVT (Swart de et al., 2003) in Parkinson's patients, only few empirical data exist on the effects of distributed practice in speech training. Therefore, research should aim at establishing the relative effects of massed versus distributed practice on learning, transfer and retention of speech quality.

- Practice variability:
  Constant practice targets only one type of a given movement (GMP), whereas variable practice targets the execution of a set of similar movements. For motor learning in general, variable practice consists, for example, of practicing a golf swing over varying distances from the hole. For speech motor learning, variable practice may consist of articulating a vocal sound while whispering or shouting, or practicing voice intensity over various distances. The Schema Theory would consider a wide range of movements, initial states and sensory consequences within a particular GMP (e.g. practicing a particular phoneme in various word positions) beneficial, since it results in a more reliable schema. That is, practice variability highlights the relations among different types of information for variations of a given task (Schmidt & Bjork, 1992). According to the Schema Theory, variable practice results in facilitated transfer to other movements within the same GMP (e.g. the same phoneme in different phonological and phonetic contexts), but not to movements of a different GMP (e.g. another phonemic category). More research into this subject is needed to determine the

benefits of variable versus constant practice in speech training. Since the DIVA model only describes the somatosensory and the feedforward subsystems in repeated productions of a single phoneme, syllable or word, its concepts seem to be confined to constant practice. The potential benefits of variable practice might be incorporated in an expanded version of the DIVA model.

- Practice schedule:
  Random practice refers to a training schedule in which different movements (i.e. GMPs) are produced across successive trials, and in which the target for the upcoming trial cannot be predicted by the learner. Blocked practice refers to a training schedule in which the learner practices a set of the same target movements before advancing to a next target. In the speech motor domain, there is preliminary support for the use of random rather than blocked practice for both intact and impaired speech motor systems (Knock, Ballard, Robin, & Schmidt, 2000). Although this seems to be in line with the alleged benefits of practice variability according to the Schema Theory, random practice is more complicated. That is, while variable practice targets the same task in various ways (e.g. repeating a word at different loudness levels), random practice targets more training dimensions (e.g. alternate practice of syllables, words and sentences, with each linguistic level focusing on a different speech dimension). This implies that learning might be optimized by first practicing in blocked order and then, when GMPS have been established, by progressing the training towards random practice (Maas et al., 2008). As far as the DIVA model is concerned, the proposed interactions between the (somato)sensory and auditory feedback subsytems are established through repeated productions of fixed items. Hence the current DIVA model does not extend beyond the notion of blocked practice.

- Attentional focus:
  There is evidence from non-speech motor learning that using an external focus of attention is better than using an internal focus. This implies that motor training involving a tool (e.g. a golf club) is easier than learning a task while focusing on body movements only. While using a tool, attention is directed at both the body movement and the tool. For speech training, this notion is supported by findings

of Palmer & Enderby (2004) and Rvachew (1994). They found that an acoustic target (i.e. external focus) in speech training results in better improvement than training speech articulators (internal focus) only. According to the DIVA model, speech motor learning also bears on the presence of an external auditory target, being vital input for the feedforward system (Guenther, 2006).

- Movement complexity:
  The effects of movement complexity are not evident (Maas et al., 2008), which might reflect a lack of adequate definition of motor complexity. Particularly in the speech motor domain, metrics for the role of task complexity in motor learning need to be further developed. Another cause of poor evidence for task complexity effects on motor learning may lie in the lack of specification of optimal challenge points for individual speakers. That is, adequate tuning of the task difficulty to the individual skills level provides the neurological patient with a realistic target to be achieved.

• **Structure of augmented feedback.**
- Feedback type:
  There are two types of augmented feedback: knowledge of results (KR) and knowledge of performance (KP). KR is information about the movement outcome in relation to the goal, and is obtained once the task is completed. KP is information on the quality of the movement pattern, and can be provided by, for instance, artificial feedback or by a therapist (Schmidt et al., 2005). Although generally undervalued in reports of speech training, the type of feedback is a vital component in the treatment of motor speech disorders. There is no evidence yet regarding what is more beneficial for speech motor training: KR (e.g. information on the achievement of a target phoneme) or KP (e.g. information on lip closure required to produce a bilabial voiceless plosive). KP may be more effective when a task is novel, whereas KR may be more effective later in therapy and for patients who are able to evaluate their own errors (Maas et al., 2008). Whereas the Schema Theory seems to focus on KR (i.e. recognition schemas), the DIVA model seems to assign a more prominent role to KP, given the importance of continuous interaction between the feedback and feedforward systems in speech motor control.

- Feedback frequency:
  Feedback frequency refers to how often augmented feedback is provided during practice. A limited frequency of feedback has benefits for motor learning in healthy speakers (Steinhauer & Grayhack, 2000) and individuals with verbal apraxia (Austerman Hula, Robin, Maas, Ballard, & Schmidt, 2008). Further research is required into the effects of limited feedback in the treatment of motor speech disorders as well as into the interactions with previously mentioned factors such as practice variability and task complexity. According to the Schema Theory, high frequency feedback is expected to be essential for establishing GMPs. Once these GMPs have been established, reducing the frequency of feedback is likely to stimulate intrinsic feedback processes, thus facilitating proper parametrization. The DIVA model also advocates high frequency feedback to obtain input for the feedforward system in the initial learning phases. In the course of practice, the feedforward control system will gain sufficient information to produce the target speech with increasing accuracy, diminishing the need to involve the feedback control system. Nevertheless, future research should address the issue of which feedback frequency is most appropriate in which stage of motor speech learning.

- Feedback timing:
  Feedback timing refers to the moment feedback is provided relative to the performance of the task. Feedback is generally given after the completion of the task (delayed), but can also be provided simultaneously with the task performance. In general motor learning, there is evidence that delayed feedback may facilitate internal movement evaluation and, thus, is most effective for retention and transfer (Salmoni, Schmidt, & Walter, 1984). In line with a reduction of feedback frequency over time, both the Schema Theory and the DIVA model would predict that, in the course of the training period, the time between performance and feedback should be increased since this would stimulate independent speech motor control. Indeed, there is preliminary evidence suggesting that delayed feedback is most beneficial in speech motor learning (Maas et al., 2008).

Implications for EST

The EST procedure and content were based on the available knowledge of motor speech training. Important aspects of prepractice, practice and feedback were implemented as follows:

- EST provides clear instructional texts appearing on the screen along with the training items.
- EST provides large numbers of practice items during an intensive training period.
- EST consists of variable practice in both blocked and random order.
- EST uses an external focus, being an acoustic target (speech example).
- Linguistic complexity varies from isolated sounds to mono- en multisyllabic words and short sentences.
- Patients are provided with delayed augmented feedback in addition to intrinsic feedback.

1.4.3 EST research program

With many unanswered questions regarding practice and feedback issues in motor speech learning, the potentials of EST as a web-based speech training device should be investigated. Considering the target group of patients with acquired neurological diseases who tend to experience motor, communicative and cognitive disabilities, these potentials not only concern the technological feasibility and the effects of EST on speech intelligibility. The 'remote nature' of this training device, requiring a large degree of independence from the patients, also warrants an investigation into the prerequisites that dysarthric speakers should meet in order to adequately use EST. In addition, users' appreciation of web-based training should be addressed. Against this background an EST research program has been developed during the past few years as depicted in Figure 1.4.

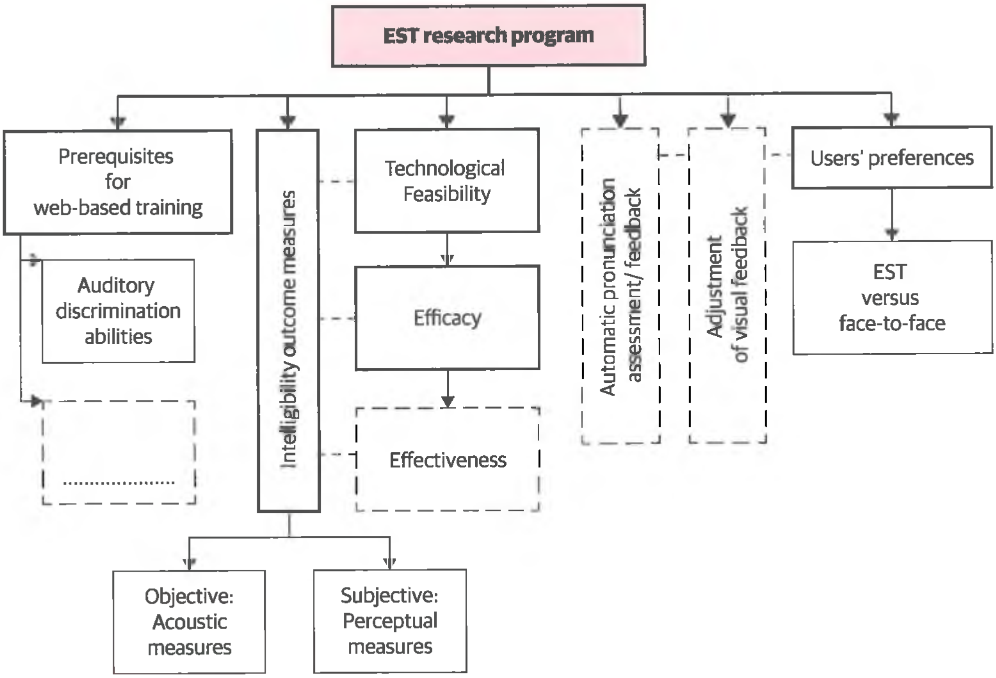


Figure 1.4

Overview of the EST research program. The uninterrupted boxes indicate the research issues which will be addressed in this thesis. The dashed boxes indicate those parts of the research program that, although crucial for EST outcome research, are beyond the scope of this dissertation.

1.5 Aims and outline of this thesis

This thesis addresses EST in patients with dysarthric speech due to stroke or Parkinson's disease. It was hypothesized that web-based speech training through EST has potentials for neurological patients with dysarthric speech. In this perspective, the following research questions are addressed in this thesis:

- Are patients with dysarthric speech due to stroke or PD able to discriminate aurally between target speech and their own speech realization?
- Are scale ratings and orthographic transcription scores of semantically

unpredictable sentences (SUS) suitable outcome measures for speech intelligibility in dysarthric speech?

- Which acoustic and perceptual outcome measures are most sensitive for establishing the effects of EST on speech intelligibility in dysarthric speakers ?
- How do dysarthric speakers appreciate speech training through EST compared to speech training in traditional face-to-face sessions?

The studies addressing the above-mentioned research questions concern initial investigations into the potentials of EST as a web-based speech training device. They typically fit into Phase I, Phase II and a transition from Phase II to Phase III of the five-phased model of clinical outcome research (Robey & Schultz, 1998), initially described by the World Health Organization (WHO, 1975). This broadly accepted model starts with Phase I in which a promising experimental treatment is tested in a few individual patients. If the experiences of Phase I warrant testing the efficacy of the treatment, Phase II consists of formulating and standardizing prescription protocols and clinical methods, validating measurement instruments, assessing factors affecting efficacy, and finding the optimal dosage of training. In Phase III the efficacy of the treatment is investigated in large samples, including controls. Quality-of-life assessments are sometimes emphasized in Phase III. When the efficacy has been established, it can be further investigated with even larger patient groups in Phase IV research. In Phase V, cost-effectiveness of the treatment is evaluated.

The thesis consists of three main parts. Part One addresses the potentials of technological developments in health care and the role of EST in patients with PD and stroke. In Chapter 2, the importance of telehealth for speech rehabilitation of neurological patients is discussed, particularly for patients with Parkinson's disease. Furthermore, the role of telehealth applications in data storage for scientific research purposes in speech and language pathology and technology is addressed. Elaborating on the therapeutical purposes of telerehabilitation applications, EST is introduced as a web-based speech training device for dysarthric patients with stroke or Parkinson's disease in Chapter 3. The infrastructure, along with the facilities for different user groups are described.

Part Two typically concerns Phase II issues of clinical outcome research: factors affecting the outcome of EST and the evaluation of the suitability of outcome measures for speech intelligibility. Chapter 4 reports on auditory discrimination capacity in neurological patients as a prerequisite for web-based speech training. Since there is ample evidence for the positive effects of an acoustic target and the vital role of auditory feedback in speech training, the development of an assessment tool for auditory discrimination skills in neurological patients is described and evaluated. The performance of neurological patients and healthy controls is compared, and the potential benefits of augmented feedback on speech quality in dysarthric patients is discussed. In Chapter 5 the suitability of scale ratings and orthographic transcription scores of semantically unpredictable sentences (henceforth 'SUS-sentences') for speech training efficacy research is evaluated. To this end, the development of five sets of SUS-sentences will be described. The equivalence of these different sets with regard to their potential intelligibility are evaluated, in order to establish whether the different sets can be used in repeated measures designs.

In Part Three, the proof of principle of speech training in dysarthric patients through EST is addressed. In line with the phases of clinical outcome research, Chapter 6 describes a case study (Phase II), exploring the effects of EST on speech intelligibility in a Parkinson's patient. The results warranted an exploratory effect study (n=8) (Phase II-III). This study is reported in Chapter 7, focusing on the evaluation of the suitability of the acoustic and perceptual outcome measures for clinical outcome research. Given the increasing role of patients in their own rehabilitation process - against the background of patient centered health care - Chapter 8 addresses patients' appreciation of web-based versus traditional face-to-face training.

In the general discussion (Chapter 9), the results of the studies that were conducted are discussed in view of the research questions mentioned above. In addition, considerations are made on the complexity of developing, evaluating and implementing telehealth applications in existing health care systems. Particular attention is paid to the prerequisites for successful employment of telerehabilitation applications.

References

American Speech-Language Hearing Association (2005). Speech-Language Pathologists Providing Clinical Services via Telepractice: Position Statement. Available at www.asha.org/telepractice (last accessed July 31, 2012).

Austerman Huia, S. N., Robin, D. A., Maas, E., Ballard, K. J., & Schmidt, R. A. (2008). Effects of feedback frequency and timing on acquisition and retention of speech skills in acquired apraxia of speech. *Journal of Speech, Language, and Hearing Research, 51*, 1088-1113.

Basso, A. (2005). How intensive/prolonged should intensive/prolonged treatment be? *Aphasiology, 19*, 975-984.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., & Geurts, A.C.H. (2010). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and eHealth, 16*, 177-180.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*, 987-993.

Borghans, I., Heijink, R., Kool, T., Lagoe, R. J., & Westert, G. P. (2008). Benchmarking and reducing length of stay in Dutch hospitals. *Health Services Research, 8*. Available at: http://www.biomedcentral.com/1472-6963/8/220

Brennan, D., Georgeadis, A., & Baron, C. (2002). Telerehabilitation tools for the provision of remote speech-language treatment. *Top Stroke Rehabilitation, 8*, 71-78.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 462-496.

Fourie, R. J. (2009). Qualitative study of the therapeutic relationship in speech and language therapy: persepectives of adults with acquired communication and swallowing disorders. *International Journal of Language and Communication Disorders, 44*, 979-999.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*, 350-365.

Gussenhoven, C. & Broeders, A. (1997). E*nglish Pronunciation for Student Teachers.* (2nd ed.) Groningen: Wolters-Noordhoff.

Hill, A. J. (2010). R*eport on the potential application of telerehabilitation to adult rehabilitation services in Scotland.* Available at: http://www.sctt.scot.nhs.uk/pdf/rehabreport.pdf.

Hill, A. J., Theodoros, D. G., Russell, T. G., Cahill, L. M., Ward, E. C., & Clark, K. M. (2006). An internet-based telerehabilitation system for the assessment of motor speech disorders: a pilot study. *American Journal of Speech-Language Pathology, 15*, 45-56.

Ho, A. K., Bradshaw, J. L., & Iansek, R. (2000). Volume perception in parkinsonian speech. *Movement Disorders, 15*, 1125-1131.

Kalf, J. G., de Swart, B. J. M., Zwarts, M. J., Munneke, M., & Bloem, B. R. (2008). Gu*idelines for speech-language therapy in Parkinson's disease. Movement Disorders, 23*, S328.

Knock, T., Ballard, K. J., Robin, D. A., & Schmidt, R. A. (2000). Influence of order of stimulus presentation on speech motor learning: A principled approach to treatment for apraxia of speech. *Aphasiology, 14*, 653-668.

Kwakkel, G. (2006). Impact of intensity of practice after stroke: issues for consideration. *Disability and*

Rehabilitation, 28, 823-830.

Kwakkel, G., Wagenaar, R. C., Twisk, J. W. R., Lankhorst, G. J., & Koetsier, J. C. (1999). Intensity of leg and arm training after primary middle-cerebral-artery-stroke: a randomised trial. *Lancet, 354*, 191-196.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J. et al. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*, 277-298.

Mahler, L. A., Ramig, L. O., & Fox, C. (2009). Intensive voice treatment for dysarthria secondary to stroke. *Journal of Medical Speech-Language Pathology.*

Mashima, P. A. & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine Journal and eHealth, 14*, 1101-1117.

Monteagudo Pena, J. L. & Moreno Gil, O. (2007). *e-Health for patient empowerment in Europe* Madrid.

Palmer, R., Enderby, P., & Cunningham, S. (2004). The effect of three practice conditions on the consistency of chronic dysarthric speech. *Journal of Medical Speech-Language Pathology, 12*, 183-188.

Palmer, R., Enderby, P., & Hawley, M. (2006). Addressing the needs of speakers with longstanding dysarthria: computerized and traditional therapy compared. *International Journal of Communication Disorders, 42* (S1), 61-79.

Parker, M., Cunningham, S., Enderby, P., Hawley, M., & Green, P. (2006). Automatic speech recognition and training for severely dysarthric users of assistive technology: The STARDUST project. *Clinical Linguistics and Phonetics, 20*, 149-156.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. Journal of Neurology, *Neurosurgery and Psychiatry, 71*, 493-498.

Revalidatie Nederland (2012). Implementatie eHealth komt op gang. Available at: http://www.revalidatie.nl/actueel/nieuwsarchief/nieuwsarchief/implementatie_ehealth_komt_op_gang [On-line].

Rijntjes, M., Haevernick, K., Barzel, A., van den Bussche, H., Ketels, G., & Weiller, C. (2009). Repeat therapy for chronic motor stroke: a pilot study for feasibility and efficacy. *Neurorehabilitation and Neural Repair, 23*, 275-80.

Robey, R. R. & Schultz, M. C. (1998). A model for conducting clinical-outcome research: an adaptation of the standard protocol for use in aphasiology. *Aphasiology, 12*, 787-810.

Rosen, M. J. (2004). Telerehabilitation. *Telemedicine Journal and e-Health, 10*, 115-117.

Rosen, M. J. (1999). Telerehabilitation. *NeuroRehabilitation, 12*, 11-26.

Rvachew, S. (2004). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*, 347-357.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*, 355-386.

Sapir, S., Spielman, J., Ramig, L. O., Hinds, S. L., Countryman, S., Fox, C. et al. (2003). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on ataxic dysarthria: a case study. *American Journal of Speech-Language Pathology, 12*, 387-399.

Sapir, S., Spielman, J., Ramig, L. O., Story, B., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT] on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *Journal of Speech, Language and Hearing Research, 50*, 899-912.

Schmeler, M. R., Schein, R. M., & McCue, M. (2009). Telerehabilitation Clinical and Vocational Applications for Assistive Technology: Research, Opportunities and Challenges. *International Journal of Telerehabilitation, 1*, 59-72.

Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schmidt, R. A. & Lee, T. D. (2005). *Motor control and learning: a behavioural emphasis.* (5th ed.) Champaign, IL: Human Kinetics Publishers.

Sellars, C., Hughes, T., & Langhorne, P. (2005). Speech and language therapy for dysarthria due to nonprogressive brain damage. *Cochrane Database of Systematic Reviews*, Art. No.: CD002088.

Steinhauer, K. & Grayhack, J. P. (2000). The role of knowledge of results in performance and learning of a voice motor task. *Journal of Voice, 14*, 137-145.

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60*, 498-500.

Theodoros, D., Hill, A., Russell, T., Ward, E., & Wootton, R. (2008). Assessing acquired language disorders in adults via the Internet. *Telemedicine and e-Health, 14*, 552-559.

Theodoros, D., Russell, T. G., Hill, A., Cahill, L., & Clark, K. (2003). Assessment of motor speech disorders online: a pilot study. *Journal of Telemedicine and Telecare, 9*, S66-S68.

Tulu, B., Chatterjee, S., & Maheshwari, M. (2007). Telemedicine taxonomy: a classification tool. *Telemedicine and e-Health, 13*, 349-358.

WHO (1975). *WHO Scientific Group on Guidelines for Evaluation of Drugs for Use in Man* Geneva: World Health Organization.

WHO (2001). *International Classification of Functioning, Disability and Health. Short version.* Geneva: World Health Organization.

WHO (2012). Rehabilitation. http://www.who.int/topics/rehabilitation/en [On-line].

Winters, J. M. (2002). Telerehabilitation research: emerging opportunities. *Annual Review of Biomedical Engineering, 4*, 287-320.

Ziegler, W. & Zierdt, A. (2008). Telediagnostic assessment of intelligibility in dysarthria: A pilot investigation of MVP-online. *Journal of Communication Disorders*, 553-577.

PART 1

POTENTIALS OF EST

CHAPTER 2

Potentials of telehealth devices for speech therapy in Parkinson's disease

L.J. Beijer, A.C.M.. Rietveld

Adapted from:
Parkinson's Disease, 2011, Book 6, Chapter 18,: pp. 379-402. Open Access Publisher InTech

2.1 Introduction

Rapidly evolving technological developments have been influencing our daily lives for a few decades now. In particular information and communication technologies enable more sophisticated and faster ways of communication than ever before. These developments have far reaching consequences for professional, domestic and leisure activities. Use of computers, whether or not with an internet connection, has become very common also in the field of scholar education and healthcare. The primary benefits in these areas are believed to lie in cost reduction and enhanced efficiency.

In this chapter we will focus on the possibilities of technological developments in healthcare, particularly for patients with Parkinson's disease. This patient group is believed to benefit considerably from innovative applications of information and communication technologies since the number of parkinsonian patients is dramatically growing due to demographic developments. Moreover, Parkinson's disease pre-eminently concerns a chronic and progressive illness, increasingly disabling these patients in almost all domains of their lives. In this chapter we will explore how telehealth technology and speech technology relates to the maintenance of their communicative competence.

2.2 General aspects of speech disorders associated with Parkinson's disease

2.2.1 General motor functions and oral motor control

Parkinson's disease is caused by the progressive impairment of neurons in an area of the brain known as substantia nigra. This is due to an imbalance in two brain chemicals (i.e. dopamine and acetylcholine) that are responsible for the transmission of nerve messages from the brain to the motor nerves in the spinal cord which control muscle movement. As a result, the communication between the substantia nigra and the corpus striatum, required for coordinating smooth and balanced muscle movement, is distorted.

The diminished functioning and coordination of respiratory, laryngeal and supralaryngeal muscles obviously affects speech, swallowing and saliva control in parkinsonian patients (Ziegler, 2003). As a consequence, the quality of speech in patients with Parkinson's disease tends to be deteriorated to some degree. As it is, dysarthria is a common manifestation of Parkinson's disease (PD) that increases in frequency and intensity with the progress of the disease (Streifler, 1984). Hypokinetic dysarthria is mainly associated with PD; mixed dysarthria tends to occur in atypical parkinsonism.

General motor symptoms such as rigidity, bradykinesia (reduced speed of muscles), tremors or trembling are reflected in typical speech symptoms of hypokinetic dysarthria. Bradykinesia associated with Parkinson's disease causes difficulty in the initiation of voluntary speech. This can result in delay in starting to talk as well as very slow speech. According to Duffy (1995), there may be freezing of movement during speech. Rigidity can also occur. Additionally, parkinsonian patients have reduced loudness, imprecise consonant production, reduced pitch variability and festinating speech. The latter can result in extremely fast speech together with short rushes of speech (Ferrand and Bloom, 1997). Perceptual features of parkinsonian speech associated with hypokinetic dysarthria, are a weak, breathy (hoarse) voice, monotone and monoloud speech, low volume, articulatory imprecision and rate disturbances (Darly et al., 1969). The syndrome of parkinsonian dysarthria is by no means homogeneous with respect to speech rate. This might be due to different consecutive stages in the development of Parkinsonian dysarthria or to different degrees of impairment (Ackermann & Ziegler, 1991). In general, more pronounced phonatory than articulatory disturbances tend to occur as far as clinical-perceptual ratings are observed.

2.2.2 Speech intelligibility

As a consequence of distorted oral motor functions and coordination, speech intelligibility in patients with PD obviously tends to be reduced. Prevalence studies point out that about 70% of home-living patients with Parkinson's disease have speech complaints (Kalf et al., 2008a), which are mainly associated with hypokinetic dysarthria. Diminished communication skills, in addition to the fact that these patients are increasingly disabled in their physical condition and motor abilities in the course of their disease, frequently lead to experiences of a deteriorating quality of life (Slawek et al., 2005). Improving or at least maintaining speech quality as long as possible is essential to enable optimal social participation and to maintain relationships. As a consequence, patients with PD are often eager to practice and improve their speech quality.

Although the majority of patients with PD are dysarthric speakers, only a minority of this group with diminished speech quality receives speech therapy (20-30%). The small percentage of PD patients under 'speech therapeutical control' might be partially due to the fact that a majority of speech language therapists consider themselves not capable to adequately treat dysarthric speakers with PD. Another reason might be the fact that speech therapy often is provided to patients that have only recently been diagnosed with PD. Once face-to-face therapy sessions have been completed, patients disappear out of a speech therapist's view. Thus, dysarthric patients tend to be deprived of speech therapy in the chronic and deteriorating course of their disease. As it is, this tendency of 'undertreatment' of dysarthric patients with PD is likely to continue in the course of the coming years. That is, the incidence and prevalence of PD will increase due to our aging population. It is estimated that in 2030 about 30% of our population will be 65 years of age or older.

2.2.3 Guidelines for diagnostic and treatment procedures

Apart from the observed 'undertreatment' of patients with PD, large variability exists in therapeutical approaches of this patients group. In the Netherlands, evidence based guidelines for diagnostic and treatment procedures for patients with PD were developed in order to provide speech-language therapists with recommendations for their clinical practice (Kalf et al., 2008b). It should be noticed that methodological quality of comparative studies is insufficient to meet the conditions of highest level of evidence (Deane et al., 2001). Therefore, evidence is mainly based on comparative studies of less methodological quality, noncomparative studies or experts' opinions. Five key points for treating dysarthric speech in patients with PD were formulated:

1) Patients with PD have basically normal motor skills, requiring to be elicited in an adequate way.
2) Hypokinesia increases when duration and complexity of motor acts increase. Therefore, complex acts should be divided into more simple acts.
3) Separate acts should therefore compensate for failing automatic motor acts
4) External cues could support initiation and continuation of motor acts.
5) Simultaneous execution of motor and cognitive tasks should be avoided, since execution of motor tasks already puts considerable demands on cognitive functions.

For diagnosis and treatment of dysarthria in patients with PD, two procedures are strongly recommended. 1) As far as diagnostic procedures are concerned, the initial situation should be assessed by documenting spontaneous speech and establish to what extent speech can be stimulated by means of maximum performance tests. 2) For treatment, the Lee Silverman Voice Treatment (LSVT) (Ramig et al., 2001) and the Pitch Limiting

Voice Treatment (PLVT) (de Swart et al., 2003) are strongly recommended. The LSVT focuses on tasks to maximize respiratory and phonatory functions in order to improve respiratory drive, vocal fold adduction, laryngeal muscle activity and synergy, laryngeal and supralaryngeal articulatory movements, and vocal tract configuration. The PLVT also aims at increasing loudness but at the same time sets vocal pitch at an adequate level. The LSVT and the PLVT produce the same increase in loudness but PLVT limits an increase in vocal pitch and claims to prevent a strained or stressed voicing (de Swart et al., 2003). Both therapy programs concern intensive training periods of four sessions weekly during a training period of four weeks. Intensive speech therapy is preferred if diagnostic results allow highly frequent and intensive training. That is, voice quality, intrinsic motivation, physical condition and cognitive abilities are vital conditions for intensive training of newly acquired speech techniques. In case a patient's condition does not allow intensive training, augmentative and alternative procedures or devices could provide a solution to the communication problems experienced by dysarthric speakers with PD.

In the next sections, we will go into more detail with respect to current trends in speech therapy for patients with PD. These result from rapidly evolving developments in information-, communication- and speech technology. Not only will these developments provide patients with new therapy facilities; they are also expected to bring about some crucial changes for health care providers (i.e. speech therapists) and influence health care processes.

2.3 Current trends in the therapy of speech disorders related to Parkinson's disease

2.3.1 Increased need for speech training

A considerable percentage of PD patients experience oral motor disorders, causing problems with swallowing, speech and saliva control. With 70% of PD patients being dysarthric, it is obvious that therapeutical interventions are required. That is, the speech of PD patients with predominantly hypokinetic dysarthria, needs treatment in order to improve speech intelligibility. Since communication skills are vital for adequate social participation, improvement of these abilities can significantly contribute to quality of life.

A number of current trends seem to influence the developments in speech therapy for parkinsonian patients. Firstly, there is an increased attention for dysarthria and its treatment. This is partially due to the results of scientific research in the field of PD, enhancing care givers' awareness of the relevance for longlasting communication skills in parkinsonian patients. Secondly, recent social and demographic developments have caused patients to be more aware of possibilities for treatment and set themselves on assertive in their call for adequate information. Patient centred health care has even gained considerable importance for reimbursement companies that find themselves increasingly confronted with clients searching for the best quality of care. Apart from this, one of the most serious financial crises since years urges the economic health care society to treat the growing number of elderly patients with neurological diseases with less financial means. It is obviously a challenge to maintain a sound balance between the need (and call) for speech training on the one hand, and the availability of professionals and financial means for speech training on the other hand. Particularly with current speech training programs for PD such as the LSVT (Ramig et al., 2001) and the Dutch PLVT (de Swart et al., 2003), involving intensive speech training for several weeks to enhance speech intelligibility, it becomes clear that traditional speech therapy does no longer meet the actual needs of our current health care society.

2.3.2 Telehealth in the field of speech-language pathology

Telehealth applications, resulting from recently developed information and communication technologies in health care, could provide solutions to overcome barriers of access to therapy services caused by factors such as decreasing financial resources, shortage of professionals and increasing number of clients. The terms 'telemedicine' and 'telehealth' are sometimes used interchangeably. Telemedicine is considered a subset of telehealth. Telemedicine uses communication networks for delivery of healthcare services and medical education from one geographical location to another, primarily to address challenges like uneven distribution and shortage of infrastructural and human resources (Sood et al., 2007). 'Telehealth' is a broader term and does not necessarily involve clinical services. It can be defined as the use of telecommunication technologies both to provide health care services and to enable access to medical information for training and educating health care professionals and consumers. As such, telehealth concerns all applications of information and communication technologies, enabling the retrieval, recording, management and transmission of information to support health care. In this chapter, we refer to this latter definition when discussing telehealth.

Mashima and Doarn's (2008) overview of telehealth activities in the field of speech-language pathology provide a strong foundation for broader applications of telehealth technologies in this area. Also telehealth applications for treatment of patients with neurogenic communication disorders have been reported. Theodoros et al. (2006) report an online speech training for PD patients which turned out to be effective. In Australia 10 patients with PD followed the LSVT online using video conferencing, during a four-week program of intensive training, involving 16 therapy sessions. Comparison of sound pressure level, pitch measurements and perceptual ratings from audio recordings pre- and posttreatment, containing participants' reading and conversational monologue, showed significant improvements, comparable to previously reported outcomes for the LSVT when delivered face-to-face. This example shows that remote diagnosis and treatment of speech in parkinsonian patients has vital benefits, in particular for patients who are less mobile and easily fatigued due to their deteriorated physical condition. Ziegler and Zierdt (2008) report an online version of a computer-based intelligibility assessment tool: the Munich Intelligibility Profile. The web-based MVP-version is reported to have potentials for dysarthric speech of patients with PD and other underlying neurological diseases such as stroke.

2.3.3 E-learning based Speech Therapy (EST)

Quite recently, in the Netherlands a web-based speech training device 'E-learning based Speech Therapy' (EST) has been developed (Beijer et al., 2010a). EST primarily aims at patients with dysarthric speech resulting from acquired neurological impairment such as stroke and Parkinson's disease. According to our clinical experience, these patients suffer from their deteriorating quality of speech. Particularly in the chronic phase of their disease, once therapy sessions have been completed, the lack of practice results in diminished speech intelligibility. With verbal communication being a vital condition for adequate social participation, diminished abilities in this field can be considerably invalidating. A vital benefit of EST is the possibility to follow a tailor-made speech training program in the patients' home environment. That is, time, energy and costs normally involved with speech training can be reduced for these patients who tend to be less mobile and easily fatigued due to their physical condition. In addition, the possibility to practice speech in the home environment at any moment, allows intensive speech training, which is known to be effective in patients with acquired neurological diseases (Kwakkel et al., 1999). Repetitive training in chronic phases also has been proven to have positive effects on speech intelligibility (Rijntjes et al.., 2009).

Since telehealth application tend to differ in many respects, Tulu et al., (2007) made an effort to provide insight into the large number of innovative web-based devices that are presented. They introduced a taxonomy of telehealth applications along five dimensions: communication infrastructure, delivery options, application purpose, application area and environmental setting. According to this classification, EST concerns a store-and-forward web application for treatment (i.e. training) purposes in the area of speech pathology, that is commonly used in the home environment.

The keystone of the EST infrastructure is formed by a central server. The server hosts two types of audio files: target speech files in MP3 format and recorded speech files uploaded by patients in wav format (Figure 1.1 in Chapter 1). A desktop computer or a laptop with internet connection provides users with access to the server. Using their EST therapist account, therapists are able to remotely provide their patients with a tailor-made speech training program, which is compiled from audio examples of target speech, stored at a central server. Patients have access to this program using their client account. In the EST training procedure patients listen to audio examples of target speech which is downloaded from the server. Subsequently they imitate the audio example, in order to approach the target speech. The target and the own speech are then aurally compared. Finally the patients' speech is uploaded and stored at the server (Figure 2.1.)

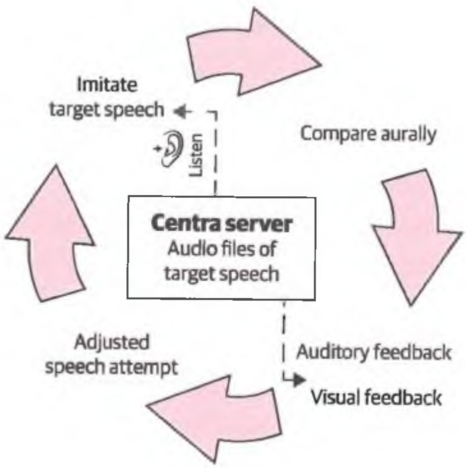


Figure 2.1:

EST training procedure (reproduced with credit of Telemedicine and e-Health)

Obviously, this training procedure puts considerable demands on patients' auditory speech discrimination skills. However, indications have been reported that patients with PD experience problems with estimating the own speech volume (Ho et al., 2000) and with auditory speech discrimination (Beijer, Rietveld & van Stiphout, 2011). Although this diminished auditory discrimination might be caused by cognitive problems and hearing loss, these patients would benefit from additional visual feedback on their own speech realization. That is, visualization of speech might support them in the auditory discrimination task of the EST training procedure. Although this visualization is already implemented in EST, the abstract graphs (Figure 1.1. in Chapter 1) and the delayed, post hoc display of visual feedback appeared not suitable for all patients (Beijer et al., 2010b). Therefore, the development of an intuitive visualisation of loudness and pitch is currently underway in order to apply to patients with various backgrounds (i.e. educational levels, age, gender). Not only should the graphic form of the visual feedback apply to the patients, indicating into what direction a new speech attempt should be adjusted to approach the target. It should also be assessed to what extent different visualisations contribute to the improvement of speech intelligibility. In section 2.4.4. we will go into some detail with respect to visual feedback on pitch and intensity (loudness).

2.3.4 EST research issues

It will be clear that innovative web-based applications for diagnostic and treatment purposes should be evaluated from several perspectives. First of all, the technological feasibility should be proven. Secondly, patients as well as therapists should be able to operate the web-based devices. Hence, user satisfaction should be evaluated since this is obviously vital for successful implementation. The term 'user satisfaction' needs to be accurately defined to ensure comparison of user satisfaction across time and across different web-based devices. This brings us to the need to establish minimum user requirements regarding physical condition, motor coordination skills and auditory or cognitive abilities. Assessing these conditions for successful use of web-based devices is vital for parkinsonian patients who tend to experience constraints in more domains than communication or speech alone. Thirdly the efficacy and the effectiveness of EST should be evaluated. This brings us to the vital issue of reliable outcome measures for treatment outcomes. In the case of parkinsonian speech, these treatment outcomes primarily concern speech intelligibility. Most of these outcome measures concern subjective, perceptual measures of speech quality (FDA, rate scaling, etc.) Along with health care reimbursers' call for objective outcome measures however, current trends point into the direction of objective acoustical measures of speech quality as a vital outcome measure for speech intelligibility in addition to traditional perceptual measures.

Employment of web-based applications for diagnosis and treatment tend to go perfectly along with the need for speech technological developments. That is, speech data can be easily collected, thus generating an automatic data base of pathological speech. We will elaborate on this in section 2.4.5.

2.4 The role of speech technology in speech therapy

2.4.1 Background

Since more than 25 years phoneticians, speech technologists and speech therapists have systematically investigated the phonetic correlates of speech disorders. These investigations were carried out with a number of explicit or implicit objectives: a) to corroborate subjective judgements of speech therapists, b) to find objective evidence for progress as a result of therapy, c) to facilitate the distinction of subgroups of pathologies, and d) to find evidence for theories on the nature of pathologies, which could not be obtained on the basis of subjective measurements. As has been the case in phonetics sciences, the progress of computer and information technology made available a number of additional applications, which form the core of the current chapter:

a) Gathering objective evidence based on acoustic and/or physiological data,
b) The development of systems which can be used by patients to obtain direct or indirect feedback on their realizations in a training program,
c) The implementation of feedback systems in telehealth applications in order to facilitate intensive training at home.

In the following we will focus on a number of applications of speech technology to be used in the assessment and treatment of dysarthria in general and that of dysartria associated with m. Parkinson in particular. We should be aware of the fact that phonetics and the associated speech technology are language bound, that is to say that phenomena which are relevant in one language, may be irrelevant in another.

Kent et al. (1999) published a seminal overview of acoustic correlates of quite a number of phenomena associated with dysarthria. The overview distinguishes the conventional phonetic components of the speech production process: Initiation, Phonation, Articulation, Velopharyngeal functioning and Prosody. As is the case with most acoustic correlates of speech disorders, stochastic relations between perceptually distinctive disorders on one side and acoustic correlates on the other are more evident than inferential procedures which boil down to statements like: the F1 and F2 (first and second formants) of segment X are higher/lower than 'normal', so we can be sure that this segment was not realized in a canonical way. The fact that trade-off relations exist between speech production and speech perception is the nuisance factor. Trade-off relations in phonetics occur when an effect in one domain - say segment duration - can compensate for the absence of a feature in another domain, say voicing. In English, for instance, a relatively long pre-consonantal vowel can perceptually compensate for an obstruent which is incorrectly voiceless. The presence of this kind of trade-off relations is an obstacle in finding clear and unambiguous acoustic correlates of the perception of speech and speech disorders. This fact is not a direct problem in group studies, which aim at finding tendencies in signal characteristics between pathological groups and a control group. In set-ups in which the aim is to provide stable and robust feedback to a patient, the presence of trade-offs can be disturbing.

Providing instrumental feedback to speakers has quite a long history. As a matter of fact, there are two parallel developments. One development focuses on learning a foreign language ("L2"), and the other on correcting speech disorders. It is quite obvious why these developments are parallel: the dimensions on which deviations of target speech can be projected are - most of the time - equal or similar: prosodic dimensions, dimensions of segmental quality, phonatory dimensions and dimensions of velopharyngeal functioning. Like in speech pathology, it is hardly ever the case that all dimensions are equally relevant. For French as L2, nasality is more important than for English or Dutch. Intonation and tone is extremely important for languages like Chinese, and much less important for English, French or Dutch. These facts have directed research both in systems which provide feedback in L2 learning and in speech pathology. Until now, it has proven not to be worth the effort in this context to assess all characteristics of speech which are imitations of target speech. It is better to direct efforts to specific segments which are known to be vulnerable and or relevant in L2 learning and speech pathology. This brings us to two different lines in feedback:

a) Direct, quasi RT feedback on the realization of global parameters like intensity, tempo and intonation and parameters associated with segmental quality, and
b) Indirect, post-hoc feedback on the realization of speech parameters.

Direct feedback is meant to help the patient in non-face-to-face training sessions. Indirect, post-hoc feedback is often only needed when the therapist has to have access to assessment scores; it is only available after quite a number of speech materials have been collected.

There might be a misunderstanding when it is decided to provide web-based speech therapy, in the sense that it is often assumed that a computer system which provides feedback is immediately applicable in an e-health application. That is not true. Supervised training/learning often cannot be directly applied in an environment in which direct assistance is absent. Supervised learning is much more robust than its non-supervised counterpart. An example is provided by Carmichael (2007). In his study, which aimed at the development of objective acoustic measures for the Frenchay Dysarthria Assessment Procedure (Enderby, 1980), the calibration of the 'loudness' measurements might be somewhat complicated to be performed at home. It involves the use of a Sound meter at a standard distance. In his set-up the test administrator performs the calibration procedure. In order to avoid this kind of calibration at home, we opted in our eHealth application the production of a long nasal consonant [mmmm]. As the production of this consonant does not involve any mouth opening and variable jaw movements, the radiated sound can be assumed to be quite constant, and to function as a reliable calibration.

Speech technology in the context of speech pathology can be divided in a number of approaches, which also depend on the objectives to be achieved. In this chapter we restrict our review to applications in assessment, therapy and training. We distinguish five dimensions on which the approaches can vary:

Dimension I:	Either the parameters focus on global parameters like intensity, tempo and pitch, or on characteristics which reflect segmental quality.
Dimension II:	There are two types of results to be obtained, viz. global assessment, or direct feedback.
Dimension III:	Types of speech: the assessment of free speech, or the assessment of read, known speech.
Dimension IV:	The inclusion of physiological parameters, like reflexes and respiration.
Dimension V:	The inclusion of facial expressions as parameter(s).

A more general dimension is the user-interface. Of course, the interface for the therapist requires less attention, but the one for the patient asks for robustness and psychological validity.

Dimension I

In specific therapies, like the PLVT (de Swart et al., 2003), global parameters are of great importance. As explained in section 1.3 of this chapter, it involves two therapy goals: “speaking loud” while not increasing pitch at the same time. The rationale is that speaking loud generally leads to an increase of articulatory precision, while increasing pitch above habitual level may harm the vocal cords.

Dimension II

Global assessment - not to be confused with global parameters - involves the assessment of speech on a long-term scale. That is, direct feedback is not provided, only feedback after some amount of speech materials has been realized, recorded and analyzed. In an application for direct feedback, the user is provided with quasi real time feedback on the quality of the speech parameters at issue: global parameters like intensity and F0, or feedback on specific segments, like vowels or consonants.

Dimension III

Very often, known and consequently read speech will be used in assessment and therapy sessions. The automatic recognition of speech and the detection of deviations from it, are enormously facilitated by the use of this kind of texts (it implies what is called forced recognition). The drawback is that the use of read speech may decrease the ecological validity of the measures and indices thus obtained.

Dimension IV

As is well-known, the initiation phase in speech, which refers mainly to respiration, is crucial for the generation of speech. There are, to our knowledge, no applications available yet which provide assessment and feedback on initiation (respiration) parameters.

Dimension V

In a number of speech pathologies the assessment of facial expressions is a relevant issue. This is also the case with dysarthria. The recognition and assessment of facial expression demands dedicated software, which is quite difficult to tune to the demands of the patient and/or therapist.

2.4.2 Realization of assessment and feedback systems

For the realization of feedback and assessment on global parameters (F0, Intensity), relatively simple algorithms are needed, often implemented in current software packages for signal analysis, like PRAAT (Boersma & Weenink, 2011). The problem there is not the analysis of the parameters itself, but the display of the results and the feedback on deviations from the goal values. No significant changes in the detection of the global speech parameters are to be expected, but work has to be done in order to provide displays which facilitate insight in possible errors and stimulates improvements.

For the realization of feedback and assessment of segmental quality, speech technology comes to play. There are a number of approaches, depending again, on the objectives of the application: direct/indirect feedback on the realization of each target speech sound (phoneme), direct/indirect feedback on the realization of single words, direct/indirect feedback on the realization of a short text, direct/indirect feedback on fixed or free texts, and feedback on the overall intelligibility of words and texts. The two main technical approaches are: the analysis of speech based on Automatic Speech Recognition (ASR) and ASR-free analysis of speech (Middag et al. 2010). If ASR is used, the Hidden Markov Model (HMM) is the main tool. HMMs constitute the default tool for speech recognition (ASR: Automatic Speech Recognition), although other approaches are also possible (Middag et al., 2010). Hidden Markov Models are based on probabilities of states of speech segments and transitions from one state to another, or to the same state. In light of the popularity of HMMs, we present a short account of this approach which has been previously applied in several research projects (Parker, Cunningham, Enderby, Hawley & Green, 2006; Cucchiarini, Neri & Strik, 2009). To illustrate HMMs we give a fictitious example of the use of an HMM for the recognition of the vowel /i/. The acoustic parameter used is the second formant (F2) only. In that respect the example is already fictitious: HMMs hardly ever use formants as acoustic parameters, let alone just one formant. The default parameters used to reflect the spectra of the sound segments are Mel-Frequency-Cepstral Coefficients (MFCCs), which also take into account human perception (Davis & Mermelstein, 1980).

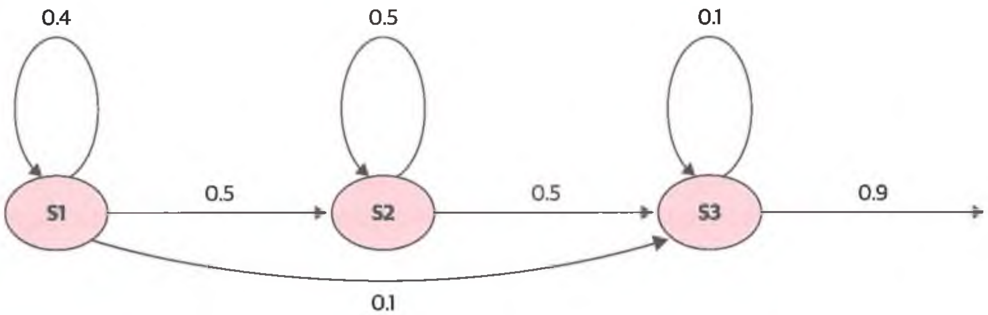


Figure 2. 2:
A hypothetical Hidden Markov Model of the vowel /i/. (Reproduced with permission from Rietveld & van Heuven, 2009).

In Figure 2.2 we display a model of the second formant (F2) of the vowel /i/. We see an inner loop from state 2 to state 2, which occurs with a probability of 0.5; this means that the model has a relatively high probability of staying in state 2, which boils down to the realization of a long vowel. The probability of going from state 1 (initial state) to the final state (S3) is small, only 0.1. The probabilities of obtaining discrete values of F2 for state 2 (low, rather low, rather high and high.) (E1, E2, E3, E4) are 0.05, 0.10, 0.15 and 0.70 respectively. Thus the probability of finding a high frequency of F2 in the middle of the vowel /i/is rather high.

In reality we observe small speech frames, with - in our restricted and hypothetical example - only values of F2. The sequence of values of the F2, for instance E2, E4, E4, E4, E3 (each frame covering 20 ms), has to be compared with the probabilities implied by the model. The model which has the largest probability of having generated the observed sequence of states, will be labelled as the 'realized segment'. Of course, there will be differences in confidence that a sequence X should be labelled as segment /x/.

A confidence index is a possible measure of the quality of the realized segment. This procedure is used in an ASR application for the detection of errors in L2 (Cucchiarini et al., 2009). There is a complication. In most speech recognition algorithms, a so-called 'language model' plays a role. That language model contains transitional probabilities of going from one word to another. In Dutch, for instance, the probability of finding a word with the neutral gender - like 'house' - is extremely low after the non-neutral determiner 'de' ('the'). In some ASR approaches language models even pertain to phoneme sequences. In Dutch, for instance, the sequence /l r/ is very low, whereas the probability of observing a sequence /s t/ is relatively high. Of course, this language knowledge should be used in speech recognition, but not in a system that aims at assessing the quality of speech segments. Knowledge of the language model is a well-known obstacle to the subjective assessment of speech. We all know that in 'the cat p...' 'p' will be followed with a high probability by 'urrs'. That is why subjective measurement has an intrinsic problem: the expectation of the listener.

HMMs are used in a number of formats, depending on:

- the number of states used in modelling speech segments,
- the amount of training material needed,
- the dimensionality of the statistical distributions of the parameters used.

A word-based account of errors is often not informative, even if the language model in the HMM is "switched off". The reason is that segment mispronunciations have to be weighted in order to obtain a valid error score for an utterance (Preston, Ramsdell, Oller, Edwards & Tobin, 2011). That is why most systems developed for providing feedback on the adequacy of pronunciation are segment based. Before a robust ASR system can be set up to provide feedback on speech performance, it has to be established to which extent target speech segments meet the following criteria, which, as a matter of fact, are quite similar to the criteria used in systems for error detection in second language acquisition (see Cucchiarini et al. 2009).

a) The influence of an incorrect realization has an impact on intelligibility and communication; this implies that for every language different segments and features are important. Tonal movements are less important for languages like English or German, but crucial in Chinese.
b) The errors are perceptually salient;
c) The errors are frequent;
d) The errors occur in the speech of relatively many speakers;
e) The errors are persistent;
f) Robust automatic error detection is possible;
g) Unambiguous feedback is available.

2.4.3 Speech materials to be used

An often neglected subject is the nature of the speech materials to be used. Of course, the materials should contain language samples which are prone to be incorrectly realized - see above -, but there are also other aspects which should be taken care of. There are a multitude of factors which affect the realization of speech segments and global parameters of speech. We mention: the prosodic position: in the English word rhododendron, for example, 'den' carries word stress. Word stress affects duration, intensity and spectral characteristics, the length of the utterance: the longer an utterance is, the shorter the speech segments which make it up are (the 'I' in 'stride' is shorter than in 'side', the distinction between function words and content words (for instance 'in' vs. 'bin') is influential in speech tempo, intensity and spectral characteristics. If speech materials are to be used in subsequent assessment procedures, it is worthwhile to have the speech segments realized in balanced conditions.

2.4.4 Technical requirements

The American Telemedicine Association published a valuable list of "Core Standards for Telemedicine Operations" (2007). For speech applications an additional number of technical requirements have to be met, as a function of the goals of the application. Most of them are self-evident. We mention a number of requirements, and will give some more information on requirements which are less self-evident:

1) *The presence of a reliable and robust server,* with personnel that can answer technical questions at well-defined time intervals and can update the system as required by IT-developments (firewalls, browsers etc.);
2) *A cross-platform browser-based application* which delivers uncompromised viewing of applications;
3) *A clear distribution of roles* with an associated system for authorizations: user, therapist and administrator;
4) *Quick uploading* of target utterances;
5) *A psychologically valid and quick presentation of feedback with sufficient screen resolution.*

The configuration of visual feedback for speech is not self-evident, and needs some scrutiny in order to adapt it to the user population. In this domain, speech technologists should be supplemented with experts in the integration of auditory and visual perception (Sadakata et al., 2008). For patients with neurogenic communication disorders such as PD, who are likely to suffer from other disorders than distorted speech alone, visual or cognitive distortions might be a serious constraint in the perception of visual displays that aim at providing feedback on different speech dimensions, such as pitch and intensity (loudness). This is particularly the case when two or more speech dimensions of a dysarthric speaker are displayed, such as pitch and intensity in the case of the PLVT for parkinsonian speakers. Rather than abstract graphs, which are difficult to interprete for a large number of predominantl elderly speakers, visualization should be simple and intuitive. That is, the form of visual feedback should apply to patients and give cues for approaching and adequate realization of speech. As an example, how should one display the time course of pitch and intensity, as a simple graph (see picture 2.3a), or as a picture which might intuitively be more appealing (picture 2.3b)? The solution might obviously lie in an integrated, multidimensional depicture of more speech dimensions.
Currently, in the Netherlands, web-based experiments are set up in order to evaluate what graphic form does appeal to healthy controls. In addition, it will be evaluated whether or not preferences for visuals forms in healthy controls also goes for neurological patients such as speakers with PD or after stroke.

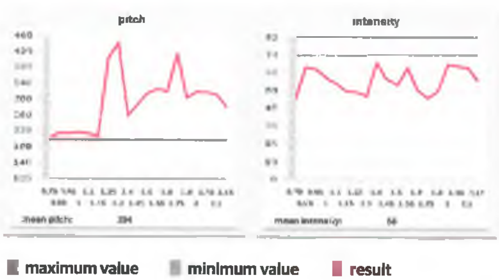


Figure 2.3a:

Separate displays of the time course of pitch and intensity

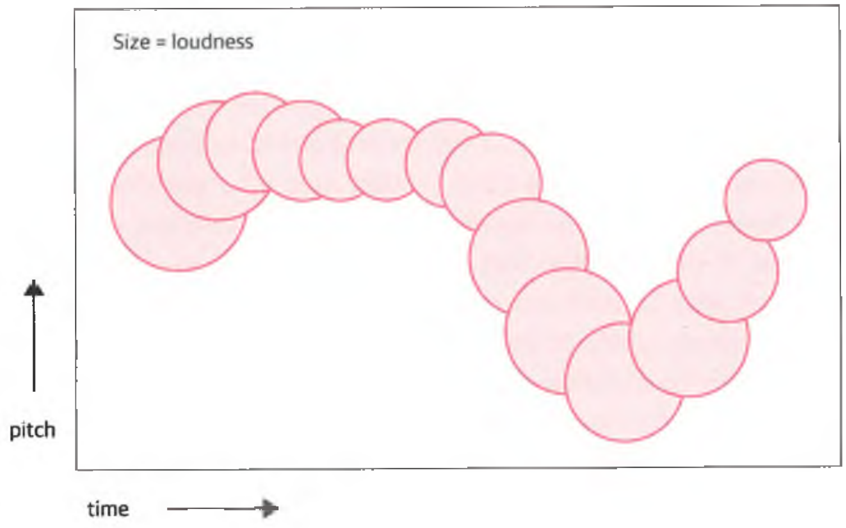


Figure 2.3b:

Integrated display of the time course of pitch and intensity

6) *Easy use of the PC/laptop;*
7) *Adjustable text fonts;*
8) *Possibility to personalize protocols for exercises;*
9) *Privacy guarantees;*
10) *Adequate format of speech files.*

If (subsets of) realized utterances are stored for subsequent assessment by a speech therapist or an automated computer procedure, the format of the speech files should be suited for those procedures. The main element of the format is the sampling frequency (22.05 kHz or 44.1 kHz). Preferably no signal coding should be used to reduce the amount of data. MP3 coding and the associated data reduction (with bit rates of 128 or 192) does not have any effects on perception, but may lead to some effects on the spectral representation. WAV-files have an advantage: they are files without any data-coding. Important factors in the decision on the sampling frequency and the possible data reduction are the characteristics of the parameters to be extracted from the signal. The upper bound of the frequency range relevant for the acoustic description of vowel-like sounds is around 3 kHz. For the analysis of fricatives - for instance speech sounds like /s/ and /S/ - we need a wider frequency range, with an upper bound of at least 6 kHz (Olive et al., 1993).

11) *Robust and well-defined recording conditions.*
The basic principle underlying e-health applications is that patients can use the application at home. Conditions at home vary to a great extent. Some people will use the application in a quiet office, others in a kitchen with neon tubes, or in a garden with traffic noise in the background. In the following figure we show the waveform of an utterance with an added 50 Hz-signal; the latter is not a pure sinus wave. The sinusoidal signal might hinder subjective judgement, and create biases in the spectral representation of the realized utterances, see panel (b) in Figure 2.4, where we see a strong 50 Hz component and the associated harmonics. The presence of the harmonics is due to the fact that the sinus wave was not 'pure'.

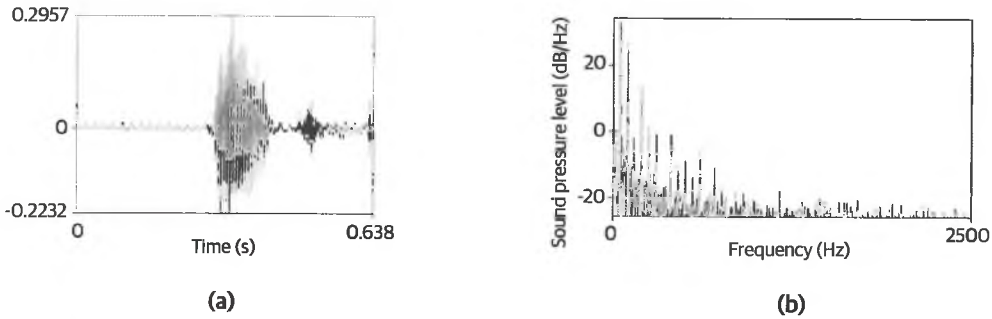


Figure 2.4:
(a) Waveform of a fragment of the Dutch word 'buitenboord' ([b↓ʏt↔b...]:'Englsh: 'outbord') with an added 50 Hz signal generated by an external source and b) spectrum of the (intended silent) initial part of the waveform.

In many approaches to the (semi-automatic) assessment of the segmental quality of speech segments, the silent interval associated with the closing phase of stop consonants like /p, t/ k/ is relevant. If this interval is filled with some 'humming' it may be an indication that the speaker was not able to firmly close his/her lips for the stop consonant (Kent et al., 1999).

2.4.5 Secondary outcomes of telehealth applications for speech technology

Apart from therapeutical aims, which mainly focus on the benefits for clients and therapists, telehealth applications such as EST provide a vital source of data for researchers in the field of speech pathology and speech technology. That is, uploading patients' speech by means of web-based systems, automatically generates a data base of pathological speech (Figure 2.6). This data base is vital for clinical outcome research in the field of speech pathology. That is, a data base allows perceptual and acoustical measurements of speech across time in order to evaluate therapy outcomes. This will increasingly gain importance in the context of decreasing financial resources for health care, where evidence based treatments finally will prevail. Although guidelines for diagnosis and treatment have been formulated (section 2.2.3), these are not based on the highest level of evidence. Objective outcome measures on the basis of a central data base of pathological speech is likely to enhance evidence based guidelines. Government policies and hence, requirements of health care reimbursers will be based on objective therapy results, to be derived from data sources as generated by web-based applications such as EST. For therapists the objective speech data over time of an individual patient is expected to provide useful information to evaluate therapy results and to adjust therapy focus if necessary.

In addition to its relevance for clinical outcome research, a data base of pathological speech contains vital information for speech technological and language technological research. In general, speech and language technology has considerably gained importance in health care during the last few decades. Particularly for patients with communicative problems this research area is of vital importance. These communicative problems can be due to cognitive disorders (e.g. aphasia or dyslexia), sensory disorders (blindness or hearing loss) or voice and speech disorders (e.g. dysartrhria, stuttering, dysphonia). Speech and language technology also applies to the needs of patients with communication problems in a broader sense. That is, constraints in the interaction with their environment as a result of motor system disorders such as Repetitive Strain Injury (RSI) or movement disorders such as paralysis after stroke or distorted arm movement coordination due to PD. Only recently, a Dutch report on needs and future possibilities for speech and language technologies for patients with communicative problems appeared (Ruiter et al., 2010).

Applications of speech and language technology are expected to contribute to more efficient and more effective health care. Patients can stay longer in their home environment without putting demands on health care givers and, hence, on financial resources. For example., patients with PD could benefit from speech synthesis applications for text-to-speech conversions, facilitating patients with severely diminished speech intelligibility in their verbal communication. Automatic error detection could provide parkinsonian patients who are eager to practice their speech in their home environment using web applications such as EST, with automatic feedback on segmental speech quality (i.e. articulation of speech sounds). An ASR application in dysarthric speech for example would lie in the field of domotica. PD patients with severe motor constraints could gain considerable independence from remotely (i.e. speech) controlled domestic equipment.

In general, speech synthesis, as applied in text-to speech conversions, is usually relatively simple and is not dependent on features of pathological speech. Automatic speech recognition (ASR) and automatic error detection of pathological speech however, are complex issues in the field of speech technology. This is primarily due to the large variability within and between pathological speakers, in particular in the case of neurogenic speech disorders such as dysarthria. Hence, large amounts of data are required for the development of ASR and automatic error detection in pathological (i.e. dysarthric speech). Applications of automatic error detections concern for instance feedback on segmental speech quality in EST, in addition to feedback on loudness and overall pitch. This would enhance patients' independent web-based speech training.

Obviously, apart from research in the field of speech technology, an automatically generated data base of pathological speech, is an essential source for additional fundamental research into acoustical features of parkinsonian dysarthria for instance. Outcomes of acoustical studies might even lead to adjustments of speech training programs for patients with PD.

A data base of pathological speech should contain speech at various linguistic levels. Audio recordings stored at the data base should be adequately annotated. That is, identification of standardized speech tasks, orthographic and phonetic annotation and linguistic level should be well documented. In addition, anonymous speaker identification should be ensured. An adequately structured data base should facilitate researchers' search for audio files of pathological speech. In Belgium, the Corpus of Pathological and Normal Speech (COPAS) (Middag et al., 2010) has been collected. Researchers employ the COPAS data base for the development of an automated intelligibility assessment, based on phonological features. These phonological features refer to articulatory dimensions. This information should reveal underlying articulation problems in dysarthric speakers. The

Nemours Data Base of Dysarthric Speech is another example of a corpus of pathological speech (Menendez-Pidal et al. 1996). It should be noticed however that the COPAS and the Nemours data bases were not generated by a web-based system, whereas a data base generated by means of EST involves upload and storage of audio files by means of a telehealth application. Vital conditions must be met to ensure audio recordings with adequate quality for perceptual and acoustical assessment. Obviously, a data base of pathological speech is language specific. Cross-language comparisons however should be enabled by similar structures of speech data bases for different languages.

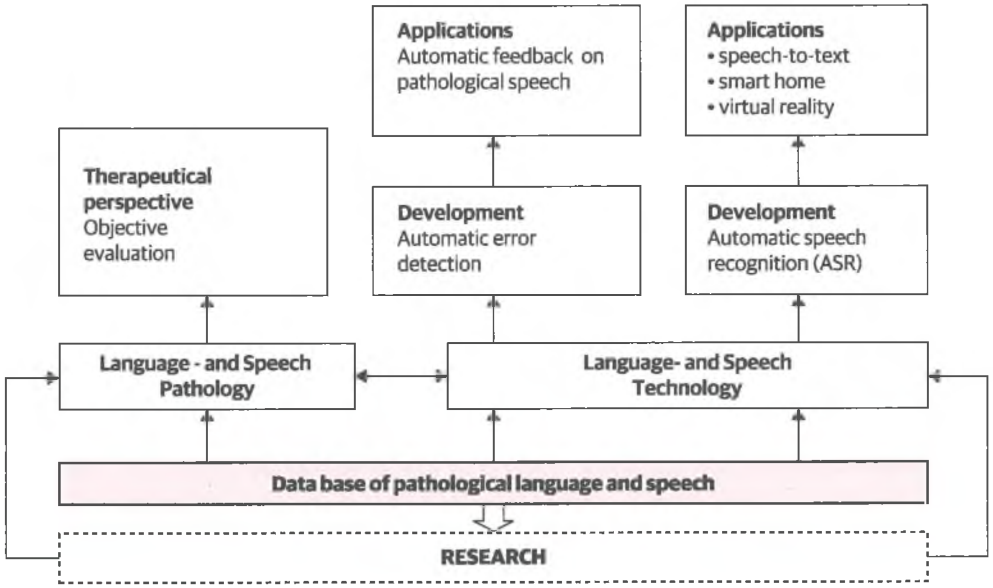


Figure 2.5:

A data base of pathological speech as a vital source for clinical outcome research and speech and language technological research.

2.5 Research areas with respect to the development, implementation and evaluation of telehealth applications for speech training of patients with Parkinson's disease

Telerehabilitation has a potential in many fields. We are still in the first phase of a development which may revolutionize medical care and cure. The heterogeneous applications – be they in psychiatry, asthma or diabetes care, speech and language therapy – share a number of factors which have to be fulfilled in order to warrant success, but which are not always met yet. A dangerous aspect of the phase we are in now is that we focus on technology and just admire its realized or promised possibilities. Thus we might overlook the key human factors for telerehabilitation applications in general as reviewed by Brennan and Barker (2008). In this section we give an overview of the four research areas formulated by the American Telemedicine Association (Krupinsky et al., 2007), and will zoom in on those aspects which are relevant for telehealth applications for people with Parkinson's disease.

a) *Attention should be paid to infrastructure definition and integration of various infrastructure components of web-based devices*

Of course, the definition of infrastructure and integration of various infrastructure components of web-based devices is a prerequisite for the application and evaluation of results obtained with web-based devices (see section 2.3.3). The central component of the applications in our field remains the availability of robust speech recognition and/or error detection systems, if at least providing automatic feedback on realized utterances in speech training is (one of) the goal(s) of the web application. Both speech and language pathology and speech and language technology are language bound. That is, they share underlying principles (HMMs, for instance, are used for very diverse languages like Chinese, English or Russian), and pathological reduction of vowels, consonants and pitch excursions occur in all languages, but the details of the technologies have to be developed for every specific language and the phenomena associated with speech pathology of specific languages have to be studied. In this context much attention has to be paid to the form of feedback, as was already pointed out in section 2.4.4.

Speech disorders and the associated symptoms may show considerable variability between and within speakers. While intensive training may help some patients to partially recover and improve their speech skills, other patients will show no improvement or perhaps even deterioration. Therefore, novel speech recognition and natural language processing techniques that can cope with the dynamics of the speech disorder, will have to be developed.

b) *Clinical utility of telehealth should be established*

The establishment of the clinical utility of a telehealth application for speech therapy of patients with PD is not a simple one. The first prerequisite is the presence of a robust infrastructure and a robust feedback system. While students using a feedback system on the quality of their pronunciation of a second language may be "robust" themselves, we cannot assume the same extent of robustness in patients with PD. That is why a very small number of studies have been conducted on the clinical utility of e-health applications in our field. In order to test the clinical utility in a phase III or phase IV studies, quite a number of conditions have to be fulfilled:

- A clear definition of the effect one wants to attain with the application. There are different possible and/or positive effects: (1) after a pre-specified time interval the speech quality of the telehealth group has more improved than that of the control group, (2) after the time interval the speech quality of the telehealth group was more stable than that of the control group, (3) in the long run, maintenance of achieved results is made possible by the telehealth application. Each of these possible outcomes has to be crossed with outcomes regarding the cost-effectiveness of the treatment (see under d) and user satisfaction.
- Relatively large numbers of patients are needed in order to ensure a high enough power to detect possible differences between a treatment and a control group. In light of the heterogeneity of the patients with respect to a large number of relevant background variables (age, SES, cognitive and motor skills, hearing and vision, computer skills, mobility, home situation), matching on these variables is a crucial issue in the effect studies to be carried out. As it is not very simple to include large numbers of patients, it is difficult to obtain a complete view over the importance of these variables (see under c).
- An important aspect of effect studies is a clear definition of the outcome variables to be used. For telehealth applications for patients with Parkinson's disease we mention five variables which are directly related to speech dimensions:

- Articulatory precision
- Intelligibility
- Naturalness of speech
- Speech effort
- Listening comfort

A lot of research is needed to find generally accepted operationalizations of the above-mentioned variables. Reviewing the relevant international journals in this domain makes clear that research is still going on, and that final results are not in view. This is in contrast to related questions on the intelligibility and naturalness of synthesized speech, where researchers agreed on a number of well-documented protocols to assess these aspects of computer speech (see the Blizzard Challenge, a yearly competition among speech synthesis systems based on corpora, see http://www.cs.cmu.edu/~awb/). For a review of problems encountered in subjective methods to assess intelligibility we refer to Beijer, Clapham & Rietveld (in preparation) and Hustad (2006).

Even if the correct operationalizations are available, a number of other questions have to be answered before ecologically valid effectiveness studies can be carried out. Here are two examples of the questions still to be answered: (1) Articulatory precision can be achieved at the cost of naturalness ("speak loud", the message of the Lee-Silverman therapy): what is more important: articulatory precision or naturalness? (2) Intelligibility is obviously related to articulatory precision, but to what extent can outcomes of current intelligibility tests and tests of articulatory precision be generalized to daily life?

c) *Human and ergonomic factors should be taken into account in research activities.*

At first sight the condition under c) is stating the obvious; however, enthusiasm for technology might obscure the importance of these factors. It is well known that diseases like PD may come along with other problems: comorbidity is not uncommon (Lowit et al., 2005). The problems may be such that a telehealth application is not suitable for a patient, neither in daily practice, nor in a research setting. For a research setting which aims at finding evidence for the effectiveness of an application itself - out of the social/psychological context - the problem may be that some of the participants are not suited to fully appreciate the application. There are a multitude of possible reasons for this; we mention impaired auditory processing and impaired vision as possible and obvious obstacles to the use of a telehealth application. In Chapter 4 we describe an Auditory Discrimination Test on a number of speech dimensions (Beijer, Rietveld, & van Stiphout,

2011) to assess participants' suitability to use auditory feedback.

There are also less obvious factors which should be taken into account, and which are often also region-determined. In densely populated areas like the Netherlands, with short distances, a patient has the choice to opt for face-to-face sessions or to stay at home with an eHealth application. A number of aspects may influence the choice: (1) finances - in the Dutch context hardly ever a factor for a patient, as even taxi expenses will often be reimbursed (2) the wish to see other people at a regular basis; this opportunity is provided by face-to-face sessions and not by a telehealth application (3) mobility; some people are home-bound, while others may be mobile, (4) the need to be intelligible for people others than direct partners.

d) *Economic analysis should point out whether the balance of costs and benefits is beneficial to the actual economic and social situation.*

This research area becomes an important one against the background of an ageing population and limited financial resources. A prerequisite for cost-effectiveness studies is the availability of effectiveness measures accepted by the community of speech and language pathologists and therapists. The question how to decide whether a number of beneficial units of web-based speech therapy (less effort for a listener, less repetitions needed to achieve complete understanding, less absence from home, less transport, better maintenance of communication etc.) are in a positive trade-off relation with additional costs (implementation and maintenance of infrastructure, availability of a help-desk etc.) is a matter of politics and society.

2.6 Telehealth for speech therapy in stroke

It will be obvious that the potentials of web-based speech training such as EST also apply to dysarthrias caused by stroke. Although speech features in spastic, flaccid, ataxic or mixed dysarthrias — depending of the site and the extent of the lesion caused by a stroke — differ from those commonly observed in patients with in hypokinetic dysarthrias due to PD (Enderby, Cantrell, John, Pickstone, Fryer & Palmer, 2010), emphasis on adequate voice intensity and, directly or indirectly, on improved articulation, are general interventions in speech therapy. Therefore, the role of speech technology and the research areas with regard to telehealth applications as discussed in the sections 2.4 and 2.5., also apply to EST for stroke patients with dysarthric speech.

2.7 Conclusion

Employment of telehealth devices for dysarthric patients with PD seems promising with respect to their possibilities to practice their speech independently. As such, they might provide a solution for the foreseen imbalance between the need (and call) for speech training on the one hand, and the reduced availability of therapeutical resources on the other hand. Although technological feasibility of various web-based training devices has been established, user requirements for parkinsonian patients with frequently observed deficits in cognitive and motor functioning demand further adjustments. Apart from therapeutical goals, web-based training devices such as EST provide the possibility of generating a data base of pathological speech. This data base not only provides required information for clinical outcome research. It is also of vital importance for the development of automatic speech recognition and automatic error detection of pathological (i.e. parkinsonian) speech.

Acknowledgements
The authors wish to thank Dike Ottoy for her corrections of the english text and Gaetano Ambrosino for his assistance with the figures.

References

Ackermann, H. & Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery & Neuropsychiatry, 54 (12),* 1093-1098.

American Telemedicine Association. http://atmeda.org/news/definition.html. Last accessed 20.2.2011.

Beijer. L. Clapham, R. & Rietveld, T. Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease. In preparation.

Beijer, L., Rietveld, T., van Beers, M, Slangen, R., van den Heuvel, H., de Swart, B. & Geurts, S. (2010a). E-learning based Speech Therapy (EST) as a web application for speech training. *Telemedicine and e-Health, 16(2),* 177-180.

Beijer, L., Rietveld, T., Hoskam, V., Geurts, A. & de Swart, B. (2010b). Evaluating the feasibility and the potential efficacy of E-learning-based Speech Therapy for speech training in dysarthric patients with Parkinson's Disease: A aase study. *Telemedicine and e-Health, 16(6),* 732-738.

Beijer. L., van Stiphout, A. & Rietveld, A. (2011). Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44*, 701-718.

Blizzard Challenge: http://www.cs.cmu.edu/~awb/.

Boersma, P. & Weenink, D. (2011). *Praat: Doing Phonetics by Computer.* www.praat.org. Downloaded 8.3.2011.

Brennan, D.M. & Barker, L.M. (2008). Human factors in de the development and implementation of telerehabilitation systems. *Journal of Telemedicine and Telecare,14,* 55-58.

Carmichael, J.N. (2007). Introducing Objective Acoustic Metrics for the Frenchay Dysarthria Assessment Procedure. Unpublished PhD-thesis, University of Sheffield.

Cucchiarini, C., Neri, A., Strik, H. (2009). Oral Proficiency Training in Dutch L2: the Contribution of ASR-based Corrective Feedback. *Speech Communication, 51 (10),* 853-863.

Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP, 28,* 357-366.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*, 246-269.

Deane, K.H.O., Whurr, R., Playford, E.D., Ben-Shlomo, Y., Clarke, C.E. (2001). Speech and language therapy versus placebo or no intervention for dysarthria in Parkinson's disease. Cochrane *Database of Systematic Reviews, Issue 2.* Art. No.: CD002814.

Duffy, J.R. (1995). *Motor Speech Disorders: Substrates, differential diagnosis and management.* St. Louis: Mosby-Year Book

Enderby, P. (1980). Frenchay Dysarthria Assessment, *British Journal of Disorders of Communication, 21,* 165-173.

Enderby, P., Cantrell, A., John, A., Pickstone, C., Freyer, K. & Palmer, R. (2010). Guidance for providers of speech and language therapy services: dysarthria. *Asia Pacific Journal of Speech, Language, and Hearing,* 13 (3), 171-190.

Ferrand, C. & Bloom, R. (1997). *Introduction to Organic and Neurogenic Disorders of Communication.* Needham Heights, MA: Allyn and Bacon.

Ho, A.K., Bradshaw, J.L. & Iansek, R. (2000).Volume perception in parkinsonian speech. *Movement Disorders* 15 (6), 1125-1131.

Hustad, K.C. (2006). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatrica et Loopaedica, 58,* 217-228.

Kalf, J.G., de Swart BJM, Zwarts MJ, Munneke M, Bloem BR. (2008a). Frequency of oral motor impairments in Parkinson's disease and implications for referral to speech therapists. *Movement Disorders*;23:S328.

Kalf, J.G., de Swart BJM, Bloem BR, Munneke M. Guidelines for speech-language therapy in Parkinson's disease (2008b). *Movement Disorders;*23:S328.

Kent, R.D., Weismer, G., Kent, J.F., Vorperian, H.K. & Duffy, J.R. (1999). Acoustic studies of dysarthric speech: methods, progress, and potential. *Journal of Communication Disorders*, 32, 141-186.

Krupinsky, E, Dimmick, S., Grigsby, J., Mogel, G., Puskin, D., Speedie, S., Stamm, B., Wakefield, B., Whited, J., Whitten, P., Yellowlees, P. Research recommendations for the American Telemedicine Association. *Telemedicine and e-Health*, 12(5), 579-589.

Kwakkel, G., Wagenaar, R.C., Twisk, J.W.R., Lankhorst G.J., Koetsier, J.C. (!999). Intensity of leg- and arm-training after primary middle-cerebral- artery stroke. *Lancet*, 354 (9174), 191-196.

Lowit, A., Howell, P. & Brendel, B. (2005). Cognitive impairment in Parkinson'disease. Is it a unified phenomenon? *Journal of Brain Impairment*, 6 (3), 191-204.

Mashima, P. & Doarn, C. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health*, 14, 1101-1109.

Menéndez-Pidal, X., Polikoff, J.B., Peters, S.M., Leonzio, J.E. & Bunnell, H.T. (1996). The Nemours Database of Dysarthric Speech. *Proceedings of the Fourth International Conference on Spoken Language Processing, October 3-6,* Philadelphia, PA, USA.

Middag, C., Saeys, Y. & Maryens, J-P. (2010). Towards an ASR-free objective analysis of pathological speech. *Proceedings of Interspeech 2010,* Makuhari, Japan, 294-297.

Neri, A., Cucchiarini, C. & Strik, H. (2006). Selecting segmental errors in L2 Dutch for optimal pronunciation training. *International Review of Applied Linguistics*, 44, 357-404.

Olive, J.P., Greenwood, A. & Coleman, J. (1993). *Acoustics of American English Speech.* New York: Springer Verlag.

Parker, M., Cunningham, S., Enderby, P., Hawley, M. & Green, P. (2006). Automatic speech recognition and training for severely dysarthric users of assistive technology: The STARDUST project. *Clinical Linguistics and Phonetics*, 20, 149-156.

Preston, J.L., Ramsdell, H.L., Oller, D.K, Edwards, M.L. & Tobin, S.J. (2011). Developing a Weighted Measure of Speech Sound Accuracy. *Journal of Speech, Language, and Hearing Research*, 54, 1-18.

Ramig, L.O., Sapir, S. Countryman, S., Pawlas, A.A., O'Brien, C., Hoehn, M. & Thompson, L.L. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow up. *Journal of Neurology Neurosurgery and Psychiatry*, 71, 493-498.

Rietveld, T. & Stoltenberg, I. (2005). *Taal- en spraaktechnologie en communicatieve beperkingen.* Rapport uitgebracht in opdracht van de Nederlandse Taalunie ('Language and Speech Technology for people with Communication Disorders', in Dutch).

Rietveld, T. & van Heuven, V.J. (2009). Algemene Fonetiek. Bussum: Coutinho, 3rd edition ('General Phonetics', in Dutch).

Rijntjes, M., Haevernick K., Barzel A., van den Bussche H., Ketels G., & Weiller C. (2009). Repeat therapy for chronic motor stroke: a pilot study for feasibility and efficacy. *Neurorehab Neural Repair.* 23(3), 275-80.

Ruiter, M.B., Rietveld, A.C.M., Cucchiarini, C., Krahmer, E.J. & Strik, H. (2010). Human Language Technology and Communicative Disabilities: Requirements and Possibilities for the Future in Dutch. *Proceedings 7th International Conference of Language Resources and Evaluation,* LREC, 2839-2846.

Sadakata M., Hoppe D., Brandmeyer A., Timmers R. & Desain P. (2008). Real-Time Visual Feedback for Learning to Perform Short Rhytms with Variations in Timing and Loudness. *Journal of New Music Research,* 37, 207-220.

Sanders, E. Ruiter, E., Beijer, L. & Strik, H. (2002). Automatic recognition of Dutch dysarthric speech: A pilot study. *Proceedings of ICSLP-2002,* Denver, USA, 661-664.

Slawek, J., Derejko, M., & Lass, P. (2005). Factors affecting the quality of life of patients with idiopathic Parkinson's disease – a cross-sectional study in an outpatient clinic attendees. *Parkinsonism & Related Disorders,* 11 (7), 465-468

Sood, S., Mbarika V., Jugoo, S., Dookhy, R., Doarn, C.R., Prakash, N., & Merrell, R.C. (2007). What is telemedicine? A collection 104 peer-reviewed perspectives and theoretical underpinnings. *Telemedicine and e-Health,*13 (5), 573-590.

Strik, H., Truong, K., de Wet, F., Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication,* 51 (10), 845-852.

Swart, B.J.M. de, Willemse, S.C., Maassen, B.A.M. & Horstink, M.W.I.M. (2003). Improvement of voicing in patients with Parkinson's disease by speech therapy. *Neurology,* 60, 498-500.

Streifler, M. & Hofman, S. (1984).Disorders of verbal expression in parkinsonism. *Advances in Neurology,* 40, 385-393.

Teasell, R.W. & Kalra, L. (2004). Advances in stroke 2003. What's new in stroke rehabilitation? *Stroke,* February, pp. 383-385.

Von Campenhausen S, Bornschein B, Wick R, Botzel K, Sampaio C, Poewe C, et al. (2005). Prevalence and incidence of Parkinson's disease in Europe. *European Neuropsychopharmacology,* 15(4), 473-490.

Theodoros, D.G., Constantinescu, G., Russell, T.G., Ward, E.C., Wilson, S.J.& Wootton, R.(2006). Treatin the speech disorder in Parkinson's disease online. *Journal of Telemedicine and Telecare,* 12: S3, 99-91

Tulu, B. Chatterjee, S. and Maheshwari, M. (2007). Telemedicine Taxonomy: a classification tooi. *Telemedicine and e-Health,* 13 (3), 349-358.

Ziegler, W. (2003). Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech. *Aphasiology,* 17 (1), 3-36.

Ziegler, W. & Zierdt, A. (2008). Telediagnostic assessment of intelligibility in dysarthria: A pilot investigation of MVP-online. *Journal of Communication Disorders,* 41(6), 553-577.

CHAPTER 3

E-learning based Speech Therapy (EST): a web application for speech training

L.J. Beijer, A.C.M. Rietveld, M.M.A. van Beers, R.M.L. Slangen, H. van den Heuvel, B.J.M. de Swart, A.C.H. Geurts

Adapted from: *Telemedicine and e-Health, 2010, 16(2): 177-80*

Abstract

In the Netherlands, a web application for speech training, E-learning based Speech Therapy (EST), has been developed for patients with dysarthria, a speech disorder resulting from acquired neurological impairments such as stroke or Parkinson's disease. In this report, the EST infrastructure and its potentials for both therapists and patients are elucidated. EST provides patients with dysarthria the opportunity to engage in intensive speech training in their own environment., in addition to undergoing the traditional face-to-face therapy. Moreover, patients with chronic dysarthria can use EST to independently maintain the quality of their speech once face-to-face sessions with their therapist have been completed. This telerehabilitation application allows therapists to remotely compose speech training programs tailored to suit each individual patient. Moreover, therapists can remotely monitor and evaluate changes in the patient's speech. In addition to its value as a device for composing, monitoring and carrying out web-based speech training, the EST system compiles a database of dysarthric speech. This database is vital for further scientific research in this area.

3.1 Introduction

In the past decades, considerable attention has been paid to the definition of telemedicine (Sood et al., 2007) and to the evaluation of telemedicine technologies (DeChant, Tohme, Mun, Hayes, & Schulman, 1996). The effect of home-based telehealth on clinical care outcomes (Dellifraine & Dansky, 2008) and users' attitudes toward diverse applications of home telehealth have also been investigated (Piron et al., 2009; Giansanti, Tiberi, Silvestri, & Maccioni, 2009). The potentials of telehealth in speech-language pathology have been emphasized only quite recently (Hill & Theodoros, 2002; Theodoros, 2008). Applications for remotely assessing speech and language disorders have been explored (Theodoros, Russell, Hill, Cahill, & Clark, 2003; Hill et al., 2006; Ziegler & Zierdt, 2008; Theodoros, Hill, Russell, Ward, & Wootton, 2008), and the potentials of telerehabilitation for treating people with communication disorders following acquired neurological impairments have been increasingly studied (Brennan, Georgeadis, & Baron, 2002; Theodoros et al., 2006). There is a need for telerehabilitation applications that enable these patients to engage in independent training in their home environment, (Theodoros, 2008) facilitating intensive training. Highly intensive training has been found to be effective in both motor rehabilitation after stroke (Kwakkel, Wagenaar, Twisk, Lankhorst, & Koetsier, 1999) and speech and language rehabilitation after stroke and Parkinson's disease (Ramig et al., 2001; de Swart, Willemse, Maassen, & Horstink, 2003; Bhogal SK, Teasell, & Speechley, 2003; Teasell & Kalra, 2004).

In the Netherlands, a web-based speech training device, 'E-learning based Speech Therapy' (EST) (available at http://www.spraaktraining.nl/), has been developed for patients with dysarthria following acquired neurological impairments. This communication describes the EST infrastructure and aims at elucidating the potentials of this telerehabilitation application.

3.2 Materials and methods

3.2.1 EST central server

A central server forms the keystone of the EST infrastructure (Figure 3.1). The server hosts two types of recorded speech audio files: target speech files in MP3 format and recorded speech files uploaded by patients (clients) in WAV format. The latter (semi-)automatically create a central database. Connected to the central server, the speech processing program 'Praat' (Boersma & Weenink, 2009) enables analyses of overall pitch and intensity of speech audio files uploaded by clients.

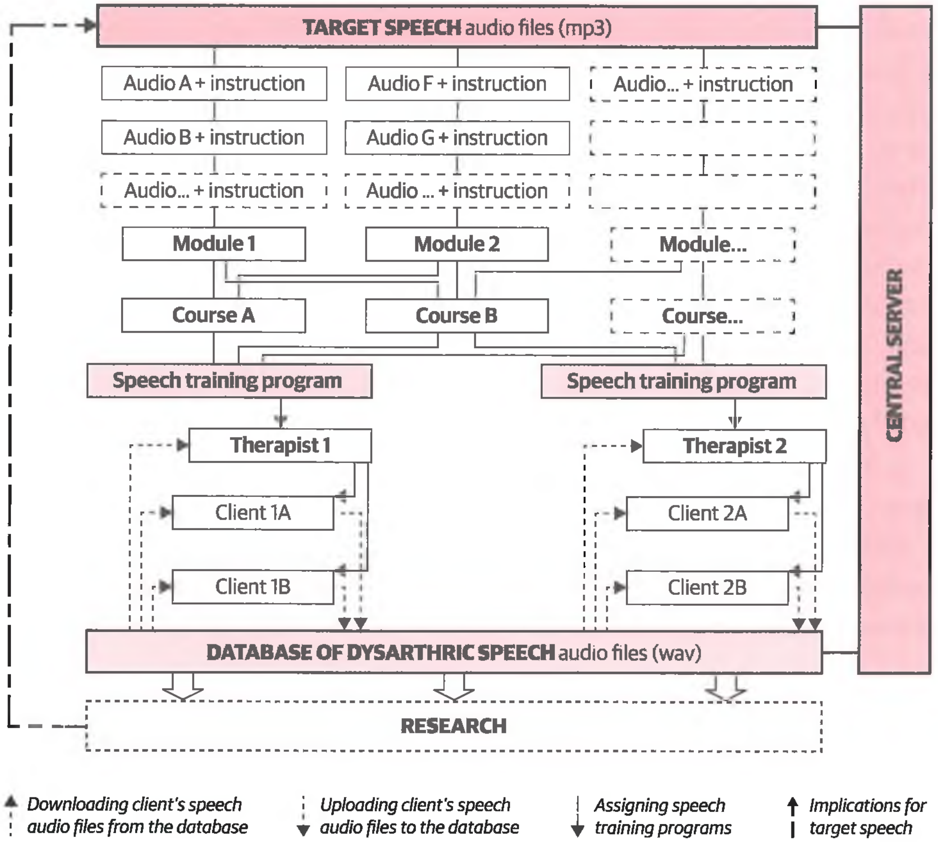


Figure 3.1:

Infrastructure of the EST system

3.2.2 Application requirements

A desktop computer or laptop with an internet connection (bandwidth at least 256 Kbit/sec) provides users with access to the server. Clients' computers require adobe shockwave player and adobe flash player, as the current EST client application uses components from these software packages. The flash application is involved in the user interface and communication with the shockwave component. The shockwave player serves as an audio recorder and file transfer protocol. This file transfer protocol facilitates file exchange between computers by standardizing procedures that differ between operating systems.

3.2.3 Architecture of EST speech training programs

A speech training program consists of one or more courses. Each course contains at least one training module. Each module has a fixed combination of audio files and instructional text to carry out the exercises. The training programs may be modified by changing the combination of courses and modules.

3.3 Results

3.3.1 Users

Administrators

Administrators supply the prerequisites for therapists and clients to operate the system. They assign user accounts to therapists and clients, and they are authorized to edit general and instructional text throughout the EST navigation system. Administrators build training modules, and establish a module's 'passive' or 'active' mode. A passive mode can temporarily prevent therapist access to a module (for instance when a module is under construction). Finally, administrators create test modules for purposes of speech assessment and evaluation. A test module contains text associated with the speech stimuli for the client to produce, record, and upload.

Speech therapists

Therapists are authorized to compose tailor-made speech training programs and to adjust the active or passive mode of each course in the program. Moreover, they assign target values to the speech parameters 'overall pitch' and 'intensity' for each client. To evaluate a client's speech, the therapist can select the 'analysis' function to compare the target and the realized values for these parameters. Clients' speech files can be downloaded using Praat for further acoustical analyses.

Clients

Clients have access to their individual speech training program composed by their therapist. For each client, the training procedure consists of the following steps:

1) Selecting a training item from the prescribed course.
2) Listening to a target speech sample, downloaded from the central server, using a headset.
3) Recording the imitation attempt via the headset's microphone connected to the laptop or computer.
4) Comparing the target and the recorded sample by selecting 'compare'. This comparison is based on auditory feedback only.
5) Deciding whether the speech attempt will be uploaded to the central server by selecting the 'save' button.
6) Once the recorded speech has been uploaded, a client's auditory discrimination is supported by automatic visual feedback for intensity and overall pitch.
7) Determining whether a new speech attempt is required to approach the target speech.

3.3.2 Speech database

A database of dysarthric speech is (semi-)automatically compiled by uploading clients' speech to the central server. The database contains speech audio files uploaded by clients during their training sessions as well as during test modules. The test modules provide realizations of standardized speech materials. Each uploaded audio file is accompanied by an XML file containing the target values of pitch (Hz) and intensity (dB), and another XML file containing realized values of pitch and intensity as a function of time.

The database provides therapists with centrally stored speech recordings produced by individual clients, thus facilitating objective evaluation of therapy effects. An additional benefit of this database is that large amounts of dysarthric speech become available for further scientific research in the field of speech pathology and speech technology.

3.4 Discussion

This report shows the potentials of EST for telerehabilitation. The technical feasibility of the described EST system has already been verified empirically. Unlike traditional face-to-face therapy, EST provides less opportunity for clinical observations and therapeutic recommendations to improve clients' speech. This might seem inconsistent with the American Speech-Language Hearing Association standard for telehealth, which states that the quality of services delivered via telepractice must be consistent with that of face-to-face delivery (American Speech-Language Hearing Association, 2005). However, the aim of EST is not to substitute, but to support and intensify speech regular speech training. In addition, EST enables patients with chronic dysarthria to independently maintain speech quality after completing their regular therapy. With the aging population in mind, EST might provide a solution for the maintenance of a sound balance between the need for, and the availability of, speech training for patients with chronic dysarthria.

3.5 Future directions

The efficacy and cost-effectiveness of EST, as well as users' satisfaction, need to be evaluated. To this end, clinical studies including dysarthric patients with various neurological impairments are in progress.

Acknowledgements

The development of the EST system was funded by the Sint Maartenskliniek (Department of Research Development and Education) in Nijmegen, the Netherlands, and by the ICT Innovation Fund from Radboud University Nijmegen, the Netherlands. The authors wish to thank Renee Clapham for her comments on the English text and Gaetano Ambrosino for his assistance with the figure depicting infrastructure included in this article.

References

American Speech-Language Hearing Association (2005). Speech-Language Pathologists Providing Clinical Services via Telepractice: Position Statement [position statement 2005. Available at www.asha.org/teleparctice (last accessed August 24, 2009).

Bhogal SK, Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, 34, 987-993.

Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.13) [computer program]. Available at www.praat.org/ (last accessed August 21, 2009).

Brennan, D., Georgeadis, A., & Baron, C. (2002). Telerehabilitation tools for the provision of remote speech-language treatment. *Topics in Stroke Rehabilitation, 8*, 71-78.

DeChant, H. K., Tohme, W. G., Mun, S. K., Hayes, W. S., & Schulman, K. A. (1996). Health systems evaluation of telemedicine: A staged approach. *Telemedicine Journal, 2*, 303-312.

Dellifraine, J. L. & Dansky, K. H. (2008). Home-based telehealth: a review and meta-analysis. *Journal of Telemedicine and Telecare, 14*, 62-66.

Giansanti, D., Tiberi, Y., Silvestri, G., & Maccioni, G. (2009). Toward the integration of novel wearable step-counters in gait. *Telemedicine and e-Health, 15*, 105-111.

Hill, A. & Theodoros, D. (2002). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare, 8*, 187-196.

Hill, A. J., Theodoros, D. G., Russell, T. G., Cahill, L. M., Ward, E. C., & Clark, K. M. (2006). An internet-based telerehabilitation system for the assessment of motor speech disorders: a pilot study. *American Journal of Speech-Language Pathology, 15*, 45-56.

Kwakkel, G., Wagenaar, R. C., Twisk, J. W. R., Lankhorst, G. J., & Koetsier, J. C. (1999). Intensity of leg and arm training after primary middle-cerebral-artery-stroke: a randomised trial. *Lancet, 354*, 191-196.

Piron, L., Turolla, A., Tonin, P., Piccione, F., Lain, L., & Dam, M. (2009). Satisfaction with care in post-stroke patients undergoing a telerehabiliation programma at home. *Journal of Telemedicine and Telecare, 14*, 257-260.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry, 71*, 493-498.

Sood, S., Mbarika, V., Jugoo, S., Dooky, R., Doarn, C., Prakash, N. et al. (2007). What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemedicine and e-Health, 13*, 573-590.

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60*, 498-500.

Teasell, R. W. & Kalra, L. (2004). What's new in stroke rehabilitation? Advances in Stroke 2003. *Stroke, 35*, 383-385.

Theodoros, D. (2008). Telerehabiliation for service delivery in speech-language pathology. *Journal of Telemedicine and Telecare, 14*, 221-224.

Theodoros, D., Constantinescu G., Russell, T., Ward, E., Wilson, S., & Wootton, R. (2006). Treating the speech disorder in Parkinson's disease online. *Journal of Telemedicine and Telecare, 12*, 88-91.

Theodoros, D., Hill, A., Russell, T., Ward, E., & Wootton, R. (2008). Assessing acquired language disorders via the Internet. *Telemedicine and e-Health, 14*, 552-559.

Theodoros, D., Russell, T. G., Hill, A., Cahill, L., & Clark, K. (2003). Assessment of motor speech disorders online: a pilot study. *Journal of Telemedicine and Telecare, 9*, S66-S68.

Ziegler, W. & Zierdt, A. (2008). Telediagnostic assessment of intelligibility in dysarthria: A pilot investigation of MVP-online. *Journal of Communication Disorders*, 553-577.

PART 2

PREREQUISITES FOR EVALUATING EST

CHAPTER 4

Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers.

L.J. Beijer, A.C.M. Rietveld, A.J.L. van Stiphout

Adapted from: *Journal of Communication Disorders*, 2011, 44: 701-718.

Abstract

Background: Web-based speech training for dysarthric speakers, such as E-learning based Speech Therapy (EST), puts considerable demands on auditory discrimination abilities.

Aims: To discuss the development and the evaluation of an auditory discrimination test (ADT) for the assessment of auditory speech discrimination skills in Dutch adult dysarthric speakers as a prelude to EST.

Method: Five ADT subtests were developed, each addressing a vital speech dimension in speech therapy: articulation (segmental elements), intensity, overall pitch, speech rate and intonation. A healthy control group of 36 participants performed a 'same-different task' in each subtest. ADT items yielding scores of at least 80% but below 100% correctly responding healthy controls were considered sensitive to diminished auditory discrimination. Subsequently, the ADT was carried out by 14 neurological patients with dysarthric speech and 14 matched healthy controls. Score percentages, sensitivity indices and reaction times (ms) on only sensitive items were compared.

Results: The majority of the ADT items met the 'minimal 80% - below 100% criterion' in the healthy control group. The neurological participants performed lower on all outcome measures across all subtests than the healthy controls, although not all of these differences achieved statistical significance.

Conclusions: The results of the healthy control group show that the majority of the ADT items meet our criterion for sensitivity to diminished auditory discrimination. The poorer performance of dysarthric patients across all subtests supports the sensitivity of the ADT. However, further research involving larger and more homogeneous groups of neurological patients is required.

4.1 Introduction

One of the requirements for speech training is the ability to aurally discriminate between distinct (i.e. correct and incorrect) realisations of speech (Palmer, Enderby, & Cunningham, 2004; Rvachew, 2004). This especially holds for web-based speech training as it puts considerable demands on independent auditory discrimination. E-learning based Speech Therapy (EST) is an example of a web application for speech training. EST has recently been developed for Dutch patients with dysarthric speech following acquired neurological impairment (Beijer, Rietveld, Hoskam, Geurts, & de Swart, 2010a; Beijer et al., 2010b). Being one of the applications in the growing field of telehealth in speech-language pathology (Hill & Theodoros, 2002; Mashima & Doarn, 2008), EST provides neurological patients with the opportunity to intensively train their speech in their home environment. In addition, this web application enables them to independently maintain the achieved speech quality (i.e. in the absence of their speech therapist) once traditional face-to-face therapy has ceased. In order to benefit from EST, dysarthric speakers must be able to differentiate between their own speech and the target speech. That is, the EST speech training procedure requires auditory discrimination judgements between the target speech and the patient's own speech, which results from the patient's attempt to imitate the target. Despite the vital role of auditory discrimination in web-based speech training, a tool that assesses this ability for speech dimensions relevant to speech training does not exist for Dutch and most other languages.

In general, although telerehabilitation applications in various fields of healthcare have been explored for some decades now (Dellifraine & Dansky, 2008; Piron et al., 2009), little attention has been paid to the issue of minimum user requirements. This is remarkable since, obviously, for telepractice devices to be effective, user requirements must be considered. For computer use in general, a valid and comprehensive assessment of patients' abilities is vital to establish whether assistive technology is warranted (Hoppestad, 2006). That is, frequently observed symptoms in neurological patients such as poor vision, diminished attention, delayed information processing, impaired motor control and diminished comprehension may adversely affect a person's ability to interact with telerehabilitation equipment. Additional to general user conditions, in the case of web-based speech training such as EST sufficient auditory discrimination abilities in speech are required. Therefore, assessing auditory speech discrimination before enrolling into EST might prevent dysarthric patients with inadequate auditory discrimination abilities from disappointing experiences or premature therapy abandonment. This is particularly vital for patients with dysarthric speech following acquired neurological impairment, since they are at risk for auditory processing disorders (Häusler & Levine, 2000; Grawemeyer, Cox, & Lum, 2000; Ilvonen et al., 2001; Bamiou, Musiek, & Luxon, 2001; Bamiou, Musiek, & Luxon, 2003; Bamiou et al., 2006; Guehl et al., 2008; Bamiou, 2008). That is, as the acoustic information of speech signals travels within the central auditory nervous system, the processing of the signal undergoes several levels of serial and parallel 'bottom-up' processing influenced by high level cognitive 'top-down' processes. Because of the multiple locations of the operational sites for acoustic processing, central auditory processing disorders in neurological patients vary and can also be quite subtle (Demanez & Demanez, 2003). Diminished auditory discrimination abilities may be one of the behavioral phenomena resulting from the impaired process of analyzing intensity, pitch and temporal features of incoming acoustic speech signals. Another risk factor for diminished auditory processing (e.g. discrimination) abilities in patients with acquired neurological impairment, is the fact that the majority are of an elderly age, tending to go along with increased hearing loss and cognitive decline (Willott, 1991; Rooij van & Plomp, 1991; Pichora-Fuller & Souza, 2003; Golding, Carter, Mitchell, & Hood, 2004; Golding, Taylor, Cupples, & Mitchell, 2006; Pichora-Fuller & Singh, 2006b). Obviously, elderly people with neurological diseases may therefore experience additional difficulties in performing auditory discrimination tasks .

In our study, we conducted two experiments: one that concerns the development of an assessment tool for auditory speech discrimination and one that evaluates how people with acquired neurological impairment perform differently to healthy controls on the developed assessment tool. The rationale for developing an auditory speech discrimination test was that we needed a test which focuses on the tasks implied by the web-based speech training through EST. Neither the common audiological screening nor more elaborate and demanding tests on auditory processing disorders (Neijenhuis, Stollman, Snik, & Van der Broek, 2001) are suited for the neurological patients at issue.

In this paper we describe the development of an auditory discrimination test (ADT) for the purpose of assessing auditory speech discrimination skills in Dutch dysarthric patients as a prelude to EST. It was assumed that test items with scores above chance level and below ceiling level in a healthy control group are sensitive to diminished auditory speech discrimination. We hypothesized that patients with dysarthric speech following acquired neurological impairment will show diminished performance on the

ADT compared to matched healthy controls. In our study two research questions were addressed:

1) Which items in the auditory discrimination test meet the criterion of minimally 80% and below 100% score in a healthy control group?
2) Do neurological patients perform poorer on the auditory discrimination task than healthy controls?

In the next sections two experiments will be reported. Ethical approval was obtained by the Dutch Committee of Medical and Health Research Ethics.

4.2 Experiment 1

In Experiment 1 the ADT was developed. The aim of Experiment 1 was to develop the ADT and identify suitable test items with scores above chance level (50%) and below ceiling level (100%) in a healthy control group. Items that met this criterion were considered to be sensitive to diminished auditory discrimination (explained in section 4.2.1).

4.2.1 Materials and methods

Given the clinical purpose of assessing auditory speech discrimination abilities in patients who plan to enroll in web-based speech training through EST, some vital conditions were taken into account for the development of the ADT.

Firstly, conditions for practical use were considered. Therefore, the ADT equipment needed to consist of a portable laptop and a headset to allow 'mobile use'. In addition, since patients with various backgrounds (e.g. different education levels, literacy levels and/or computer experience) will perform the test, the test procedure had to consist of a simple 'same-different' task. Moreover, the duration of the test had to fit within a speech therapy session of 30 minutes, implying a limited number of test items.

Secondly, the ecological validity of the ADT needed to be taken into account. That is, the auditory discrimination tasks in the ADT should represent the auditory discrimination tasks in speech training procedures. Primarily being a prelude to web-based speech training, the ADT needed address the same speech dimensions as those to be trained in order to improve speech intelligibility: articulation (segmental speech elements), intensity (loudness), overall pitch, speech rate and intonation. With the exception of 'overall pitch', researchers frequently report the contribution of these dimensions to the intelligibility of speech (Metz, Samar, Schiavetti, Sitler, & Whitehead, 1985; Maassen & Povel, 1985; Yorkston, Hammen, Beukelman, & Traynor, 1990; Metz, Samar, Schiavetti, & Sitler, 1990; Kent, 1992; Laures & Weismer, 1999; De Bodt, Huici Hernandez-Diaz, & Van de Heyning, 2002; Rietveld & van Heuven, 2009b). Auditory discrimination of 'overall pitch' was included in the ADT since this speech dimension is relevant for voice hygiene reasons in the speech intelligibility training programs provided by EST. In particular, 'overall pitch' is a vital component of the Pitch Limiting Voice Treatment (PLVT) for patients with Parkinson´s disease (Swart de, Willemse, Maassen, & Horstink, 2003). The PLVT is a further development of the Lee Silverman Voice Treatment (Ramig et al., 2001). For each of the five speech dimensions, a separate auditory discrimination subtest has been developed. The speech materials in the ADT contain both words and sentences, similar to the commonly employed linguistic levels in speech training.

The third condition concerned a methodological issue in the ADT development. That is, in order to differentiate between normal and diminished auditory discrimination abilities, each subtest of the ADT had to contain items that are sensitive to diminished auditory discrimination. To this end a reference score of 80% was established for 'normal performance'. In fact, we used the 80% criterion three times: once to establish the Just Noticeable Difference (JND) per item, and twice to establish the number of correct items per subtest and the number of healthy controls needed to identify suitable items.

The 80% score applied to each separate item was based on the frequently used criterion applied in procedures to establish the Just Noticeable Difference(JND) of changes in the physical characteristics of a stimulus: detectable 75% of the time (Durrant & Lovrinic, 1977). As the percentage of 75 is often used in laboratory conditions, we decided to use a slightly higher percentage for the clinical purpose of the ADT and increase it to 80% ('quasi JND'). Items with scores below 80% correctly scoring healthy controls were considered to reduce the specificity of the test. In addition to a minimum score for each test item, we decided to avoid items with a ceiling score of 100% in the healthy control group as this might imply a lack of sensitivity to diminished discrimination. We consider items with scores within the range of at least 80% and below 100% as being sensitive to diminished auditory speech discrimination.

Once the sensitivity of the test items was established, each subtest should allow adequate performance by healthy controls. We defined adequate performance as a test score of at least 80% by healthy participants. An effect size of 30% was employed (being the difference between 80% score and a chance level score of 50% on a test); with test power set at the conventional 80% and an adopted one-tailed significance level of 5%, a sample size of 15 items was needed according to the program SamplePower (IBM SPSS). Thus, a compromise was achieved between a considerable effect size and test power on the one hand, and an acceptable number of test items for neurological patients.

Furthermore, under the assumption that the conditions for both separate items and for test score per healthy control participant were met, we decided that the majority of healthy controls (at least 80%) should score adequately on a test. The number of healthy controls required to establish this condition was also determined by a sample size calculation. This was based on an effect size of 30%, being the difference between the majority of controls (i.e. 80%) and chance level (50%), adopting a significance level of 5%. Based on these calculations, we required at least15 healthy control participants.

Speech materials

The speech materials used as auditory stimuli in the five subtests of the ADT contain both words and sentences. Since the ADT aimed to assess auditory speech discrimination skills, semantic distraction had to be avoided. Therefore, ADT items at word level needed to contain words that occur in the CELEX corpus of written Dutch (Baayen, Piepenbrok, & Gulikers, 1995) with a minimum frequency of 50. For test items at sentence level, we selected sentences that are often applied in auditory testing in the Netherlands (Plomp & Mimpen, 1979). The Plomp and Mimpen sentences offer the benefits of i) representing conversational speech, ii) being of acceptable length, varying from seven to nine syllables and iii) containing non-distracting semantic content. Voiceless stops and voiceless fricatives were avoided as much as possible, since these phonetic features are less suitable for manipulations of speech dimensions in various subtests. The speech materials for each subtest are presented in appendix A.

The speech materials for the ADT were produced by both a 66 year old male and a 62 year old female speaker, whose ages corresponded to that of the majority of dysarthric patients with acquired neurological impairments. Both were native speakers of Dutch and had an academic degree. The speech was digitally recorded in a sound studio at a sampling frequency of 44.1 KHz. For the speech recording, a Sennheiser MKH418 P48 microphone with vocal pop filter and the computer program Adobe Audition, version 1.5, were used.

Subtests

With each subtest containing 15 experimental items for methodological statistical reasons (Section 2.1.), we decided to divide this number in seven equal pairs ('equal items') and eight unequal pairs ('unequal items'). Although there were an unbalanced number of equal and unequal pairs we decided not to increase the number of test items to avoid unnecessarily burdensome tasks in neurological patients. That is, patients will have to perform five subtests in total, with each subtest being preceded by two example items. To obtain equal and unequal items, speech materials were manipulated using the speech processing program 'Praat' (Boersma & Weenink, 2008). In all subtests, in an equal pair the default stimulus (i.e. the original speech recording, adjusted to a mean intensity of 65 dB in 'Praat') was presented twice. In an unequal pair, the default version was presented together with a manipulated version. Thus, these unequal pairs contained different values of the speech dimension at issue (e.g. different intensities in the subtest 'intensity', different overall pitch in the subtest 'overall pitch'). The magnitude difference in unequal pairs was based on reported Just Noticeable Differences (JNDs) (Rietveld & van Heuven, 2009a).

For the subtests 'overall pitch' and 'intonation', we developed both a male and a female version for reasons of ecological validity. Distinct versions of these subtests for both genders were based on the observed differences in mean pitch and in magnitude of pitch change between male and female voices (Kraayeveld, 1997). As a consequence, listeners tend to adapt their judgement about pitch change dependent on the speaker's gender (Gussenhoven & Rietveld, 1998). These differences should be reflected in a male and a female version for the subtests 'overall pitch' and 'intonation'. For the other subtests, separate male and female versions were not developed since evidence to do so was considered poor. For instance, for 'speech rate', studies investigating the effect of gender present conflicting results. Van Borsel and De Maesschalck (2008) did not find gender differences in overall speech rate for Dutch, although their conclusion concurs with other findings on gender differences in Dutch (Quené, 2005; Binnenpoorte, Van Bael, den Os, & Boves, 2005; Quene, 2008). For the subtests 'segmental elements' and 'loudness', little evidence was found in favour of separate male and female versions either. For 'segmental elements', an additional reason for using one version only (i.e the male version), is the fact that phoneme manipulation through spectral interpolation does not always yield equivalent results with different (i.e. male and female) speakers. The subtests of the ADT are discussed in detail below.

Segmental elements

The first subtest of the ADT ,'segmental elements', seeks to assess a listener's ability to discriminate between intracategorical phonemes (Liberman, Harris, Hoffman, & Griffith, 1957). This discrimination within phoneme boundaries represents the perception of distinct (i.e. correct or incorrect) realisations of phonemes and is vital in speech training. The stimuli in the subtest 'segmental elements' contained frequently occurring monosyllabic words. Unequal pairs contained two distinct stimuli: the default stimulus (i.e. the original recording adjusted to 65 dB) and the manipulated stimulus in which one phoneme from the default stimulus had been manipulated. The manipulated phoneme was either a consonant or a vowel. Through spectral interpolation of the target phoneme ('master') with another phoneme ('slave') by means of the program 'Provo' (Van Hessen, 1987), eight interpolated sounds were created. Three trained listeners, who were all native speakers of Dutch and had an academic degree in phonetics or speech pathology, selected one interpolated sound as the most suitable representation of a distorted version of the target phoneme. For an overview of the default phonemes and the phonemes used to obtain an interpolated version in the unequal pairs, we refer to Table 4.1.

Loudness

The second subtest, 'loudness', is designed to assess a listener's ability to discriminate between distinct mean intensities of speech. Seven pairs of words and eight pairs of sentences were used as speech stimuli for auditory presentation. In the default stimuli the mean intensity was adjusted to 65 dB using 'Praat' (Boersma & Weenink, 2008), whereas the mean intensity in the manipulated stimuli was adjusted to 71 dB. Thus, the difference between mean intensities in unequal pairs was 6 dB. The magnitude of this difference was based on the reported Just Noticeable Differences (JNDs) that vary from 1 to 5 dB (Sorin, 1981) as well as on clinical experiences in therapeutically relevant steps in intensity training. In only one unequal pair, containing the word [ʋi:l] (´wheel´), the difference was increased to 9 dB (65 and 74 dB respectively). This was based on the results of a preceding test trial with healthy participants, which showed poor discrimination performance between a mean intensity of 65 and 71 dB for this utterance.

Table 4.1:

Overview of the utterances in unequal pairs in the subtest 'segmental elements'. The unequal pairs contain the utterance with the default phoneme and the utterance with the interpolated version of the phoneme. The 'master' (default phoneme) was interpolated with the 'slave'(distinct phoneme). Orth. = orthographic transcription; phon. = phonetic transcription. English translations of the utterances between brackets.

CONSONANTS					VOWELS				
Utterance		Phoneme			Utterance		Phoneme		
orth.	phon.	default	interpolation		orth.	phon.	default	interpolation	
			master	*slave*				*master*	*slave*
/soep/ *(soup)*	[sup]	[s]	[s]	[ʃ]	/kaak/ *(jaw)*	[kaːk]	[aː]	[aː]	[i]
/soms/ *(sometimes)*	[soms]	[s]	[s]	[ʃ]	/piep/ *(peep)*	[pɪp]	[i]	[i]	[u]
/das/ *(tie)*	[dɑs]	[d]	[d]	[h]	/map/ *(map)*	[mɑp]	[ɑ]	[ɑ]	[i]
					/miep/ *(name)*	[mɪp]	[i]	[i]	[ɑ]
					/kaak/ *(jaw)*	[kaːk]	[aː]	[aː]	[ə]

Overall pitch

The goal of the subtest 'overall pitch' is to assess a listener's ability to discriminate between distinct pitch versions of speech. As mentioned before, both a male and a female version of this subtest were developed. To rule out the influence of differences in speech rate between the male and the female speech recordings, the speech rate for the male and the female speaker was equalized for all utterances. The items in this subtest contained 7 pairs of words and 8 pairs of sentences.

To obtain unequal pairs, the overall pitch was manipulated using the PSOLA technique in 'Praat'. This technique does not affect the naturalness of the manipulated utterances (Moulines & Charpentier, 1990). Unequal pairs contained the speaker's original (i.e. default) overall pitch combined with either a higher pitch version or a lowered pitch version. In the male version of the subtest, the higher pitch versions were 4 semitones higher than the original pitch. In the female version, the pitch was raised by 3 semitones (in one case 5 semitones) to obtain higher pitch versions. For unequal pairs containing lowered pitch versions, the lower version differed 6 semitones from the default pitch for females, and 7 semitones for males. An additional pitch manipulation had to be made in these lowered pitch pairs. That is, unlike all other default stimuli, the default pitch in these unequal items was raised by 2 semitones for both the male and the female version. This was done to avoid unnatural sounding stimuli due to a lowering of 6 or 7 semitones by pitch manipulation. Increasing the default pitch by 2 semitones allowed an equal increase of the lowered pitch, thus maintaining a natural sounding lowered pitch. Thus, for the unequal items containing lowered pitch versions, the slightly increased pitch version of the original utterance served as the default pitch.

The frequent use of the semitone scale in psychoacoustic research and its availability in software for speech manipulation allowed its use in our study. The fact that the stepsize on the semitone scale for the male speaker was somewhat larger than for the female speaker is associated with the lower fundamental frequencies in male voices than in female voices. It is in accordance with the findings of (Hermes & Gestel van, 1991) who compared the semitone scale with the scale of Equivalent Rectangular Bandwith (ERB). They concluded that equally perceived pitch changes require larger step sizes in lower pitch regions than in the higher pitch regions. This also explains the larger step sizes for pitch lowering than for pitch rising.

Speech rate

The fourth subtest of the ADT, 'speech rate', is designed to assess a listener's ability to discriminate between distinct speech rates. All auditory stimuli for this subtest consisted of sentences as sentences are generally used in speech rate training. The unequal pairs contained a sentence at natural (default) rate and a manipulated sentence with either a slower or a faster speech rate. Since rate changes in natural speech tend to co-occur with variations in other speech dimensions, such as pitch and intensity (Smith, 2002), the slower and faster rate versions in the unequal pairs were based on natural rate changes in speech. To this end, the male speaker was asked to produce the sentences at different speech rates. To visualize different speech rates for the speaker, a 7-point scale representing speech rates varying from '-3' (very slow) to '+ 3' (very fast) was used. A value of '0' was assigned to his natural speech rate. The speaker was asked to produce a speech rate comparable to a value of '0' (natural), '-1' (slightly slower), '0' (natural) and '+1' (slightly faster) successively. In the unequal pairs, the slower versions contained the

'-1 version' (slightly slower), additionally manipulated to an overall-deceleration of 5% (Quene, 2007). For unequal pairs containing the faster versions, the '+1 version' (slightly faster) was additionally manipulated to an overall acceleration of 5%. For an overview of the original and the modified speech rates, we refer to Appendix 4B.

Intonation

The last ADT subtest 'intonation' investigates a listener's ability to discriminate between relative prominences of sentence accents. The differences were generated by varying the magnitudes of the pitch changes associated with the sentence accents (Rietveld & Gussenhoven, 1985). For the manipulation of sentence accents, only pitch rising was used as in most Germanic languages, like English, German and Dutch, a high pitch target is a common way of signalling a sentence accent. Lowering pitch values has hardly any effect on the perceived prominence. The range in which the lowering could be realized for the low pitch accents (symbolized as L*H in terms of the autosegmental theory of intonation) is also very limited (Gussenhoven & Rietveld, 2000).

A male version as well as a female version was created for this subtest. Unequal pairs in this subtest contained the default utterance as well as a version with a manipulation of the last sentence accent (Figure 4.1). In both the male and the female version of the stimuli, the speaker's original intonation was used as the default utterance. In the unequal pairs, the default version was presented along with a manipulated version, in which the last sentence accent was increased compared to the default utterances. For the male speaker the increase was by 6 or 7 semitones compared to the default utterances. For the female version, the last sentence accent was increased 4 or 5 semitones compared to the default utterances.

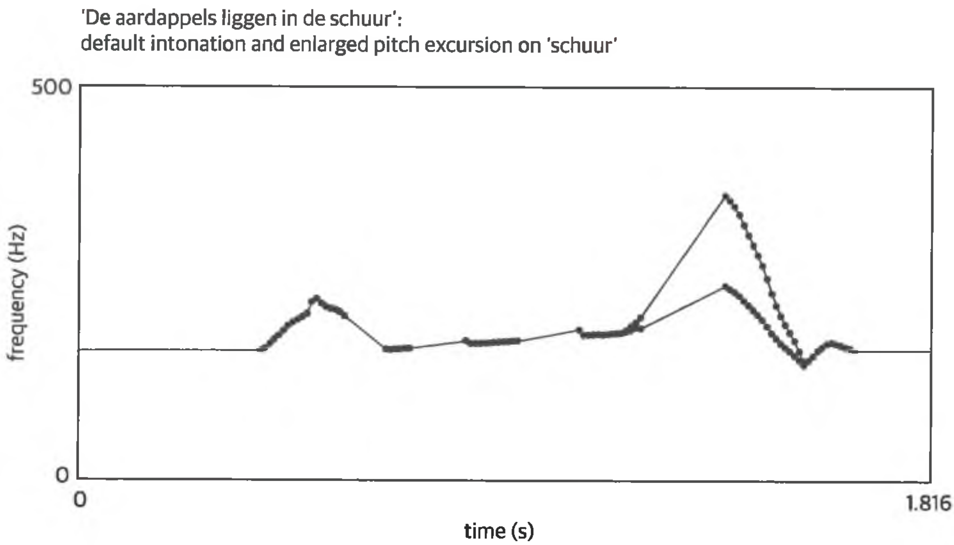


Figure 4.1. :

An unequal pair in the female version of the subtest 'intonation', containing two distinct versions of the utterance ' De aardappels liggen in de schuur' [də 'a:rdapəls lIɣə In də 'sʏʏr] ('The potatoes are in the barn') in the female version of the subtest 'intonation'. The default version contains the original intonation of the female speaker. The manipulated version contains an enlarged pitch excursion (7 semitones) on the last sentence accent 'schuur'.

Stimuli presentation

The ADT was presented using the E-prime software tool, version 1.2 (Schneider, Eschman, & Zuccolotto, 2002), allowing a standardized test procedure. The test items were delivered binaurally by means of headphones (Sennheiser PC 131). Participants were not allowed to adjust the volume to their own preferred level since this might cause loudness recruitment. This phenomenon is frequently observed in elderly people with sensorineural hearing loss. It is characterized by a faster growth of perceived loudness level with increasing intensity level (Moore, 2003) than in a normal hearing person, and is known to affect speech perception. During the aural presentation of equal or unequal pairs of the utterance (i.e. the identical or the non-identical realization), the corresponding orthographic representation of the target utterance was displayed on the laptop screen. Within each test item (i.e.

between the two stimuli of a pair), the interstimulus interval (ISI) was 500 ms.

Unlike all other subtests, in the subtest ‘segmental elements’ each test item was presented twice. This was done to provide the listener with the opportunity to focus on discrimination within phoneme boundaries. This is a complex task, since, for communicative purposes, intercategorical rather than intracategorical phoneme perception is required (Liberman, et al., 1957). Hence, auditory discrimination within phoneme boundaries, as is the case in the subtest ´segmental elements´, is complicated. This finding is supported by the poor performance of healthy participants in a preceding test trial, in which test items of ‘segmental elements’ were only presented once. The opportunity to listen to each test item twice, allows participants to focus on discrimination within phoneme boundaries.

All ADT subtests were presented in a fixed order: ‘segmental elements’, ‘loudness’, ‘overall-pitch’, ‘speech rate’ and ‘intonation’. ‘Segmental elements’ was presented first, since, due to the complexity of the intracategorical phoneme discrimination task, this subtest requires a participant’s utmost attention. ‘Segmental elements’ was followed by ‘loudness’ as it was considered a less complex task. The subtests ‘overall pitch’ and ‘intonation’ were separated by the subtest ‘speech rate’ to avoid confusion of these similar speech dimensions. Each subtest was followed by a pause of 60 seconds to minimize effects of fatigue on test performance. Within each subtest, two distinct versions of item order existed to rule out an effect of item order.

Participants

Thirty-six healthy control participants who were all native speakers of Dutch completed the ADT. Their age ranged from 55 to 74 years. This age range covers the age of the majority of patients with stroke and Parkinson’s disease in the Netherlands, being the neurological target group for the ADT. Participants were divided into four age categories: 55-59 years (n=8); 60-64 years (n=10); 65-69 years (n=9) and 70-74 years (n=9). The participants were divided as equally as possible for gender, age group and education level. Nineteen males and 17 females were included. Three education levels were represented. Education levels were based on the self-reported highest level of completed education. Level 1 represents primary school and lower vocational education (n=11), level 3 represents education resulting in a bachelor or academic degree (n=13) and level 2 was assigned to education levels higher than level 1 and lower than level 3 (n=12). Eighteen participants carried out the ADT with item order 1 and 18 participants performed item order 2.

Procedure

Before carrying out the ADT, the participants received written task instructions and were allowed to ask questions about the test procedure. Female listeners were presented with the female version whereas male listeners were presented with the male version of the ADT. Instructional text appeared on the laptop screen prior to each subtest to ensure that participants focused on the speech dimension at issue. The two training items in each subtest were followed by a written message on the screen containing the correct response. No feedback was given for the 15 experimental items in each subtest. Using the E-prime software tool, for each participant the responses ('same - different') and the reaction times (ms) on all items were registered in a digital data file.

Once the ADT was completed a pure tone audiogram was conducted. Based on these audiological results, the mean value of the Fletcher Index in both ears (henceforth: mean FI) was established, using an 'Interacoustics AS 208' screening audiometer. The mean FI for both ears varied between 3 and 27 dB with a mean of 12.81 dB (SD = 5.24), indicating a normal hearing level for the frequency area relevant for speech perception.

4.2.2 Data analyses

By using E-prime, an automatically generated data file was provided for each participant. Score percentages and reaction times for all subtests were calculated

4.2.3 Results

Out of 75 items across the five subtests, 21 items did not meet the sensitivity criterion. That is, four (unequal) items did not achieve the 80% criterion and 17 items (nine equal and eight unequal pairs) showed a 100% score. In the subtest 'segmental elements' one unequal item did not reach the 80% criterion, whereas a 100% score was achieved on one equal and one unequal item. In the subtest 'loudness' all items reached a minimum score of 80%, yet seven items did not fall within the score range of the sensitivity criterion: three unequal and four equal items showed a 100% score. In 'overall pitch' two unequal items did not reach an 80% score, whereas a maximum score of 100% was reached by one unequal and two equal items . In 'speech rate' one equal and three unequal items showed a ceiling score. In these unequal pairs utterances at default rate were compared to decelatered versions. 'Intonation' showed a score below 80% on one unequal item (default sentence accent compared to increased sentence accent with 5 semitones and 7 semitones in the female and the male version respectively). A 100% score was achieved on an equal item. In Appendix A, for all subtests the items with scores below 80% and the items with maximum scores of 100% are marked.

4.2.4 Discussion

The results of Experiment 1 showed that all subtests contained items that did not meet the predetermined sensitivity criterion of at least 80% and below 100% scoring healthy controls. For this reason, we could only evaluate this predetermined sensitivity criterion for each separate item and thus identify suitable items. Since all subtests contained unsuitable items we were not able to evaluate the other 80% criteria: the required 80% score of a healthy participant in a subtest with 15 suitable items for sufficient performance, and the minimum of 80% healthy participants that sufficiently performed a subtest with 15 suitable items (Section 4.2.1). Nevertheless, in order to explore the sensitivity of this newly developed assessment tool for auditory speech discrimination, the performance of healthy and matched neurological participants was investigated for the suitable (i.e. sensitive) items in Experiment 2.

4.3 Experiment 2

With Experiment 2 we sought to explore the ADT's sensitivity to diminished auditory discrimination performance. To this end, we compared the performance of neurological and healthy participants for only items that were identified as suitable (i.e. sensitive) in Experiment 1. That is, 12 items in 'segmental elements' (six equal and six unequal pairs), eight items for 'loudness' (three equal and five unequal pairs), 10 items for 'overall pitch' (five equal and five unequal pairs), 11 items for 'speech rate' (six equal and five unequal pairs) and 13 items for 'intonation' (six equal and seven unequal pairs).

4.3.1 Materials and methods

Participants

Fourteen patients with dysarthric speech due to acquired neurological impairment were matched for age, gender and education level with 14 healthy controls. Age matching was carried out per age category, each category covering five years (e.g. 40-44 years, 45-49 years, 50-54 years and so forth). All participants were native speakers of Dutch. The neurological patients were dysarthric speakers due to stroke (n=10), Parkinson's disease (n=3) or encephalitis (n=1). Patients with aphasia were excluded to avoid interference of language comprehension deficits with diminished auditory discrimination skills. Assessment of cognitive functioning and arm-hand coordination, which might affect the ADT performance, were conducted for all neurological patients. Table 4.2 shows an overview of the matched healthy and neurological participants in Experiment 2.

The mean age in the healthy group was 61.8 yrs. (SD=9.4) and in the neurological group 62.4 yrs. (SD= 10.1). A paired t-test (two-tailed) for matched samples (df=13) yielded a t-value of 1.29 and did not achieve statistical significance (p= .22). Matching of education level was based on the participants' self reports on highest level of completed education. The median was 3, being the highest education level.

Procedure

Following the same procedure as in Experiment 1, participants of both the healthy and the neurological group completed the ADT and an audiogram was obtained. The mean FI in the healthy controls varied from 1 to 16 dB (M = 9.9, SD = 4.9) and from 1 to 49 dB (M = 20.6, SD = 14.5) in the neurological participants. A paired t-test (one-tailed) for matched samples (df = 13) yielded a t-value of 2.80, being statistically significant (p = 0.008).

Although unsuitable test items were identified in Experiment 1, neurological participants were asked to perform the ADT including the unsuitable items. By providing an equal number of items across all subtests we avoided interfering effects of distinct task load between subtests. For statistical analyses however, only suitable items were included.

4.3.2 Data analyses

For all subtests mean score percentages, mean number of 'hits' (i.e. adequately identified unequal items) and mean reaction times (ms) were calculated for all matched healthy and neurological participants. With matched pairs of participants being involved, paired t-tests for matched samples (henceforth: t-tests for matched samples) were carried out for these three outcome measures for all subtests. In addition, across all participants (n=28) correlations were investigated for score percentages and hearing loss (mean FI), for number of hits and hearing loss, and for reaction times and hearing loss. Univariate ANOVA's for repeated measures, allowing the introduction of 'mean FI' as a covariate, were conducted for all outcome measures on all subtests (Kirk, 1995; Rietveld & van Hout, 2005). In case of statistical significance, Hochberg's procedure for multiple

Table 4 2:

Demographic data for the 14 matched pairs of healthy and neurological participants. Edu = education level; Neur. Imp. = neuro Time p.o. = Time post onset . Reported concomitant constraints in neurological participants under 'cognition' and 'arm-hand co

	Healthy controls			Neurological participants					
Pair	Gender	Age	Edu	Gender	Age	Edu	Neur. Imp.	Time p.o.	Cognition
1	male	65	2	male	68	2	ischaemic stroke LH	3 mth.	slightly diminished attention
2	female	66	1	female	69	1	encephalitis	5 mth.	no abnormalities observed
3	male	65	1	male	65	1	ischaemic stroke LH	3 mth	no abnormalities observed
4	male	56	2	male	57	2	ischaemic stroke RH with hemorrhagic component	3 mth.	visual hemineglect left, diminished attention
5	male	60	3	male	60	3	hemorraghic stroke RH	3 mth.	diminished attention
6	male	72	2	male	70	2	Parkinson's disease	12 yrs.	delayed information processing
7	female	43	3	female	44	3	ischaemic stroke LH	2 mth.	no abnormalities observed
8	male	75	3	male	78	3	Parkinson's disease	1 yr	delayed information processing
9	male	69	3	male	69	3	Parkinson's disease	9 yr	no abnormalities observed
10	female	47	3	female	47	3	ischaemic stroke LH	2 mth	no abnormalities observed
11	male	63	3	male	63	3	ischaemic stroke RH	5 mth.	diminished attention, visual hemineglect, delayed information processing
12	male	71	3	male	74	3	ischaemic stroke RH	1 mth	diminished attention, visual hemineglect
13	male	60	3	male	60	3	hemorraghic stroke RH	2 mth.	mild visual hemineglect
14	female	53	3	female	50	3	hemorraghic cerebellar stroke	3 mth.	diminished attention

outcomes of significance was applied in order to avoid an increased probability of a Type I error (Benjamin & Hochberg, 1995). Finally, since the neurological group contained participants with diverse disease states (Table 2), we investigated whether the matched pairs interacted with the magnitude of differences within pairs. Therefore, Tukey's test of Nonadditivity (Rietveld & van Hout, 2005) was conducted for score percentages, number of hits and reaction times.

4.3.3 Results

Score percentages

For all subtests, a tendency of lower score percentages in neurological participants than in their matched healthy controls was observed, although this pattern is less prominent for 'loudness' (Figure 4.2a-4.2e).

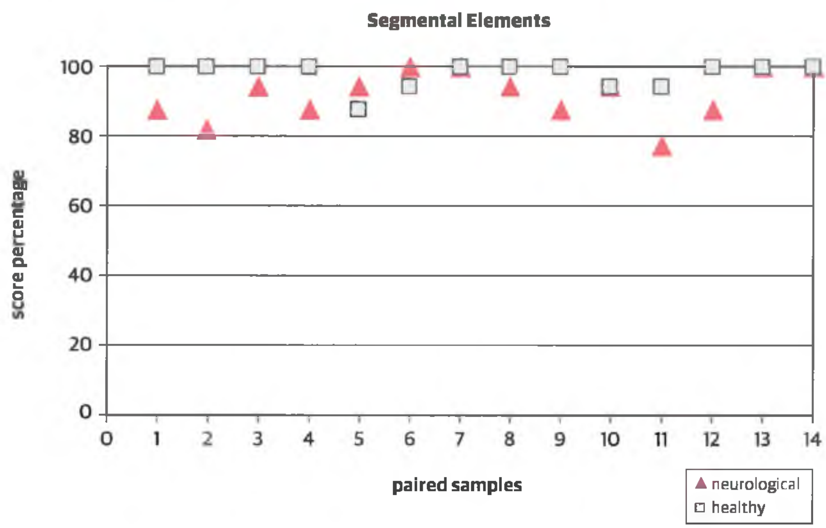


Figure 4.2.a.:

Score percentages on the subtest 'segmental elements' (containing 12 suitable items) of the matched pairs of healthy and neurological participants.

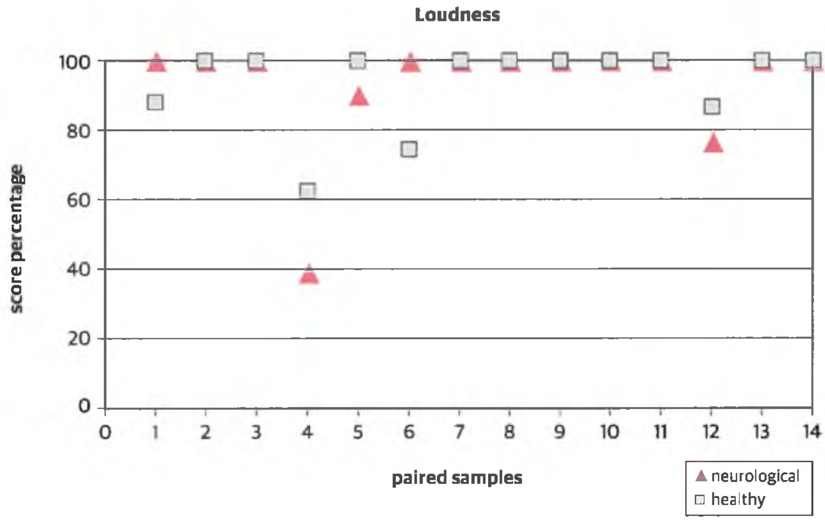


Figure 4.2.b.:

Score percentages on the subtest 'loudness' (containing 8 suitable items) of the matched pairs of healthy and neurological participants.

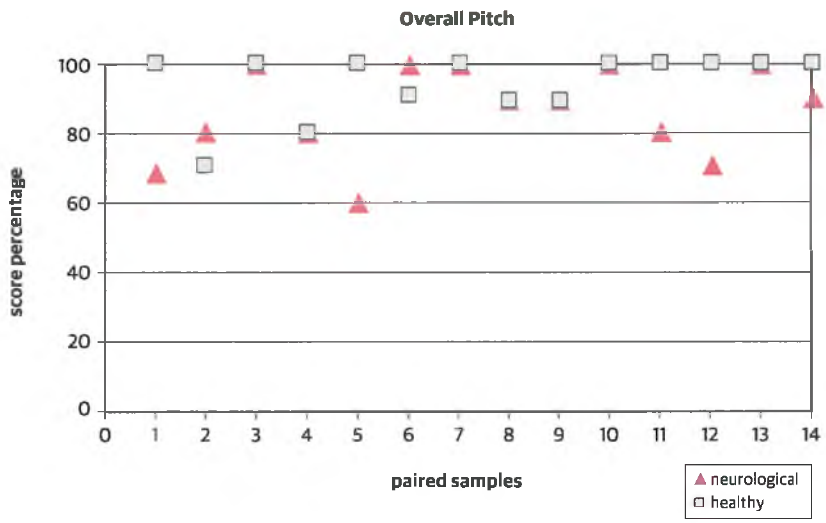


Figure 4.2.c.:

Score percentages on the subtest 'overall pitch' (containing 10 suitable items) of the matched pairs of healthy and neurological participants.

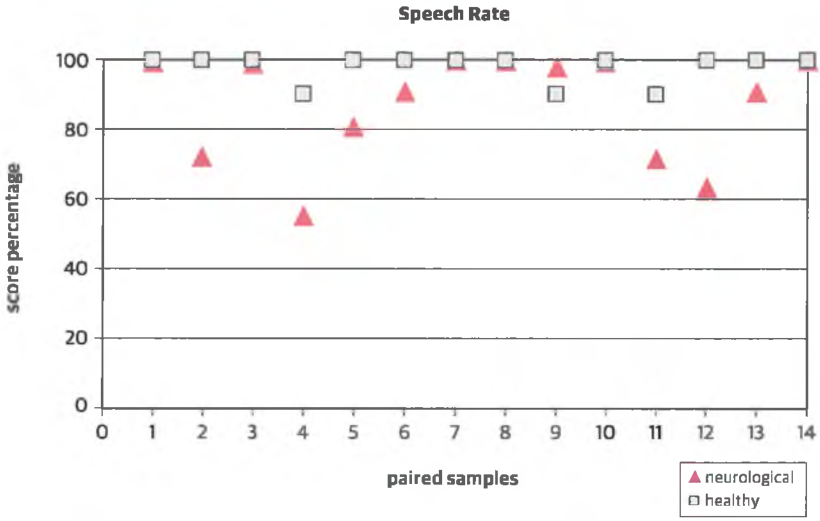


Figure 4.2.d.:

Score percentages on the subtest 'speech rate' (containing 11 suitable items) of the matched pairs of healthy and neurological participants

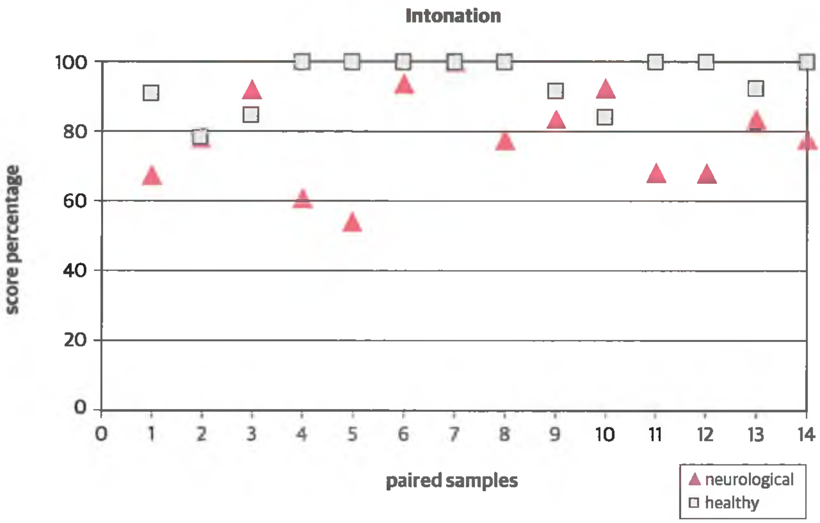


Figure 4.2.e:

Score percentages on the subtest 'intonation' (containing 13 suitable items) of the matched pairs of healthy and neurological participants.

Mean scores for the healthy and the neurological group on suitable items are presented in Figure 4.3.

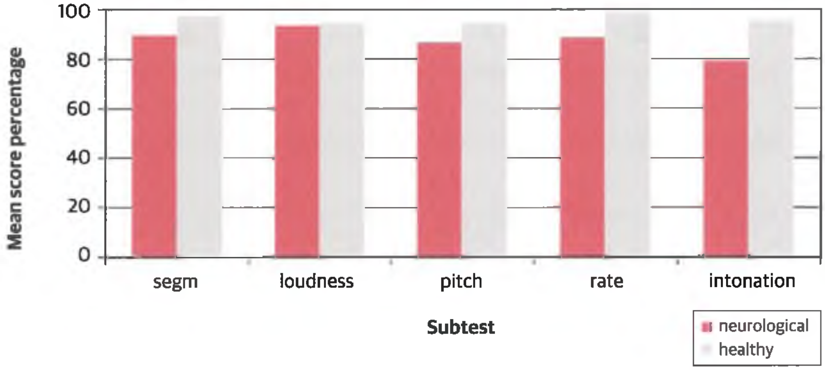


Figure 4.3.:

Mean percentages per subtest (only suitable, i.e,. sensitive items) for the neurological group (n=14) and the healthy group (n=14).

One-tailed t-tests for matched samples (df=13) showed that the lower performance of the neurological groups reached statistical significance for 'segmental elements' (t = 2.74 p = .009, effect size (ES) = 0.60), 'overall pitch' (t = 1.86, p = .043, ES = 0.49), 'speech rate' (t = 2.7, p = .010, ES = 0.59), and 'intonation' (t = 3.50, p = .002, ES = 0.70).

With the 'mean FI' in the neurological participants being significantly lower than in the healthy controls, the correlation between hearing loss and scores was established. For all subtests, Pearson correlations between score percentages and mean FI were calculated across all participants (n=28). Significant correlations were found between 'mean FI' and the score percentages in the subtests 'segmental elements' (r = -.510, p (one-tailed) = .003),' overall pitch' (r = -.373, p (one-tailed) = .025) and 'intonation' (r = -.476, p (one-tailed) = .005). To rule out the effect of hearing loss the Fletcher Index was introduced as a covariate in the ANOVA's for repeated measures, resulting in significant differences for 'speech rate' and 'intonation' only (Table 4.3). Applying Hochberg's procedure for multiple outcomes resulted in nonsignificant p-values for these subtests however.

Table 4.3:

Results of ANOVA's for repeated measures of score percentages in all subtests with 'mean Fletcher Index' as a covariate. Adjusted p-values resulting from Hochberg's procedure between brackets.

Subtest	Mean (SD)		$F_{1,12}$	p
	Neurological	Healthy		
Segmental Elements	88.79 (10.00)	97.07 (5.27)	2.06	.18 (.54)
Loudness	92.93 (17.35)	93.86 (11.62)	0.21	.66 (.66)
Overall Pitch	86.43 (13.36)	93.57 (9.29)	0.92	.36 (.66)
Speech Rate	87.79 (15.63)	98.07 (3.83)	0.63	.03* (.12)
Intonation	78.57 (13.14)	94.50 (7.58)	5.79	.03* (.12)

** = statistically significant (p< .05)*

Tukey's test for Nonadditivity yielded significant results for 'segmental elements' ($F_{1,13}$ = 5.641, p = .035), 'loudness'($F_{1,13}$ = 4.792, p= .049) and 'speech rate' ($F_{1,13}$ = 52.927, p = .000). For these subtests score differences within pairs were not equal over pairs.

Sensitivity

One-tailed t-tests for matched samples (df=13) showed that neurological participants scored significantly fewer hits than their matched healthy controls for 'segmental elements' (t = 3.02, p = .005, ES = .64) 'speech rate' (t = 2.46, p = .029, ES = .56) and 'intonation' (t = 3.30, p = .003, ES = .46).

Across all participants (n=28) significant correlations were found between 'mean FI' and the number of hits in the subtests 'segmental elements' (r = -.421, p (one-tailed) = .001),' and 'intonation' (r = -.384, p (one-tailed) = .002). The introduction of the Fletcher Index as a covariate in univariate ANOVA's for repeated measures, yielded significant differences for 'intonation' only. Hochberg's procedure for multiple outcomes yielded a nonsignificant p-value for 'intonation' (Table 4.4).

Table 4.4:

Subtest	Mean (SD)		$F_{1,12}$	p
	Neurological	Healthy		
Segmental Elements	5.00 (0.96)	5.79 (0.43)	1.68	.09 (.35)
Loudness	4.43 (1.40)	4.71 (0.83)	0.00	.96 (.96)
Overall Pitch	4.07 (1.21)	4.36 (1.08)	0.04	.84 (.96)
Speech Rate	4.50 (0.76)	5.00 (0.00)	4.23	.62 (.96)
Intonation	4.71 (1.49)	6.36 (1.01)	5.05	.04* (.20)

** = statistically significant (p< .05)*

Tukey's test for Nonadditivity yielded significant results for 'segmental elements' ($F_{1,13}$ = 10.23, p = .008) and 'loudness' ($F_{1,13}$ = 6.30, p = .003).

Reaction times

T-tests for matched samples (one-tailed) with df =13 showed significant longer reaction times (ms) for neurological patients in all subtests: 'segmental elements' (t = 2.48, p = .014, ES = .57), 'loudness' (t = 1.80, p = .048. ES = .45), 'overall pitch' (t = 1.96, p = .036, ES = .48), 'speech rate' (t = 2.17, p = .025, ES = .52) and 'intonation' (t = 3.11, p = .004, ES = .65).

Across all participants (n=28) significant correlations were found between 'mean FI' and reaction times in the subtests 'segmental elements' (r = .327, p (one-tailed) = .045),' and 'loudness' (r = .194, p (one-tailed) = .02). Introducing the Fletcher Index as a covariate in univariate ANOVA's for repeated measures, yielded significant differences for 'intonation' only (Table 4.5). Adjusted p-levels resulting from Hochberg's procedure for multiple outcomes yielded a nonsignificant p-value for 'intonation'.

Tukey's test for Nonadditivity showed significant interactions between pairs and differences in reaction times for 'loudness' ($F_{1,13}$ = 14.678, p = .002), 'overall pitch' ($F_{1,13}$ = 19.830, p = .001), 'speech rate' ($F_{1,13}$ = 51.279, p = .000) and 'intonation' ($F_{1,13}$ = 11.712, p = .005).

Table 4.5:

Results of univariate ANOVA's for repeated measures for reaction times (ms) in all subtests with 'mean Fletcher Index' as a covariate. Adjusted p-values resulting from Hochberg's procedure between brackets.

Subtest	Mean (SD)		$F_{1,12}$	p
	Neurological	Healthy		
Segmental Elements	4753 (2583)	2747 (1492)	3.96	.07 (.15)
Loudness	2081 (853)	1670 (333)	3.13	.10 (.15)
Overall Pitch	2352 (821)	1850 (295)	2.41	.15 (.15)
Speech Rate	2691 (751)	2216 (178)	3.81	.08 (.15)
Intonation	2974 (666)	2420 (283)	5.27	.04* (.15)

4.4 General discussion

This paper reports the development and the evaluation of an auditory discrimination test (ADT) for the purpose of assessing auditory speech discrimination in Dutch patients with dysarthric speech due to acquired neurological impairment. The main reason for constructing an auditory discrimination test is a recently developed web application for speech training in neurological patients: E-learning based Speech Therapy (EST) (Beijer et al., 2010a; Beijer et al., 2010b). The EST training procedure requires auditory discrimination between distinct realizations of speech. It is therefore vital to assess neurological patients' ability to perform an auditory discrimination task. The five subtests which make up the ADT address the speech dimensions similar to those to be trained by means of EST: segmental speech elements, intensity, overall pitch, speech rate and intonation.

4.4.1 Evaluation of test items

From the results of Experiment 1 it became obvious that the development of test items with scores within the predetermined range of 'minimal 80% - below 100%' correctly responding healthy controls, is a complicated and time consuming task. Across all subtests, four items did not achieve a minimum score of 80% correctly responding healthy controls, although the differences in these unequal pairs were well above reported

JNDs. The relatively poor discrimination on these unequal items might be partially due to the interference of test stimuli's phonetic characteristics with 'normal' thresholds for auditory perception of variations along speech dimensions. That is, the magnitude of manipulated differences in the ADT's unequal pairs were based on reported JNDs which were established using sinus waves or specific speech materials, such as isolated vowels (Rietveld & van Heuven, 2009), whereas the manipulated speech materials in the ADT contained different and linguistically more complex materials. For instance, in the subtest 'overall pitch' the unequal items with a score below 80% contain a considerable number of nasals, which are known to impede detection of pitch change (Straatman, Rietveld, Beijen, Mylanus, & Mens, 2010). Possibly, this interference of phonetic features with auditory perception thresholds accounts for the relatively poor detection of differences in overall pitch in unequal pairs. In fact, using JNDs as a reference for establishing the magnitude of noticeable differences in ADT items containing unequal pairs might be ecologically invalid. Another explanation for this presumed lack of ecological validity could be found in the fact that for the establishment of JNDs, procedures were used that differed from the auditory discrimination task in the ADT. That is, for JND experiments, multiple presentation of stimuli tends to be employed, whereas the ADT discrimination task only allowed one single item presentation. We argue that the observation that reported JNDs for many speech dimensions will not be reached in tests such as the ADT constitutes an important lesson: the characteristics of the speech materials (phonetic features, language, speaker characteristics) interact with the target speech dimensions and consequently determine (quasi-) JNDs to some extent. This also explains the observation that the difference of 6dB in the unequal pair [wi:l] (´wheel´) in the subtest 'loudness' was not perceived by healthy controls and had to be increased to 9dB.

Apart from items with minimum scores, maximum scores were observed in 17 items across all subtests. These 100% scores occurred in seven items of the subtest 'loudness' (i.e. four equal and three unequal) pairs. This relatively large number of items with a ceiling score in this subtest could be explained by the fact that listeners are familiar with the dimension 'loudness', being a frequently occurring sound dimension in daily life. 'Speech rate' contained four items with 100% scores. One of these items concerns an equal pair, whereas three items consist of unequal pairs in which utterances at default rate were presented along with decelerated versions. The difference between the default and the decelerated version is apparently perceived very well. Possibly, decelerated speech tends to be perceptually somewhat artificial since this is not a natural manner of speaking, unlike frequently occurring accelerated speech in daily life. The remaining six items with maximum scores are divided over the other three subtests: they concern four equal pairs (one in 'segmental elements', two in 'overall pitch' and one in 'intonation') and two unequal pairs (one in 'segmental elements' and one in 'overall pitch'). In general, the occurrence of maximum scores in equal pairs is obvious and difficult to avoid. The differences in the unequal pairs with ceiling scores in 'segmental elements' and 'overall pitch' are obviously too large to go unnoticed by listeners. A larger number of healthy control participants would be required to draw firmer conclusions with respect to the potential lack of sensitivity of these items.

4.4.2 Sensitivity of the test

Despite the complications in constructing test items yielding scores within the range of minimally 80% and below 100% correctly responding healthy controls, the results of Experiment 2 are encouraging with respect to the ADT's sensitivity to diminished auditory discrimination. With only suitable items being analyzed, score percentages, number of hits and reactions times were consistently lower in the neurological than in the healthy group across all subtests. Nevertheless, these results should be interpreted with care. That is, t-tests for paired samples for score percentages showed that the lower scores in neurological patients achieved significance level for all subtests except 'loudness'. Even ruling out the effect of hearing loss by conducting univariate ANOVA's for repeated measures with 'mean FI' as a covariate still resulted in significant p-values for 'speech rate' and 'intonation'. However, adjustment of p-values to avoid increase of probability of Type I error due to multiple outcomes, resulted in nonsignificant differences in these two subtests. These results imply that hearing loss accounts to a considerable extent for the lower scores in neurological patients. Hence, not auditory discrimination per se, but physiologically induced hearing loss might cause poorer performance in neurological participants. Hearing loss might even be an indicative variable for neurological impairment. This is in line with recent research results indicating that neurological diseases such as stroke might be associated with moderate to severe hearing loss (Gopinath, Schneider, Rochtchina, Leeder, & Mitchell, 2009). This could explain the observed larger hearing loss in the neurological participants compared to matched healthy controls. It should also be noted that, despite the lack of statistical significance, the tendency of neurological participants' lower scores in 'speech rate' corresponds to reported deficits in duration discrimination in stroke patients (Bamiou, Musiek, Stow, Cipolotti, Brown, & Luxon, 2006) and patients with Parkinson's disease (Guehl et al., 2008). In fact, these authors argue that the poorer perception of temporal variation in neurological patients might not be caused by diminished auditory discrimination per se, but could represent an underlying impairment in general cognitive functions. That is, rather than genuine hearing loss alone, the reported cognitive impairments in our neurological patients might also play a role. Diminished attention and delayed information processing could negatively affect both the performance on audiometric tests and auditory information processing required in the ADT. This is in agreement with the assumption of cognitive impairments contributing to auditory discrimination problems in elderly neurological patients (Section 4.1) (Rooij van et al., 1991; Golding et al., 2004; Golding et al., 2006; Pichora-Fuller & Singh, 2006a). In our study, cognitive deficits in the neurological participants might be reflected by their lower number of adequately identified unequal pairs (hits), being an index for sensitivity. T-tests for matched samples for 'segmental elements', 'speech rate' and 'intonation' yielded significantly lower scores for neurological patients, although univariate ANOVA's for repeated measures with hearing loss as a covariate resulted in statistical significance for 'intonation' only. Adjustment of p-values resulted in nonsignificant differences for this subtest, however. The same pattern was observed for reaction times. T-tests for matched samples showed significantly longer reaction times for neurological participants on all subtests, whereas the introduction of 'mean FI' as a covariate in the univariate ANOVA's for repeated measures resulted in statistical significance for 'intonation' only. Adjustment of p-values to control for the effect of multiple outcomes resulted in a nonsignificant p-value for 'intonation'. The observed longer reaction times in the neurological patients across all subtests support the hypothesis of underlying cognitive deficits that play a role. The effect of reported diminished arm-hand functioning in neurological participants might also account for the longer reaction times.

4.4.3 Clinical relevance

Although the neurological patients in our study performed below the healthy controls, the majority responded correctly to at least 80% of the suitable items in most subtests. Neurological participants' overall performance indicated sufficient auditory discrimination for independent aural comparison of their own speech with the target speech in the EST training procedure. This would imply that diminished auditory discrimination skills in dysarthric speakers are not likely to affect the suitability of independent web-based speech training such as EST. However, this is not in line with other reported research outcomes which indicate that patients with Parkinson's disease have deficits in discrimination of duration (Guehl et al., 2008). These findings raise doubts about speech perception in these neurological patients which might reflect impairment of memory or attention. In addition to the small and heterogeneous neurological group in our study, these findings emphasize the need for further research into auditory discrimination and concomitant deficits that might affect the suitability of EST. From a therapeutic point of view it is vital to establish whether additional visual feedback on patients' speech in the EST training procedure is required. Visual feedback might support neurological patients who are unable to discriminate aurally between their own speech and the target speech. Moreover, the visualisation of speech dimensions might provide patients with cues for adjustments in order to approach the target speech. Finally, these visual cues might even compensate for cognitive impairments, which are likely to impede neurological patients' performance on the web-based speech training procedure.

4.4.4 Limitations

The results of Experiment 1 are limited in their generalization to other languages since the ADT has been designed for Dutch speakers only. Indeed, large cross-language differences exist in the sensitivity for variations along speech dimensions, such as differences in the processing of pitch change between Dutch and British English due to the distinct standard pitch ranges in both languages (De Pijper, 1983; Chen, Gussenhoven, & Rietveld, 2001). Therefore, constructing an auditory speech discrimination test requires language specific adjustments.

Experiment 2 did not allow firm conclusions with respect to ADT performance of specific neurological patient groups. The fact that several test items were not considered sensitive and were excluded, resulted in less data for comparison and, hence, less test power in Experiment 2. Moreover, the neurological group in Experiment 2 was not homogeneous. This could be justified from the perspective of ecological validity. That is, in clinical practice potential EST users in whom auditory speech discrimination should be assessed also concern dysarthric speakers with different neurological causes. However, this variety of neurological diseases induces different concomitant constraints such as hearing loss, motor coordination and cognitive deficits and are likely to affect ADT performance. Tukey's test for Nonadditivity established interaction between matched pairs and differences within pairs for several outcome measures and subtests, thus confirming the variability in our neurological group. However, we did not control for the effect of different neurological diseases and concomitant constraints. Experiment 2 should therefore be characterized as exploratory. Nevertheless, the results of our study seem to indicate the sensitivity of an auditory speech discrimination test and are therefore encouraging for future research.

Another limitation of our study was the fixed order presentation of the five subtests. The fact that the last presented subtest 'intonation' showed lower and more variable scores in neurological participants than in the other subtests might be the result of an order effect. Possibly, the reported cognitive constraints such as diminished attention in our neurological participants are responsible for this observation.

A methodological limitation of our study concerns the fact that in Experiment 2 neurological participants beyond the age range of 55 to 74 years were included. This was due to the limited availability of patients who were fit to participate. Three neurological participants were below 55 years whereas one was older than 74 years. Consequently, these patients were matched with healthy controls of the same age category, resulting in four matched pairs (i.e. pair 7, 8, 10 and 14 in Table 4.2) beyond the age range of the control group in Experiment 1. We are aware that from a methodological point of view it would have been more appropriate to include only matched pairs of participants within the age range applying to the healthy controls in Experiment 1. Future research should aim at inclusion of larger numbers of pairs within the same age category as the healthy reference group.

It could be argued that hearing loss in participants should have been established in advance in order to match neurological and healthy participants for hearing loss. However, hearing loss might be associated with neurological disease. That is, recent research results indicate that neurological diseases such as stroke are frequently associated with hearing loss (Lee, 2009; Gopinath et al., 2009; Kim & Lee, 2010). The larger hearing loss in the neurological group might be caused by vascular involvement in case of stroke and is not necessarily an independent constraint in our neurological participants. In order to confirm the association between neurological disease and sensorineural hearing loss however, it would have been appropriate to establish the absence of premorbid hearing loss in the neurological participants. In addition, it would have been more accurate to explicitly define 'age adjusted normal hearing' in the healthy controls as an explicit inclusion criterion for our study.

4.4.5 Directions for future research

Overall, the current study provides a sound basis for language specific development of an assessment tool for auditory speech discrimination. Further research should address the development of sufficient items (n=15) per subtest that meet the sensitivity criterion of minimally 80% and below 100% scoring healthy controls. In addition, comparison of larger numbers of matched neurological and healthy participants is needed to establish whether auditory discrimination in neurological patients is significantly diminished. Furthermore, included neurological participants should be more homogeneous with respect to underlying neurological disease to control for interfering effects of various concomitant deficits. In addition, future research should address the assessment of cognitive functions, motor arm-hand functioning or any other constraints in neurological patients that are likely to affect ADT performance. In fact, clarity about the contribution of additional deficits to auditory discrimination performance is of vital clinical relevance. It provides cues to support patients carrying out the EST training procedure.

On the longer term, language-specific development of an auditory speech discrimination test would provide the opportunity for research addressing cross-language comparison of auditory discrimination skills in neurological patients.

4.5 Conclusions

An empirical method was used for the development of an auditory discrimination test to assess auditory speech discrimination skills in neurological patients with acquired dysarthria as a prelude to E-learning based Speech Therapy (EST). Several items of the ADT did not meet the sensitivity criterion that was determined in advance and were therefore considered unsuitable. Neurological participants' poorer performance on the remaining suitable items compared to their matched healthy controls are encouraging to further investigate the ADT's sensitivity to diminished auditory speech discrimination. The development of sufficient suitable items per subtest should be addressed in future research.

The level of auditory discrimination performance in neurological patients is likely to be influenced by many individual factors. The effect of hearing loss and other concomitant constraints should be investigated in studies with larger and more homogeneous groups of neurological patients. The clinical relevance of the results is lying in establishing the need to provide neurological patients with visual feedback to support auditory speech discrimination in the EST training procedure.

Acknowledgements

The authors wish to thank Marina Ruiter for her suggestions for adjustments on an earlier draft of this paper. In addition, the authors wish to thank Alex Teulings and Gaetano Ambrosino for their assistance with the figures and tables in this article.

References

Baayen, H., Piepenbrok, & Gulikers, L.. (1995). The CELEX database consortium (CD-ROM). Linguistic data. 2nd release. Max Planck Institute for Psycholinguistics, Nijmegen (the Netherlands).

Bamiou, D. E. (2008). Auditory processing disorders in neurological patients and patients with developmental disorders. *Annals of General Psychiatry, 7,* S12.

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2001). Aetiology and clinical presentations of auditory processing disorders -a review. *Archives of Disease in Childhood, 85*, 361-365.

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of reil) and its role in auditory processing. LIterature review. *Brain Research Reviews, 42,* 143-154.

Bamiou, D. E., Musiek, F. E., Stow, I. S. J., Cipolotti, L., Brown, M. M., & Luxon, L. M. (2006). Auditory temporal processing deficits in patients with insular stroke. *Neurology, 67,* 614-619.

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010a). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) as a web application for speech training in dysarthric patients with Parkinson's Disease: a case-study. *Telemedicine and e-Health, 16,* 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M., Slangen, R. M. L., Heuvel van den, H., Swart de, B. J. M. & Geurts, A.C.H. (2010b). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16,* 177-180.

Benjamin, Y. & Hochberg, Y. (1995). Controlling false discovery rate. *Journal of the Royal Statistical Society, 57,* 289-300.

Binnenpoorte, D., Van Bael, C., den Os, E., & Boves, L. (2005). Gender in everyday speech and language: A corpus-based study. In *Proceedings interspeech 2005.*

Boersma, P. & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.1.05) [Computer software].

Chen, A. J., Gussenhoven, C., & Rietveld, T. (2001). Language-specific effects of pitch range on the perception of universal intonational meaning. In P. Dalgaard, B. Lindberg, & Benner.H. (Eds.), (pp. 1403-1406). Aalborg.

De Bodt, M., Huici Hernandez-Diaz, M., & Van de Heyning, P. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. Journal of Communication Disorders, 35, 283-292.

De Pijper, J. R. (1983). Modelling British English Intonation. Unpublished PhD thesis, University of Utrecht.

Dellifraine, J. & Dansky, D. (2008). Home-based telehealth: a review and meta-analysis. *Journal of Telemedicine and Telecare, 14,* 62-66.

Demanez, L. & Demanez, J. P. (2003). Central auditory processing assessment. *Acta Oto-rhino-laryngologica Belgica, 57,* 243-252.

Durrant, J. D. & Lovrinic, J. H. (1977). *Bases of Hearing Science.* Baltimore: The Williams & Wilkins Company.

Golding, M., Carter, N., Mitchell, P., & Hood, L. (2004). Prevalence of central auditory processing (CAP) abnormality in an older Australian population: the Blue Mountains Hearing Study. *Journal of the American Academy of Audiology, 15,* 633-642.

Golding, M., Taylor, A., Cupples, L., & Mitchell, P. (2006). Odds of demonstrating auditory processing abnormality in the average older adult: The blue mountains hearing study. *Ear and Hearing, 27,* 129-138.

Gopinath, B., Schneider, J., Rochtchina, E., Leeder, S. R., & Mitchell, P. (2009). Association between age-related hearing loss and stroke in an older population. *Stroke, 40,* 1496-1508.

Grawemeyer, B., Cox, R., & Lum, C. (2000). AUDIX: a knowledge-based system for speech-therapeutic auditory discrimination exercises. *Studies in health technology and Informatics, 77,* 568-572.

Guehl, D., Burbaud, P., Lorenzi, C., Ramos, C., Bioulac, B., Semal, C. et al. (2008). Auditory temporal processing in Parkinson's disease. *Neuropsychologia, 46,* 2326-2335.

Gussenhoven, C. & Rietveld, A. (1998). On the speaker-dependence of the perceived prominence of F0-peaks. *Journal of Phonetics, 26,* 371-380.

Gussenhoven, C. & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech, 43,* 183-203.

Hausler, R. & Levine, R. (2000). Auditory dysfunction in stroke. *Acta Otolaryngology, 120,* 689-703.

Hermes, D. J. & Gestel van, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America, 90,* 97-102.

Hill, A. J., Theodoros, D. G., Russell, T. G., Cahill, L. M., Ward, E. C., & Clark, K. M. (2006). An internet-based telerehabilitation system for the assessment of motor speech disorders: a pilot study. *American Journal of Speech-language Pathology, 15,* 45-56.

Hill, J. & Theodoros, D. (2002). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare, 8,* 187-196.

Hoppestad, B. (2006). Essential elements for assessments of persons with severe neurological impairments for computer access utilizing assistive technology devices: A Delphi study. *Disability and Rehabilitation, 1,* 3-16.

Ilvonen, T.-M., Kujala, T., Tervaniemi, M., Salonen, O., Näätänen, R., & Pekkonen, E. (2001). The processing of sound duration after left hemisphere stroke: event-related potential and behavioural evidence. *Psychophysiology, 38,* 622-708.

Kent, R. D. (1992). *Intelligibility in Speech Disorders. Theory measurements and management.* John Benjamins Publishing Company.

Kim, J. S. & Lee, H. (2010). Inner ear dysfunction due to vertebrobasilar ischemic stroke. *Seminars in Neurology, 29,* 534-540.

Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences*. London: Brooks/Cole Publishing Company.

Kraayeveld, J. (1997). *Idiosyncrasy in prosody: Speaker and speaker group identification in Dutch using melodic and temporal information.* Unpublished PhD thesis, University of Nijmegen.

Laures, J. S. & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. Journal of Speech, Language , and Hearing Research, 42, 1148-1156.

Lee, H. (2009). Neuro-otological aspects of cerebellar stroke syndrome. *Journal of Clinical Neurology, 5,* 65-73.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358-368.

Maassen, B. & Povel, D. J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America, 78,* 877-886.

Mashima, P. S. & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and -eHealth, 14,* 1101-1117.

Metz, D. E., Samar, V. J., Schiavetti, N., & Sitler, R. W. (1990). Acoustic dimensions of hearing-impaired speakers' intelligibility: segmental and suprasegmental characteristics. *Journal of Speech and Hearing Research, 33,* 476-487.

Metz, D. E., Samar, V. J., Schiavetti, N., Sitler, R. W., & Whitehead, R. L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. *Journal of Speech and Hearing Research, 28,* 345-355.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing.* (5th ed.) London: Academic Press.

Moulines, E. & Charpentier, F. (1990). Pitch-synchronous wave-form processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453-467.

Neijenhuis, K. A., Stollman, M. H., Snik, A. F., & Van der Broek, P. (2001). Development of a central auditory test battery for adults. *Audiology, 40,* 69-77.

Palmer, R., Enderby, P., & Cunningham, S. (2004). The effect of three practice conditions on the consistency of chronic dysarthric speech. *Journal of Medical Speech-Language Pathology, 12,* 183-188.

Pichora-Fuller, M. K. & Singh, G. (2006b). Effects of age on auditory and cognitive processing: implicaions for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification, 10,* 29-59.

Pichora-Fuller, M. K. & Singh, G. (2006a). Effects of age on auditory and cognitive processing: implicaions for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification, 10,* 29-59.

Pichora-Fuller, M. K. & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42,* 2S11-2S16.

Piron, L., Turolla, A., Tonin, P., Piccione, F., Lain, L., & Dam, M. (2009). Satisfaction with care in post-stroke patients undergoing a telerehabilitation programme at home. *Journal of Telemedicine and Telecare, 14,* 257-260.

Plomp, R. & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

Quené, H. (2008). Multilevel modelling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustic Society of America, 123,* 1104-1113.

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics, 35,* 353-362.

Quené, H. (2005). Modelling of between-speaker and within-speaker variation in spontaneous speech tempo. Interspeech.Proceedings (pp. 2457-2460).

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow up. *Journal of Neurology, Neurosurgery and Psychiatry,* 493-498.

Rietveld, A. C. M. & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics, 13,* 299-308.

Rietveld, T. & van Heuven, V. (2009). *Algemene fonetiek*. Bussum: Coutinho. Third edition.

Rietveld, T. & van Hout, R. (2005). *Statistics in Language Research: Analysis of Variance.* Berlin: Mouton de Gruyter.

Rooij J. C. G. M. van & Plomp, R. (1991). Auditive and cognitive factors in speech perception by elderly listeners. *Acta Oto-Laryngologica, 111*, 177-181.

Rvachew, S. (2004). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*, 347-357.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburg.

Smith, B. (2002). Effects of speaking rate on temporal patterns of English. *Phonetica, 9*, 232-244.

Sorin, C. (1981). Functions, roles and treatments of intensity in speech. *Journal of Phonetics, 9*, 359-374.

Straatman, L. V., Rietveld, A. C. M., Beijen, J., Mylanus, E. A. M., & Mens, L. H. M. (2010). Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. *Journal of the Acoustical Society of America, 128*, 1884-1895.

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60*, 498-500.

Van Hessen, A. (1987). *Spectral interpolation* (Rep. No. 12 Progress Report Institute of Phonetics Utrecht).

Willott, J. F. (1991). *Aging and the auditory system: anatomy, physiology, and psychophysics*. London: Whurr Publishers Ltd.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*, 550-560.

Appendix 4A

Overview of the equal and unequal items in all subtests

Segmental elements:					
		Stimulus 1	**Stimulus 2**		
		** = Manipulated phonemes (underlined)*	**Default version**	**(phonetic transcripion)**	**Equal/unequal pair**
	Items				
	example a	*Piep*	*/piep/*	*[pɪ:p]*	*equal*
	example b	*Koek/ə **	*/koek/*	*[ku:k]*	*unequal*
	1	Kaak	/kaak/	[ka:k]	equal
	2	Kaak *	/kaak/	[ka:k]	unequal
	3	Piep *	/piep/	[pɪp]	unequal
	4	Kies	/kies/	[kis]	equal
	5	Map *	/map/	[mɑp]	unequal
	6	Soep *	/soep/	[sup]	unequal
	7	Bak	/bak/	[bɑk]	equal
O	8	Miep *	/miep/	[mip]	unequal
	9	Koek	/koek/	[kuk]	equal
	10	Paal	/paal/	[pa:ɫ]	equal
O	11	Dak	/dak/	[dɑk]	equal
	12	Kaak*	/kaak/	[kaˑk]	unequal
	13	Tak	/tak/	[tɑk]	equal
▶	14	Soms *	/soms/	[s:ɔms]	unequal
	15	Das *	/das/	[dɑs]	unequal

(▶ = < 80% score, O = 100% score)

Loudness				
		Stimulus 1	**Stimulus 2**	
		Default version	* = *Manipulated*	**Equal/unequal pair**
	Items			
	example a	*Manier*		*equal*
	example b	*Seizoen*	*+ 6dB **	*unequal*
O	1	Ham		equal
	2	Naam	+ 6dB *	unequal
O	3	Wiel	+ 9dB *	unequal
	4	Haring		equal
O	5	Ballon		equal
	6	Ruzie	+ 6dB *	unequal
	7	Banaan	+ 6dB *	unequal
	8	De kat van de buren is weg		equal
O	9	De wagen reed snel de berg af		equal
O	10	De kamer rook naar sigaren	+ 6dB *	unequal
O	11	Gisteren waaide het nog harder		equal
	12	De foto is mooi ingelijst	+ 6dB *	unequal
O	13	Het gras was helemaal verdroogd	+ 6dB *	unequal
	14	De kleren waren niet gewassen		equal
	15	De grijze lucht voorspelt regen	+ 6dB *	unequal

(▶ = *< 80% score,* O = *100% score)*

Overall Pitch				
		Stimulus 1	**Stimulus 2**	
		Default version	** = Manipulated (ST = semitones)*	**Equal/unequal pair**
	Items			
	example a	*Mening*	*+ 5ST (male) / +4ST (female) **	*unequal*
	example b	*De jongen ging er gauw vandoor*		*equal*
	1	Zien		equal
	2	Dam	+ 4ST (male) / + 3ST (female) *	unequal
▶	3	Doen	- 7ST (male) / - 6ST (female) *	unequal
	4	Leraar		equal
O	5	Leuning	+ 6ST (male) / + 5ST (female) *	unequal
	6	Gordijn	- 7ST (male) / - 6ST (female) *	unequal
O	7	Mobiel		equal
	8	Hij kan er nu eenmaal niets aan doen		equal
	9	De oude man was kaal geworden		equal
	10	De nieuwe fiets is gestolen	- 7ST (male) / - 6ST (female) *	unequal
O	11	Rijden onder invloed is strafbaar		equal
▶	12	De schoenen moeten verzoold worden	- 7ST (male) / - 6ST (female) *	unequal
	13	Moeizaam klom de man naar boven		equal
	14	Het regent al de hele dag	+ 4ST (male) / + 3ST (female) *	unequal
	15	Door de neus ademen is beter	+ 4ST (male) / + 3ST (female) *	unequal

(▶ = < 80% score, O = 100% score)

Speech rate				
		Stimulus 1	**Stimulus 2**	
		Default version	** = Manipulated*	**Equal/unequal pair**
	Items			
	example a	*Lawaai maakt je op den duur doof*	*faster **	*unequal*
	example b	*Ze heeft haar proefwerk slecht gemaakt*	*slower **	*unequal*
	1	Zijn leeftijd ligt boven de dertig		equal
	2	Voor het eerst was er nieuwe haring	faster *	unequal
	3	Mijn buurman heeft een auto gekocht		equal
	4	Steile trappen zijn gevaarlijk		equal
	5	Door zijn snelheid vloog hij uit de bocht	faster *	unequal
	6	Mijn broer gaat elke dag fietsen	faster *	unequal
O	7	De bomen waren helemaal kaal	slower *	unequal
O	8	Het tempo ligt voor hem veel te hoog		equal
O	9	Het geschreeuw is duidelijk hoorbaar	slower *	unequal
	10	De rivier trad buiten haar oevers		equal
	11	De appels aan de boom zijn rijp	faster *	unequal
	12	Door zijn haast maakte hij veel fouten	slower *	unequal
	13	Zijn manier van werken ligt mij niet		equal
O	14	Zijn gezicht heeft een rode kleur	slower *	unequal
	15	Dat hotel heeft een slechte naam		equal

(▶ = < *80% score,* O = *100% score)*

Intonation				
		Stimulus 1	**Stimulus 2**	
		Default version	** = Manipulated (ST = semitones)*	**Equal/unequal pair**
	Items			
	example a	De zon ging in het westen onder		equal
	example b	De jongens vechten de hele dag	+ 7ST (male) / + 7ST (female) *	unequal
	1	De aardappels liggen in de schuur	+ 7ST (male) / + 5ST (female) *	unequal
	2	De appelboom stond in volle bloei		equal
	3	De biefstuk is vandaag erg mals		equal
	4	De kast was één meter verschoven		equal
	5	De tuinman heeft het gras gemaaid	+ 7ST (male) / + 5ST (female) *	unequal
	6	De bel van de voordeur is kapot		equal
	7	De appel had een zure smaak	+ 7ST (male) / + 5ST (female) *	unequal
	8	Het vliegtuig vertrekt over een uur	+ 7ST (male) / + 7ST (female) *	unequal
O	9	De hond blafte de hele nacht		equal
	10	De sigaar ligt in de asbak		equal
▶	11	De volgende dag kwam hij ook niet	+ 7ST (male) / + 5ST (female) *	unequal
	12	Haar gezicht was zwart van het vuil	+ 7ST (male) / + 5ST (female) *	unequal
	13	Het gedicht werd voorgelezen	+ 7ST (male) / + 5ST (female) *	unequal
	14	De witte zwaan dook onder water	+ 9ST (male) / + 7ST (female) *	unequal
	15	Hij nam het pak onder zijn arm		equal

(▶ = *< 80% score*, O = *100% score)*

Appendix 4B

Appendix 4B:

Actual and modified speech rates (syll/min) in the subtest 'Speech Rate'

	Sentence	Equal/ unequal pair	Speech rate of items within pair		Difference within pair *(syll/min)*
			Item 1 *(syll/min)*	**Item 2** *(syll/ min)*	
Ex1	*Lawaai maakt je op de duur doof*	*unequal (acceleration)*	*209*	*251*	*-42*
Ex2	*Ze heeft haar proefwerk slecht gemaakt*	*unequal (deceleration)*	*199*	*165*	*34*
1	Zijn leeftijd ligt boven de dertig	equal	233	233	0
2	Voor het eerst was er nieuwe haring	unequal (acceleration)	262	331	-69
3	Mijn buurman heeft een auto gekocht	equal	241	241	0
4	Steile trappen zijn gevaarlijk	equal	229	229	0
5	Door zijn snelheid vloog hij uit de bocht.	unequal (acceleration)	204	257	-53
6	Mijn broer gaat elke dag fietsen	unequal (acceleration)	225	282	-57
7	De bomen waren helemaal kaal	unequal (deceleration)	275	222	53
8	Het tempo was voor hem veel te hoog	equal	234	234	0
9	Het geschreeuw is duidelijk hoorbaar	unequal (deceleration)	231	180	51
10	De rivier trad buiten haar oevers.	equal	243	243	0
11	De appels aan de bomen zijn rijp.	unequal (acceleration)	229	294	-65
12	Door zijn haast maakte hij veel fouten	unequal (deceleration)	231	186	45
13	Zijn manier van werken ligt mij niet	equal	229	229	0
14	Zijn gezicht heeft een rode kleur	unequal (deceleration)	221	164	57
15	Dat hotel heeft een slechte naam	equal	244	244	0

CHAPTER 5

Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease.

L.J. Beijer, R.P. Clapham, A.C.M. Rietveld

Adapted from: *Journal of Medical Speech-Language Pathology,* 2012, 20 (2): 17-34

Abstract

Purpose: Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers due to Parkinson's disease (PD).

Method: A single SUS set was developed for Experiment 1 and 5 SUS sets for Experiment 2. Each set contained ten 6-word declarative sentences (D6), ten 13-word declarative sentences (D13) and ten question sentences (Q). For Experiment 1 sentences were read aloud by 1 speaker with PD and dysarthria, whereas 5 speakers read the sentences for Experiment 2. Ten speech-language pathologists completed two intelligibility scoring tasks for the audio-recorded sentences: orthographic transcription and 10-point scale rating. In Experiment 1, the interrater reliability for both intelligibility measures was established. In Experiment 2, the effects of 'SUS set', 'listener' and 'speaker' on both measures were investigated for all sentence forms.

Results: Only orthographic transcription of D6 sentences met the requirements for efficacy research: the factors 'SUS set' and 'listener' did not affect transcription scores, but the factor 'speaker' did.

Conclusions: The initial results imply that the D6 SUS sets are equivalent with respect to potential intelligibility and sensitive to different degrees of intelligibility as measured by orthographic transcription scores, and may be suitable intelligibility measures in speech training efficacy research.

5.1. Introduction

Currently, in our aging population, numerous efforts in health care are aiming at the maintenance of a sound balance between the need for and the availability of health care. These efforts tend to yield innovative interventions which frequently employ new technologies, treatment protocols, and procedures. Obviously, when new approaches for treatment are introduced, establishing scientific evidence of the expected benefits is inevitable.

Researchers and practitioners in the field of communication disorders face critical pressure to establish scientific evidence for allegedly beneficial treatments. That is, concerns for objective accountability underpinning clinical reimbursement and research funding underscore the relevance of form and quality of clinical outcome studies (Robey & Schultz, 1998). The clinical-outcome research community conducts studies along an organized structure of a model which is mainly based on the concepts of 'efficacy' and 'effectiveness' (Robey, 2004). Efficacy can be defined as the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use. Effectiveness is defined as the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under average conditions of use. Efficacy research then is conducted under best-possible conditions, whereas effectiveness research is conducted under 'real world circumstances' (Office of Technology Assessment, 1978). The sequence of research tasks in clinical outcome studies is defined by a five-phase model. In Phase I treatment protocols are described, and selected therapeutic effects are identified if present along with an estimation of the magnitude of the effect. In Phase II the dimensions of the therapeutic effect are explored and the required preliminary tests and preparations are made for testing the protocol in a clinical trial. In Phase III a clinical trial is conducted to test efficacy. Once a treatment has been found to be efficacious, in Phase IV the effectiveness of that treatment is tested. Finally, in Phase V the cost-effectiveness is evaluated to establish whether the worth of a treatment justifies the costs (Robey, 2004).

Similar to the area of medical care, the paramedical field of speech-language pathology is facing the challenge of meeting the treatment needs for an expanding group. That is, the number of patients with neurogenic communication disorders is growing along with the aging of our population. Among other innovative intervention protocols and devices, several telerehabilitation applications are emerging as potential solutions to the threatening shortage of therapeutical resources. The benefits of remote diagnostics and therapy might decrease the 'need-and-capacity-gap'. That is, instead of spending time, energy, and money on travelling, neurological speakers are allowed to train their communication skills in their home environment as intensively as they prefer. Therapists can remotely monitor and guide their patients and, hence, increase their therapeutic capacity. Telerehabilitation devices have therefore seen a substantial growth (Brennan, Georgeadis, & Baron, 2002; Mashima & Doarn, 2008)) and are currently subject of investigations into their efficacy (Theodoros, Hill, Russell, Ward, & Wootton, 2008; Ziegler & Zierdt, 2008).

In the Netherlands, a web application for speech training in speakers with dysarthria due to stroke or Parkinson's disease (PD) has been recently developed. This web application, called E-learning based Speech Therapy (EST), enables therapists to remotely compile a tailor-made speech training program for their clients, providing these neurological speakers with diminished speech intelligibility with the possibility to practice speech in their home environment (Beijer, Rietveld, van Beers, Slangen, van den Heuvel, de Swart, & Geurts, 2010b; Beijer, Rietveld, Hoskam, Geurts, & de Swart, 2010a). The dysarthric speakers of our target population, who tend to be less mobile and easily fatigued as a result of their age and their neurological condition, are expected to benefit considerably from the enhanced accessibility of speech training. That is, according to our clinical experience, this chronic group tends to be eager to improve or at least maintain their speech quality in order to optimally participate in our verbally oriented society. However, the potentials of EST for dysarthric speakers have yet to be established.

In the context of preparing efficacy studies for EST as a web application for speech training in patients with PD and dysarthric speech (i.e. Phase I and Phase II research), speech intelligibility seems an obvious outcome measure since it is often used as an index for both dysarthria severity and treatment effects (Yorkston, Hammen, Beukelman, & Traynor, 1990; Kennedy, Strand, & Yorkston, 1994; Strand & Yorkston, 1994). At this point it is vital to distinguish 'intelligibility' from 'comprehensibility'. 'Intelligibility' refers to speaker-dependent information (i.e., the acoustic speech signal), whereas 'comprehensibility' is determined by both speaker-dependent and speaker-independent (e.g., contextual) information (Yorkston, Strand, & Kennedy, 1996). With speech intelligibility being the changeable target behavior to be observed for EST efficacy research, we need primary outcome measures that adequately register intelligibility (Beukelman, Mathy, & Yorkston, 1998; Wertz & Irwin, 2001) in our target group of dysarthric speakers caused by PD.

To obtain outcome measures for speech intelligibility, researchers tend to employ both scaling methods and identification methods, using different types of speech tasks, speech stimuli, speech sample presentation modes, response types and transcription analyses (Barreto & Ortiz, 2008). For both scaling and identification methods, the ecological validity of speech materials should be considered. The ecological validity concerns the requirement that speech materials employed for speech measurements should represent daily speech in some respects. It will be obvious that finding an appropriate balance between ecologically valid speech materials and research feasibility is a real challenge. Ideally, spontaneous speech obviously would provide the most representative sample of a patient's speech beyond the clinical setting. However, comparing different samples of spontaneous speech across and within speakers is a laborious task. Large corpora of speech would be required to compensate for possible biases due to different speech samples and to warrant comparisons between and within subjects. This is a vital disadvantage of employing spontaneous speech for intelligibility outcome measurements. Therefore, it could be argued that speech materials at sentence level, although less representative for daily communication than spontaneous speech samples, meets the condition for ecologically valid speech materials in some respects. That is, compared to isolated words, speech at sentence level is more likely to reflect natural speech with respect to utterance length and suprasegmental speech dimensions such as speech rate and prosodic patterns. In addition, in the context of EST efficacy research, sentence level materials represent the EST practice content, which also contains training at sentence level (Beijer et al., 2010b; Beijer et al., 2010a). However, despite the benefits of speech at sentence level for intelligibility outcome measurements, a vital drawback of most developed sentence level material is the lack of control for signal-independent (e.g. semantic) information, which aids signal identification by listeners (Hustad, 2007; Yorkston & Beukelman, 1981; Yorkston et al., 1996). Hence, outcome measures based on sentence level materials are liable to reflect comprehensibility rather than intelligibility.

To benefit from speech material at sentence level and at the same time reduce the disadvantage of sentence predictability based on contextual information, orthographic transcription of semantically unpredictable sentences (SUS) seems suitable. Being syntactically correct yet semantically unpredictable (e.g. 'A job went through a flight' or 'Did a port fall out of a box?'), these sentences avoid interfering effects of signal-independent information such as semantic context (Benoît, Grice, & Hazan, 1996; Jekosch, 2005), thus leaving listeners with only speaker-dependent information. We have to be aware however, that in repeated measures designs for EST efficacy research, the employment of orthographic transcription on only limited SUS materials is at risk for rater learning effects.

This potential learning effect may interfere with genuine intelligibility assessment.

Rater learning effects could be avoided by applying objective (instead of subjective) identification methods based on automatic speech recognition (ASR) technology. Adapted forms of ASR are currently being applied in computer assisted language learning (CALL), providing methods for automatic pronunciation scoring in second language (L2)-learners (van Doremalen, Cucchiarini, & Strik, 2009). Based on confidence measures such as the likelihood-based 'Goodness of Pronunciation' (GOP), automatic pronunciation scores are obtained (Witt & Young, 2000). However, confidence measures for non-native speech tend to be lower because ASR is developed to recognize native speech, with different acoustic properties (Doremalen van et al., 2009). For the same reason, confidence measures for dysarthric (i.e., deviant) speech are likely to be lower than for healthy speech. Therefore, an ASR engine should be trained with speech from healthy speakers and speech from the target population with dysarthric speech. Thus, the reference acoustic models (i.e., from the healthy speakers) should be adapted for dysarthric speech. Although developments in the area of automatic pronunciation assessment of dysarthric speech are currently underway (Middag, Martens, Van Nuffelen, & De Bodt, 2009; Sanders, Ruiter, Beijer, & Strik, 2002), robust automatic pronunciation assessment at sentence level is still lacking. Hence, as far as identification methods for intelligibility assessment are concerned, for now we have to rely on subjective scorings such as orthographic transcription scores of SUS. To reduce rater learning effects in repeated measures for efficacy research, different yet equivalent sets of semantically unpredictable sentences (henceforth: SUS sets) are required. That is, although the SUS sets contain different sentences, the equivalence of different sets with respect to potential intelligibility is required.

This article attends to the development and the suitability of orthographic transcription and intelligibility scale ratings of different SUS sets for employment in EST efficacy research in dysarthric speakers due to PD. A study containing two experiments was conducted. In Experiment 1, our research question concerned the interrater reliability of 10 raters with respect to orthographic transcription scores and intelligibility scale rating of SUS. This was a condition for conducting Experiment 2, where three research questions were at issue: 1) are SUS scores sensitive to dysarthric PD speakers with different degrees of speech intelligibility? 2) are SUS sets equivalent with respect to potential intelligibility? 3) Is 'raters' a factor that affects SUS intelligibility scores?

In the next section two experiments will be reported. Ethical approval was obtained by the Regional Committee of Medical and Health Research Ethics.

5.2 Experiment 1

Experiment 1 was conducted to establish the interrater reliability of 10 trained listeners for two subjective intelligibility scoring measures: orthographic transcription scores and intelligibility scale ratings. Sufficient interrater reliability for both outcome measures was required to allow an orthogonal latin square design in Experiment 2. Since interrater reliability for intelligibility scale ratings are an issue of discussion (Bunton, Kent, Duffy, Rosenbek, & Kent, 2007), establishment of this outcome measure was considered of particular interest. Intelligibility scorings were based on an initial set of semantically unpredictable sentences (henceforth: 'SUSO set'), which were read aloud by a dysarthric speaker with PD. The SUSO set contained three different sentence types: two declarative sentence forms (short and long) and one question form. Each syntactic type contained 10 sentences. Interrater reliability analyses were conducted for each syntactic type .

5.2.1 Materials and methods

Participants
One dysarthric speaker with PD and 10 trained listeners, speech-language pathologists (SLPs), were involved in this experiment.

Speaker.
A 70-year old male dysarthric speaker with PD participated. The speaker was recruited by a SLP with expertise in providing the Pitch Limiting Voice Treatment (PLVT) for parkinsonian speakers (Swart de, Willemse, Maassen, & Horstink, 2003), a Dutch adaptation of the Lee Silverman Voice Treatment (LSVT) (Ramig et al., 2001). The speaker was selected based on our inclusion criterion of representing a moderate-to-severe degree of hypokinetic dysarthria. Severe cognitive deficits or physical restrictions, that were exclusion criteria, were not observed.

Listeners.
Ten female speech-language pathologists (SLPs) from a local rehabilitation center participated as listeners in this study (mean age, 33 years; standard deviation (SD), 6.9; range, 25–44 yrs). Average length of employment as an SLP was 9.8 years. All SLPs had

therapeutic experience with children and adults with dysarthric speech. They did not have specific experience in perceptual judgment on parkinsonian dysarthria since these speakers were referred to an academic expertise center for Parkinson's disease nearby the rehabilitation center. All listeners were native speakers of the Dutch language. None of the listeners reported hearing concerns and all were aware that participation in Experiment 1 would lead to participation in Experiment 2.

Speech materials

An initial set of 30 semantically unpredictable sentences (SUSO) was composed and used as speech material in Experiment 1. An overview of the SUS generation process, with a number of linguistic conditions taken into consideration, is presented in Figure 5.1.

Syntactic structures.

Syntactic structures for the generation of SUS sentences were selected based on arguments for SUS generation as described in the Speech Output Assessment Package (SOAP) document (ESPRIT project 2589). In this document, five syntactic structures allowing cross-language comparison of intelligibility results for synthesized speech were addressed. For our study, only two of these structures were used. First, a declarative sentence form with an intransitive verb was selected. This structure, rather than declarative forms with transitive verbs, or forms with imperative or relative sentence structure, contains basic word constituents (i.e. nouns, verbs and prepositions) that frequently occur in everyday speech. Deviant speech realization of these content words by dysarthric PD speakers is likely to affect sentence intelligibility to a larger extent than function words with little lexical meaning. A second sentence type selected for our study concerns similar word constituents embedded in a question form, requiring interrogative prosodic expression. Both the declarative form as well as the question form, were assumed to be sensitive to deviant speech in patients with PD. That is, features of parkinsonian dysarthric speech, such as imprecise articulation, prosodic abnormalities, disturbances of speech rate and reduced speech intensity (Kempler & van Lancker, 2002; Pinto et al., 2004) are expected to affect speech intelligibility of these sentences containing content words.

Basic sentence length was based on accommodating the limitations of both the dysarthric speaker's articulatory load and the listeners' working memory capacity. To meet the limitations of the verbal working memory capacity in healthy listeners, they were presented with six-word sentences (Miller, 1956). To this end, Q-sentences and

D-sentences containing six words were developed. In addition however, for declarative forms both sentences containing six words (D6) and sentences containing 13 words (D13) were developed. In D13 sentences, two 6-word clauses were connected by a coordinating conjunction term, resulting in longer sentences and, thus, increased articulatory load of the speaker. With the speaker's realization of the entire 13-word sentence being recorded, only the second clause (containing six words) was presented to the listeners. Obviously, sentence length and syntactic complexity are - nearly by definition - confounded since longer sentences often consist of more than one clause, requiring the insertion of conjunctions and thus adding to the grammatical complexity of the sentence. In the longer D13 sentences however we limited the syntactical complexity by inserting the most simple conjunction 'en' (English: 'and') to connect two syntactically simple clauses.

Word classes.
To limit word predictability, word classes were carefully selected for inclusion in SUS. Adjectives and definite articles were avoided since, in Dutch, adjectives are inflected by the noun that they modify and nouns only take a specific definite article. Each target sentence presented to the listeners (i.e. in D13 sentences only the second clause) contained four keywords: one verb, two nouns and one preposition. The remaining two words concerned two indefinite articles, ensuring a complete sentence.

Nouns, verbs and prepositions included in the sentences were selected on the basis of word frequency, consonant cluster categories and initial phoneme class as provided by the written Dutch corpus CELEX , version 3.1 (Baayen, Piepenbrok, & Gulikers, 1995). Word frequency in the CELEX database was set at 1000, aiming at a balance between high frequency words (enhancing speech naturalness but increasing predictability) and low frequency words (decreasing speech naturalness due to potential speaker hesitation, but decreasing predictability). To limit word identification assistance and to ensure that all sentences were of basic comparable articulation difficulty, we only included monosyllabic words, which is in line with methods commonly employed in the evaluation of speech synthesis systems. Noun and verb consonant cluster categories were divided in four consonant cluster groups: single consonants (Consonant Vowel: CV, or Consonant Vowel Consonant: CVC), double consonants (CCV, CCVC or CVCC), triple consonants (CCCVC or CVCCC) and mixed consonants (CCVCC, CCCVCC or CCVCCC). The noun and verbs of each syntactic category (i.e. D6, D13 and Q) in the SUSO set contained 40% of single consonants and 54% of double consonant clusters. Triple consonants and mixed consonants both accounted for only 3% of the consonant clusters across the sentences per sentence type, reflecting the dearth of monosyllabic words with these consonant clusters in Dutch (Luyckx, Kloots, Cousse, & Gillis, 2007).

Initial phonemes.

With respect to initial phonemes, nouns and verbs were balanced across each set of the 3 syntactic sentence types. That is, each set of 10 sentences for each syntactic structure comprised 73% of nouns and verbs from the plosive group (40%) and the fricative phoneme group (33%), and the remaining 27% contained words from the other phoneme groups (i.e. 10% nasals, 10% liquids, and 7% approximates). This weighting toward fricatives and plosives reflects the two primary areas of articulation difficulty for speakers with PD, characterized by hypokinetic dysarthria (Ackermann & Ziegler, 1991) .

Word insertion procedure.

Since it appeared impossible to control for all linguistic conditions mentioned above with an automatic procedure, the SUS sentences were created by hand on the basis of word lists from the CELEX corpus. To this end, a word matrix was developed where words were grouped based on their initial phonemes (plosive, fricative, liquid, approximate and nasal). This selection process used ensured that a cross-section of words from each phoneme category was present in each sentence group. We also used two additional criteria to guide word placement. First, if a sentence already contained a noun or a verb with the same initial phoneme, the word awaiting placement was placed in the next available position within that sentence type. Second, if a sentence already contained a semantically related word, the word awaiting placement was moved to the next available position within that sentence type. Insertion of verbs, nouns and prepositions in the selected syntactic structures for sentence generation was as random as possible. Some intentional selection was necessary to meet phoneme class and consonant structure requirements. In addition, coincidentally generated compound words (e.g. the Dutch word 'nood sprong' , which is comparable to for instance the English 'sound track') were removed. The consonant structure of the prepositions was not controlled as there are only a handful of mid to high frequency Dutch monosyllabic prepositions.

All generated SUS sentences were subjectively evaluated by two Dutch native speakers (the first and the last author of this paper) to ensure that the sentences were semantically unpredictable yet syntactically correct.

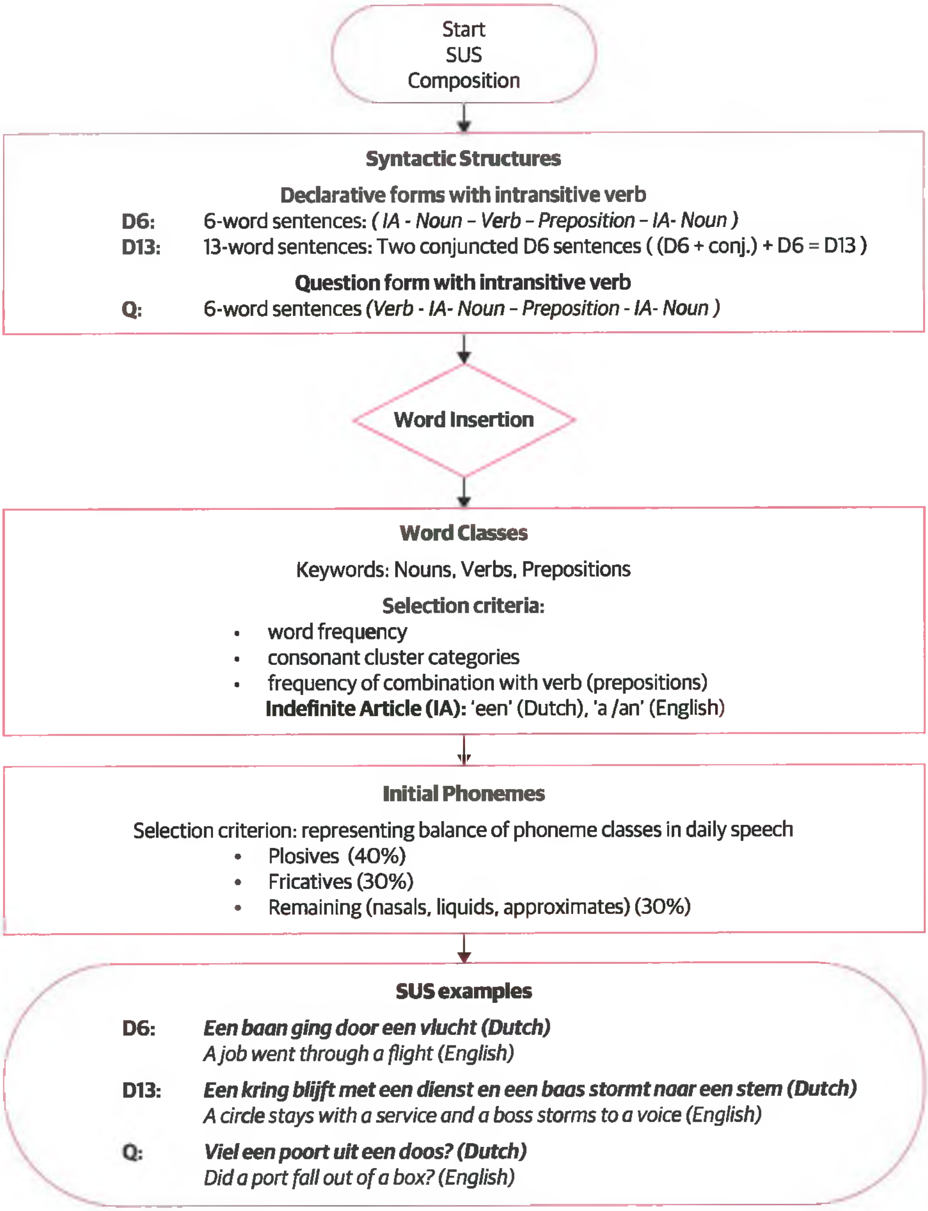


Figure 5.1:

Composition of semantically unpredictable sentences (SUS). IA = Indefinite Article

Procedure

Audio recordings.

All sentences were read aloud by a 70-year-old male speaker with PD and moderate to severe hypokinetic dysarthria. Instructions were provided in written and spoken form. The speaker was instructed to read sentences that were not comprehensible yet grammatically correct and that the sentences contained no nonsense words. He was instructed to silently read each sentence before reading it aloud. Pre-reading the sentence allowed the speaker to become familiar with each sentence, assisting in a natural reading and limiting a reading-list effect. In addition, the speaker was presented with an example sentence to become familiar with the speech material and the reading task. In case of disfluent realization of the target utterance and in case of substituting linguistic elements of the target utterance, the sentence was re-recorded. When asked to repeat a sentence, the speaker was not encouraged to improve speech quality in any way. For audio-recording, a headset with microphone (Sennheiser PC 131) was used. Mouth to microphone distance was approximately 5 cm. The digital recordings were made on a Fujitsu Siemens laptop, by means of the Audacity software, version 1.2.6 (Freeware Software Inc., 2008). The sampling frequency was 44.100 Hz. Recordings of each stimulus sentence were separated into individual sound files. In addition, all audio-files were adjusted to a mean intensity of 65 dB using 'Praat', which is a software package for speech processing (Boersma & Weenink, 2008).

Sentence presentation order.

The sentences were aurally presented to the listeners in three blocks of 10 sentences. A random presentation order of different sentence types was required to decrease the chance of syntactic priming, which is. the phenomenon that the processing of an utterance with a particular syntactic form is facilitated by a preceding utterance with a similar form (Ledoux, Traxler, & Swaab, 2007). To reduce predictability, each sentence type appeared three times in two presentation blocks and four times in the remaining block, whilst sentences of the same 'sentence type' did not appear more than twice directly after each other. Since an initial selection of sentence type was required to achieve an adequate division across presentation blocks, the sentences were in fact pseudo-randomly assigned to the different blocks.

Listening task.

All listeners were provided with a hard copy of the task instructions. They were instructed to orthographically transcribe each sentence and, in addition, rate the extent of perceived intelligibility on a 10-point scale to obtain a judgement on the difficulty involved in the transliteration of the stimulus. The listeners were familiar with a 10-point scale since it is the default assessment scale in the Dutch school system. There is ample evidence that this scale is valid in many conditions (Huinck & Rietveld, 2007). The scale ranged from '1' (extremely poor intelligibility) to '10' (extremely good intelligibility). Ratings below '6' indicate that intelligibility is qualified as insufficient. A completed example of both the transcription task and the rating task was included on the instruction sheet. The listeners were aware that the sentences contained only real words and that the sentences were grammatically correct yet had unusual content. They were unaware of the speaker's neurological condition. Each listener completed the listening experiment individually in a quiet environment via a laptop computer and a headset (Sennheiser PC 131). Sentences were presented using the E-prime software tool, version 2.0 (Schneider, Eschman, & Zuccolotto, 2007), allowing the listeners to self-pace sentence presentation. Each sentence was presented only once, since a single presentation i) may make a subjective measurement more sensitive to achieved intelligibility, and may avoid ceiling effects, and ii) is more ecologically valid, as in real life speakers do not tend to repeat themselves.

Scoring.

The first and the second author of this article scored the orthographic transcriptions by using a keyword decision diagram as depicted in Figure 5.2. The diagram represents scoring decisions for observed orthographic mismatches between target keywords and listener responses. To be scored as correct, transcribed words that did not entirely match the orthographic representation of the target, had to be existing Dutch words and match the same grammatical class as the target. Some exceptions to this initial scoring system were required to avoid interference of other variables with the assessment of speaker intelligibility. That is, decision criteria for correct scores took into account orthographic mismatches due to spelling errors and the effects of phonological and phonetic processes in normal continuous speech such as cluster reduction or assimilation (Ernestus, Baayen, & Schreuder, 2002; Pluymaekers, Ernestus, & Baayen, 2005; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009).

For the transcription task, the maximum score per sentence was 4, based on the total number of keywords per sentence. For all listeners and for each sentence type, the number of correctly identified keywords were tallied up and divided by the number of keywords possible. The proportions of correct transcriptions were expressed in percentages, which were used for analyses. Intelligibility scale ratings for all listeners and each sentence type were also expressed in percentages. As the intelligibility scale was a 10-point scale, for each sentence a maximum score of 10 could be achieved for the rating task.

5.2.2 Experimental design and data analyses

A repeated measures design was used, employing a within subject factor 'sentence type' (fixed, three levels). Orthographic transcription scores and intelligibility scale ratings were dependent variables. When appropriate Huyn-Feldt adjusted p-values are reported.

Mean scores for both orthographic transcriptions and intelligibility scale ratings were calculated for each sentence type. ANOVA's were conducted for both intelligibility outcome measures to investigate the effect of the factor 'sentence type' on the orthographic transcription scores and scale ratings. Alpha level was set at 0.05. Finally, for each sentence type interrater reliability was assessed for both transcription scores and intelligibility scale ratings. To this end, the Intraclass Correlation Coefficients (ICCs) for Average Measures was calculated, based on a two-way random effects model and absolute agreement, with both 'raters' and 'utterances' as random factor (Müller & Buttner, 1994).

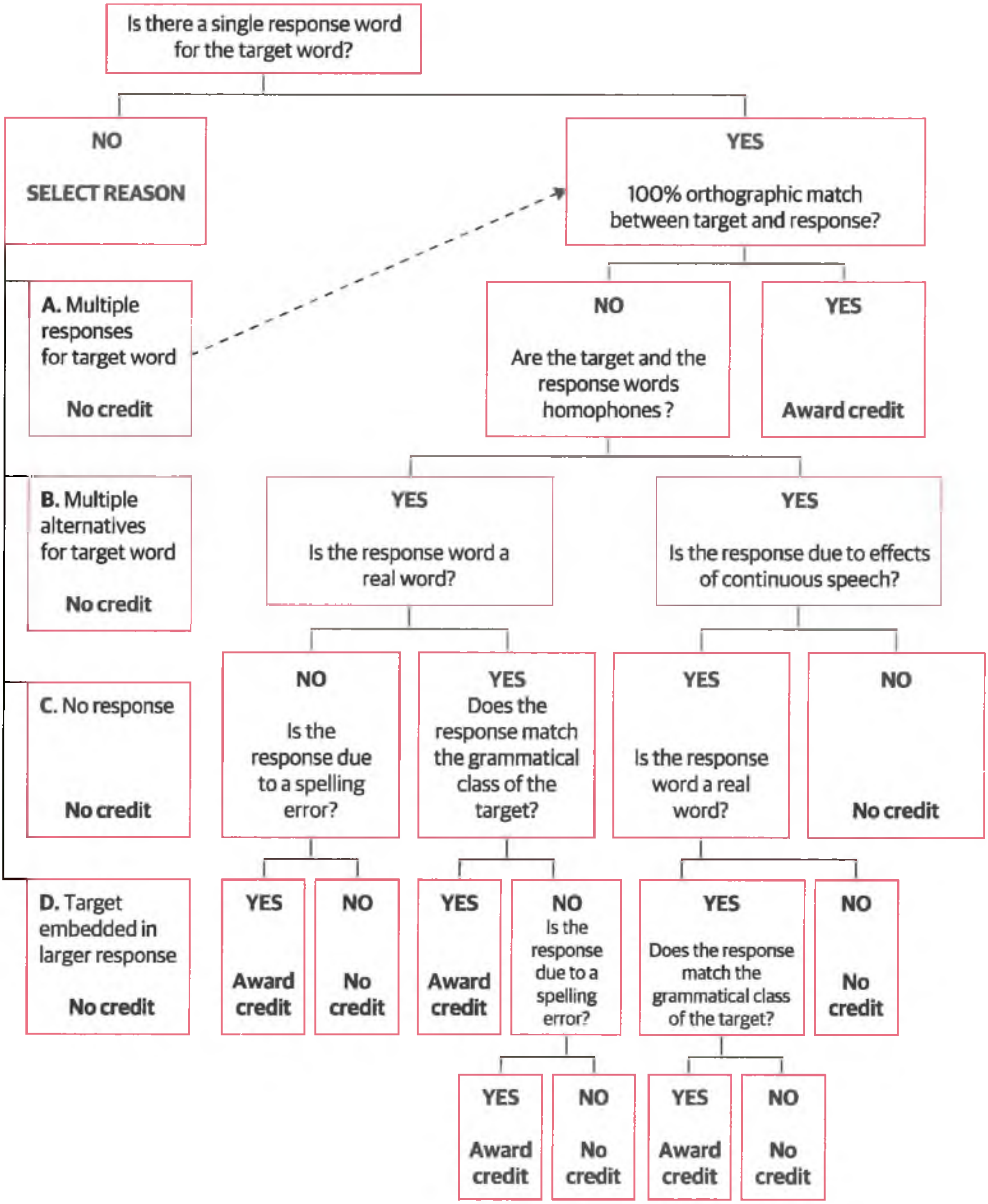


Figure 5.2:

Decision scoring diagram for the orthographic transcriptions of the keywords in the semantically unpredictable sentences (SUS).

5.2.3 Results

For D6 sentences, a mean orthographic transcription score of 86.3% (SD = 16.5%) was obtained. Also for D13 sentences we obtained a mean score of 86.3% (SD = 21.5%). For Q sentences we observed a mean score of 85% (SD = 18.5%). Mean intelligibility scale ratings of 67.2% (SD = 18.9%), 63.9% (SD = 22.2%) and 62.7% (SD = 20.1%) were calculated for D6, D13 and Q sentences respectively. Two separate ANOVA's were carried out to investigate the effect of the within-subject factor 'sentence type' on orthographic transcription scores and on intelligibility scale ratings. The results showed that 'sentence type' did not significantly affect orthographic transcription scores ($F_{2,198} < 1.0$) or intelligibility scale ratings ($F_{2,198} = 2.37$, $p = 0.096$).

The Intraclass Correlation Coefficient, as indices for interrater reliability for orthographic scores as well as for intelligibility ratings are shown in Table 5.1. All reliability results were significant, regardless sentence type or scoring task.

Table 5.1:

Intraclass Correlation Coefficients (ICCs) for the orthographic scores and intelligibility ratings on a 10-point scale

	ICC	
Sentence type	**Orthographic scores**	**Intelligibility ratings**
D6	0.75*	0.60*
D13	0.94*	0.89*
Q	0.83*	0.76*

Notes. D6 = 6-word declarative sentences, D13 = 13-word declarative sentences, Q = question sentences

** statistically significant (p ≤ 0.001)*

5.2.4 Discussion and conclusion

To warrant the employment of a latin square design in Experiment 2, the interrater reliability for both orthographic transcription and intelligibility ratings had to be sufficient. The results of Experiment 1 show that, for all sentence types, the ICC's for intelligibility ratings were lower than for orthographic transcription scores. This is in line with research findings supporting the assumption that subjective intelligibility estimates may not be a reliable substitute for the more objective intelligibility measure provided by word identification through orthographic transcription (Hustad, 2006). This is possibly due to the lack of 'fixed anchors' in listeners for speech intelligibility (Beijer, Rietveld, Hoskam et al. 2010). Nevertheless, researchers have reported relatively high interrater reliability for intelligibility scale ratings in dysarthric speakers with PD (Zraick et al., 2003) and Huntington's Disease (Zraick et al., 2004). It should be noticed, however, that those ratings took place in a highly protocolized context with standardized speech protocol. In addition, the raters were trained medical and nonmedical professionals with expertise in the relevant specific neurological diseases. In our study, although trained listeners were involved, they had neither previous experience with the speech materials to be rated nor specific expertise in dysarthric speech resulting from PD. Hence, the intelligibility ratings in our study took place in a less protocolized context with less-experienced raters, possibly causing more variability in scale ratings. It could be argued that a perceptual scaling technique such as Direct Magnitude of Estimates (DME) would have been more appropriate than 10-point scale rating, since this would provide more rigorous inferential statistical data than is possible using ordinal data. However, DME scaling has serious disadvantages: 1) the results are highly dependent on the chosen standard (Weismer & Laures, 2002) and 2) participants tend to find it difficult to do the mental calculations involved in a DME task (Rietveld & Chen, 2006).

A closer look at both the intelligibility ratings and orthographic transcription scores showed considerable variation in ICC´s for the different syntactic sentence types (i.e. D6, D13 and Q). The ICC for intelligibility scale ratings in D6 sentences was relatively low (i.e. 0.60) compared to the ICC's for intelligibility ratings for D13 (0.89) as well as Q sentences (0.76). An explanation could be found in the possibility that the syntactic structure in D13 and Q sentences are more likely to elicit explicit prosodic expression due to sentence length (i.e. double clauses) and question form respectively. These prosodic cues contribute to language processing (Maier, Haderlein, Schuster, Nkenke, & Noth, 2007) and hence, intelligibility. That is, intelligibility might be improved due to additional prosodic information. As a consequence, they are likely to assign higher scale ratings across D13 and Q sentences. This results in less variation and higher ICCs. Realizations of D6 sentences might lack this additional prosodic information which possibly compensates for poor realization of segmental speech elements. Thus, the syntactic structure of D6 sentences, with short sentence length and declarative form, might be less likely to elicit explicit prosodic expression, thus failing to add relevant information for intelligibility perception. Therefore, with less prosodic expression, the actual differences between D6 sentences might have been smaller, leading to more randomlike variation in ratings and hence, lower ICC's.

With respect to orthographic transcription scores, the ICC´s were considered high, with 0.75 for D6 being the lowest value. Similar to intelligibility scale ratings, the less-explicit prosodic expression in D6 sentences might have slightly affected orthographic scores as well.

From the results of Experiment 1 we concluded that, except for the ICC observed for intelligibility ratings in D6 sentences, ICCs for both orthographic transcription scores and intelligibility ratings were sufficient to warrant assignment of the listeners in an orthogonal latin square design, as described in Experiment 2.

5.3 Experiment 2

Experiment 2 aimed at establishing the equivalence of five SUS sets with respect to potential intelligibility for three sentence types (D6, D13 and Q), as measured by both orthographic sentence transcription and intelligibility scale ratings. Five SUS sets, read aloud by five PD speakers with dysarthria, were scored by the same listeners who participated in Experiment 1. The effect of the factor 'SUS set' was tested for both intelligibility measures. In addition, the effects of the factors 'speaker' and 'listener' were evaluated.

5.3.1 Materials and methods

Participants

Speakers.

Five male speakers with PD and hypokinetic dysarthria, including the speaker in Experiment 1, participated in Experiment 2. Similar to the speaker in Experiment 1, the speakers were recruited by SLPs who were trained in Pitch Limiting Voice Treatment (PLVT) for parkinsonian speakers (de Swart et al., 2003). Criteria for including the speakers were assessment of PD by a neurologist and the assessment of dysarthria by the speaker's SLP. Exclusion criteria concerned any circumstances that might complicate the audio recordings during one hour on two subsequent days, such as weak physical condition or severe cognitive problems. The speakers' average age was 66.6 years, ranging from 53 to

71 years. The mean time since diagnosis with PD was 11.8 years, ranging from 7 to 15 years. All speakers took their Parkinson medication two hours prior to the start of the audio recordings. We present speaker demographic data in Table 5.2.

Table 5.2

Demographic data of the speakers (PD 1 to PD 5) with hypokinetic dysartria in Experiment 2

Speaker	Gender	Age (yrs.)	Time since diagnosis (yrs.)
PD1	male	70	12
PD2	male	53	7
PD3	male	69	12
PD4	male	70	15
PD5	male	71	13

Speech intelligibility among the speakers varied according to the clinical impression of two trained listeners (i.e., authors LJB and RPC). This was confirmed by the 10 SLPs who also had participated in experiment 1. They judged intelligibility of the different speakers in a paired comparison listening task. To this end, for each speaker a speech sample was selected. The sample contained a passage from a Dutch standard text that was read aloud. The SLPs were thus provided with 10 pairs of speech samples. For each presented pair of speech samples, the listeners were asked to select the most intelligible speech sample and to rate the extent to which the intelligibility was considered better than in the alternative sample. In case of perceived equal intelligibility, listeners assigned a '0'. In case of unequal intelligibility, the listeners assigned a scale value to the most intelligible speech sample: '1' (slightly more intelligible), '2' (moderately more intelligible) or '3' (considerably more intelligible). Five SLPs rated 10 items of paired speech samples in AB order, and the other five SLPs performed the same task in both reversed item order and reversed pair order (BA). The listening task was computerized using the E-prime software tool (Schneider et al., 2007). Applying Scheffé's method for paired comparisons (Scheffé, 1952) with dedicated software developed by author ACMR[1], we found a significant main effect for 'speaker' at

1 Executable and manual in English is available on request.

the 1% significance level ($F_{4,80}$ = 89.82). The speakers were ranked from highest to lowest perceived intelligibility based on scale ratings in the following order: PD2, PD4, PD5, PD3 and PD1. Hence, speaker 2 (PD2) was considered most intelligible and speaker 1 (PD1) was perceived as least intelligible. Adopting a significance level of 1%, differences between ratings for all speakers, except for speakers 4 and 5, were statistically significant. No order effects were found at the 1% significance level ($F_{10,80}$ = 2.45).

Listeners.
The same SLPs (n=10) from Experiment 1 participated in Experiment 2.

Materials
Five additional SUS sets (SUS1 to SUS5) were constructed. Similar to the SUS0 set, each set consisted of 30 sentences, containing 10 D6 sentences, 10 D13 sentences and 10 Q sentences. Thus, 150 experimental sentences were created. For an overview of the sentences, see http://www.ostt.eu/SUS-Appendix.aspx. Conditions with respect to selected word classes, initial phonemes and the absence of semantically related key words within sentences, were similar to those as described for the SUS0 set. Minor adaptations were required, since a considerable larger number of words were needed to ensure word category balances in and across the SUS sets. Hence, in order to obtain a larger amount of words, criteria for CELEX word frequencies were lowered for both nouns and verbs. In addition, given the lack of available monosyllabic words in Dutch required to meet the same division of word initial phonemes as in the SUS0 set, this division had to be slightly adapted. That is, the percentage of nasal initial nouns and verbs decreased to 7% (opposed to 10% in the SUS0 set), while the number of approximates increased to 10% (opposed to 7 % in the SUS0 set). Furthermore, to avoid insertion of words with too low frequency of occurrence, the percentage of double consonant structure words decreased to 50% (opposed to 54% in the SUS0 set) whereas the triple consonant category exceeded the 3% criterion utilized in the SUS0 set. Both across and within SUS sets, preposition use was balanced where possible. In general, to avoid increased sentence predictability, words were not repeated across syntactic sentence type within the same SUS set.

Procedure

Audio recordings.
Each speaker was asked to read aloud the five sets of 30 unpredictable sentences. Each set contained 10 D6 sentences, 10 D13 sentences and 10 Q sentences. In total, each speaker read aloud 150 SUS sentences. Recording sessions were spread over two 1-hour sessions between 10.00 and 11.00 AM on two separate days. During audio recording, blocks of 8-9 written sentences were presented to be read aloud. After three blocks, the syntactic type of sentence alternated. The speakers were encouraged to rest as needed. We used the same recording settings and post-recording signal manipulation as for Experiment 1.

Sentence presentation order.
Similar to Experiment 1, sentences were presented to the listeners in pseudo random order so that sentence type (D6, D13 and Q) was balanced across presentation blocks and the same sentence type was not presented more than two times in a row.

Listening task.
All listeners were provided with a hard copy of the task instructions. Similar to Experiment 1, they were instructed to orthographically transcribe each sentence and also rate the extent of perceived intelligibility on a 10-point scale. Sentences were aurally provided via a Sennheiser PC 131 headset. Each session began with five practice items: one audio recording per speaker. The practice items were sentences from the initial SUS set developed in Experiment 1 and were not used in any subsequent analysis. The 150 experimental sentences were provided across three sessions on the same day. There was a compulsory 15 minute break between the sessions. Each session contained 50 sentences, each divided into two blocks of 17 sentences and one block of 16 sentences. Similar to Experiment 1, sentences were presented using E-prime, version 2.0, allowing self-paced completion of the experiment.

Scoring.
The first and the second author of this article scored the orthographic transcriptions. They followed the same scoring decision procedure as in Experiment 1. During the scoring procedure, one unsuitable Q sentence was detected: 'Zakt een maart op een straat?' ('Does a March sink on a street?'). That is, the noun 'maart' (i.e. 'March') cannot

be preceded by an article. This sentence was therefore excluded, resulting in 1490 (10 x 149) transcribed sentences to be scored. Scores for the orthographic transcription task and the intelligibility rating task were expressed in percentages.

To verify interrater scoring agreement, an independent female SLP, who was not involved in Experiment 1 or 2, re-scored 20% of all transcribed sentences that did not match the orthographic presentation of the target sentence. This resulted in 158 sentences, which were presented in random order to the independent SLP. The SLP scored the transcriptions according to the keyword decision diagram as depicted in Figure 2. Scoring decision of only four keywords in 158 sentences (i.e. 158´4 = 632 key words) differed between the independent SLP and the scorers of Experiment 2. This is a difference of only 0.6% of the keywords.

5.3.2 Experimental design and data analyses

We employed an orthogonal latin square design to test the effects of the factors 'SUS set', 'speaker' and 'listener' on both orthographic transcription scores and intelligibility scale ratings. In Table 5.3, we present the orthogonal array for the latin square design. Although the low ICC for intelligibility scale ratings in D6 sentences in Experiment 1 did not allow the assignment of listeners to different SUS sets in an orthogonal array, we did not exclude D6 sentences from scale ratings to avoid confusion in the scoring task. Hence, each listener completed both scoring tasks for all sentence types in five different SUS sets, each of which was produced by a different speaker (PD1 to PD5). However, intelligibility scale ratings in D6 sentences were not included in the data analyses.

The mean percentage of correctly scored key words and intelligibility scale ratings was calculated for each sentence type per SUS set. In addition, the mean score percentage for each speaker was established. This was done for each sentence type separately and across sentence types. Finally, for the analysis of latin squares, we carried out separate ANOVAs (on the basis of SPSS syntax) for each syntactic sentence type. Thus, the effects of the factors 'SUS set', 'speaker' and 'listener' on orthographic transcription scores and intelligibility scale ratings were tested. A significance level of 5% was adopted.

Table 5.3

Orthogonal Array Representation in the Latin Square Design Employed in Experiment 2

	SUS set				
Speaker	**SUS1**	**SUS2**	**SUS3**	**SUS4**	**SUS5**
PD1	L1/L6	L2/L7	L3/L8	L4/L9	L5/L10
PD2	L5/L10	L1/L6	L2/L7	L3/L8	L4/L9
PD3	L4/L9	L5/L10	L1/L6	L2/L7	L3/L8
PD4	L3/L8	L4/L9	L5/L10	L1/L6	L2/L7
PD5	L2/L7	L3/L8	L4/L9	L5/L10	L1/L6

Note. Each pair of listeners (L) was assigned five different SUS sets, each of which was produced by a different speaker with PD (PD1 to PD5).

5.3.3 Results

Descriptive statistics for orthographic transcriptions and scale ratings are illustrated in the tables 5.4, 5.5 and 5.6. Intelligibility scale ratings for D6 sentences were not included in the tables 5.5 and 5.6 since low interrater reliability did not warrant an orthogonal latin square array in Experiment 2.

Table 5.4:

Mean orthographic transcription scores

Sentence Type	SUS set					
	SUS1	**SUS2**	**SUS3**	**SUS4**	**SUS5**	**Across all SUS sets**
D6	86.8	84.8	85.0	87.8	83.3	85.5
D13	84.3	73.3	82.0	83.3	79.8	80.5
Q	86.5	83.3	77.3	74.7	80.8	80.5

Notes. Scores represent the mean percentage of correctly transcribed keywords per sentence type for each SUS set ((SUS1 to SUS5) and mean across all SUS sets. D6 = 6-word declarative sentences, D13 = 13-word declarative sentences, Q = question sentences.

Table 5.5:

Mean intelligibility scale ratings

Sentence Type	SUS set					
	SUS1	SUS2	SUS3	SUS4	SUS5	Across all SUS sets
D6	-	-	-	-	-	-
D13	60.2	53.9	56.2	57.5	58.5	57.3
Q	60.9	60.0	58.1	54.9	56.9	58.2

Notes. Scores represent the mean intelligibility scale ratings (percentages) per sentence type for each SUS set (SUS1 to SUS5) and across all SUS sets. Intelligibility scale ratings for D6 sentences were not included since low interrater reliability in Experiment 1 did not warrant an orthogonal latin square design for this sentence type in Experiment 2. D6 = 6-word declarative sentences; D13 = 13-word declarative sentences, Q = question sentences.

Table 5.6:

Mean orthographic transcription scores (percentages) and mean intelligibility scale ratings (percentages) for each speaker across sentence types and for each sentence type separately.

Speaker	Orthographic transcription scores				Intelligibility scale rating			
	D6	D13	Q	Mean across sentence types	D6	D13	Q	Mean across sentence types
PD1	82.5	53.1	72.8	77.3	-	51.8	54.3	53.1
PD2	89.0	65.6	85.0	87.6	-	67.5	64.2	65.6
PD3	83.8	56.3	77.3	81.5	-	58.8	53.7	56.3
PD4	83.3	52.6	79.2	79.4	-	50.3	54.8	52.6
PD5	89.0	60.8	87.9	86.9	-	57.9	63.7	60.8

Notes. Intelligibility scale ratings for D6 sentences were not included since low interrater reliability in Experiment 1 did not warrant an orthogonal latin square design for this sentence type in Experiment 2. D6 = 6-word declarative sentences, D13 = 13-word declarative sentences, Q = question sentences.

Only for orthographic transcriptions of D6 sentences, ANOVA results showed the ideal combination of requirements for clinical outcome research: no main effect for 'SUS set' was found ($F_{4,12} < 1.0$) whereas 'speaker' affected the scores significantly ($F_{4,12} = 2.77$, $p = .041$ and $\eta_p^2 = .23$). In addition, no main effect for 'listener' ($F_{4,12} = 2.20$, $p = .087$) or interaction between 'listener' and 'speaker' ($F_{16,25} < 1.0$) was observed. For orthographic transcription scores for D13 and Q sentences, we did not observe this combination of a main effect of 'speaker' without significant effects of 'SUS set' or 'listener'. For D13 sentences, 'speaker' ($F_{4,12} = 9.89$, $p = .000$, $\eta_p^2 = .52$) as well as SUS set ($F_{4,12} = 3.71$, $p = .012$, $\eta_p^2 = .29$) significantly influenced the transcription scores, whereas 'listener' did not ($F_{4,12} = 1.96$, $p = .120$). For Q sentences, we found a significant effect of 'speaker' ($F_{4,12} = 9.95$, $p = .000$, $\eta_p^2 = .52$), 'listener' ($F_{4,12} = 4.95$, $p = .003$, $\eta_p^2 = .35$) and 'SUS set' ($F_{4,12} = 5.93$, $p = .001$, $\eta_p^2 = .39$).

For intelligibility scale ratings no main effect of 'SUS set' was found for any sentence type (p not being lower than 0.779). The factor 'listener' significantly affected the scale ratings in D13 and Q sentences (D13: $F_{4,12} = 3.47$, $p = .017$, $\eta_p^2 = .27$; Q: $F_{4,12} = 6.17$, $p = .001$ $\eta_p^2 = .40$). A main effect for 'speaker' was only found for D13 sentences ($F_{4,12} = 3.61$, $p = .014$, $\eta_p^2 = .28$).

5.3.4 Discussion and conclusion

In this study, we investigated the suitability of orthographic transcription and intelligibility scale ratings of different SUS sets for employment in longitudinal research designs for speech training efficacy research, more specifically EST efficacy studies. In fact, this study on suitable intelligibility outcome measures fits into Phase I and Phase II clinical outcome studies (Robey et al., 2004), which precede Phase III efficacy studies in clinical trials. The results of Experiment 2 indicate that only orthographic transcriptions of D6 sentences seem suitable outcome measures for speech intelligibility since 1) no main effect of 'SUS set' on orthographic transcription scores was observed, suggesting equivalence of SUS sets with respect to potential intelligibility, 2) orthographic transcription scores were not affected by the factor 'listener' and 3) we found a statistically significant main effect for 'speaker', implying that our orthographic transcription method for SUS (D6) is sensitive to various degrees of intelligibility. This observed effect of 'speaker' on orthographic transcription scores is in line with the different degrees of speech intelligibility between the five dysarthric speakers with PD, as perceived by the 10 SLPs in the pairwise comparisons task. Although the rank order of speakers' intelligibility based on orthographic scores of

D6 sentences (Table 5.6) did not correspond entirely with the rank order based on the results of the pairwise comparisons task for the spoken standard text, both manners of intelligibility assessment indicated speaker 2 as most intelligible and speaker 1 as least intelligible. Intelligibility scale ratings seemed not suitable for efficacy research, since the factor 'listener' significantly affected this intelligibility measure in all sentence types. This finding supports the assumption that subjective intelligibility estimates may be less reliable than orthographic transcription (Hustad, 2006), possibly caused by the lack of 'fixed anchors' for intelligibility rating in listeners (Beijer et al. 2010a).

A limitation of the current study lies in the restricted neurological population that participated, since only speakers with PD were involved. Therefore, generalization of the suitability of orthographic SUS (D6) transcription to dysarthric speech following other neurological causes (e.g. stroke or traumatic brain injury) is not warranted. That is, distinct underlying neurological diseases tend to affect speech in various ways (Darley, Aronson, & Brown, 1969b; Darley, Aronson, & Brown, 1969a), possibly affecting the intelligibility of semantically unpredictable sentences in a different manner. Further research will have to point out whether orthographic transcription of SUS sets as intelligibility outcome measures is warranted for dysarthric speakers with other neurological diseases.

It could be argued that another limitation of our study concerns the lack of an external intelligibility measure. An external measure might have confirmed the intelligibility differences between our dysarthric speakers, as established by the initial perceptual judgements of the two trained listeners and by the outcomes of pairwise comparisons by the SLPs. From the perspective of the goal of our study however (i.e., establishing the suitability of SUS intelligibility scoring for the purpose of EST efficacy research), it is vital to identify intelligibility differences over time that may be only mild. Relative intelligibility scorings of different dysarthric speakers, as in the pairwise comparison task, might therefore be more sensitive to mild differences than formal speech intelligibility measures. Nevertheless, to allow comparisons of speech intelligibility in different studies on dysarthric speakers with PD, external intelligibility measures should be provided in future studies. This also goes for information on level of PD.

Using SUS as speech materials for intelligibility measures has some disadvantages. Obviously, these sentences do not represent everyday speech in some respects. Apart from this, these sentences may be difficult to read aloud by dysarthric speakers and could induce disfluency since they are not familiar with SUS. However, SUS elicit intelligibility scores rather than comprehensibility scores, and have suprasegmental features similar to everyday speech at sentence level. These are key benefits in intelligibility outcome research. Nevertheless, to obtain intelligibility measures based on speech materials with adequate ecological validity, SUS intelligibility scoring should be supplemented with other formal speech measurements based on speech of different linguistic levels.

A final limitation of our study concerns the fact that our results appear only to apply to the Dutch language. Nevertheless, this study gives rise to considerations addressing cross-language comparison of speech intelligibility. That is, our study provides a sound methodological basis for both cross-language generation of semantically unpredictable sentences and the development of a detailed scoring paradigm for orthographic transcriptions. The scoring paradigm employed in our study could provide a basic principle for cross-language guidelines in intelligibility assessment, enabling comparisons between different languages in clinical outcome research. Cross-language employment of a more sophisticated 'SUS scoring paradigm' might even induce a theoretical challenge. That is, similar to procedures and materials in speech and language tests which require 'localization' (i.e. adaptation to speakers' language and culture), also scoring paradigms should be adjusted to the speaker group at issue. The theoretical challenge lays in the definition of phonetic processes that are responsible for differences in scoring, due to effects of different languages. For instance, the occurrence of certain phonetic processes in one language, might be considered deviant in another language (Rietveld & van Heuven, 2009). An example of a normal phonetic process in British English is the intervocalic glottalization in the realization of the word 'better': [bE?@r] (SAMPA notation) (Wells, 1997). Another example is the voicing of stops ('better' pronounced as 'bedder') in American English: [bEd@r] (SAMPA notation). Also the elision of an ambisyllabic /t/ after after /n/ in a strong syllable ('winter' pronounced as 'winner') ([wIn@r]) (SAMPA notation) is considered a normal phonetic process in American English. These normal phonetic phenomenons in British and American English respectively, are characterized as deviant in Dutch however. Obviously, language-specific phonetic processes account for decisions on orthographic transcription scoring. Accurate definition of these phonetic processes contributes to a more detailed scoring paradigm that allows cross-language comparison of speech intelligibility in dysarthric speakers. Thus enlarging the amount of evidence for therapy effects, cross-language intelligibility studies are vital for clinical outcome research in the long term.

Although the results of our study suggest that orthographic transcription of SUS may be a suitable intelligibility measure for EST efficacy research, it remains a relatively time-consuming method. Therefore, as already mentioned in the introduction of this paper, it is tempting to develop automatic pronunciation assessment at sentence level. To this end, ASR engines should be trained with dysarthric speech to achieve considerable confidence measures of pronunciation. Moreover, once acceptable confidence measures by ASR are obtained, correlations between automatic pronunciation scores and subjective intelligibility scoring should be established. That is, although De Bodt and his colleagues report the major role of articulation in intelligibility perception of dysarthric speech (De Bodt, Hernandez-Diaz Huici, & Van De Heyning, 2002), other speech dimensions such as speech rate and intonation are also reported to contribute to intelligibility in dysarthric speech (Blanchet & Snyder, 2011; Tamplin, 2011). This implies that pronunciation scores alone do not account entirely for perceived speech intelligibility. Nevertheless, insight into correlations between pronunciation scores and perceived intelligibility might shed light on the potentials of automatic pronunciation assessment as an intelligibility measurement method in dysarthric speech. In addition to the benefits of accurate and objective intelligibility assessment, this would save time, energy and costs currently involved with subjective intelligibility scoring.

For now, as long as automatic pronunciation assessment at sentence level is not available, intelligibility assessment by orthographic transcription of our D6 SUS sets seems an appropriate contribution to intelligibility outcome measures in EST efficacy research for dysarthric speakers with PD. Future research will have to point out whether SUS transcription scores correlate with other intelligibility measures and with level of PD.

Acknowledgements

The authors wish to thank the speakers in this study for their willingness to spend time and energy reading all the speech materials. In addition, the authors want to thank Joop Kerkhoff (Radboud University Nijmegen) for his technical assistance and the SLPs from the Sint Maartenskliniek in Nijmegen for their participation in this study.

References

Ackermann, H. & Ziegler, W. (1991). Articulatory deficits in parkonsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry, 54,* 1093-1098.

Baayen, R., Piepenbrok, R., & Gulikers, L. (1995). The CELEX Lexical Database Consortium (CD-ROM). Linguistic Data. 2nd release.

Barreto, S. d. S. & Ortiz, K. Z. (2008). Intelligibility measurements in speech disorders: a critical review of the literature. *Pro-Fono Revista de Actualização Cientifica, 20,* 201-206.

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010a). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) as a web application for speech training in dysarthric patients with Parkinson's Disease: a case study. *Telemedicine and e-Health,* 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M. et al. (2010b). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16,* 177-180.

Benoît, C., Grice, M., & Hazan, V. (1996). The SUS test: a method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences. *Speech Communication, 18,* 381-392.

Beukelman, D. R., Mathy, P., & Yorkston, K. (1998). Oucome measurement in motor speech disorders. In C.M.Frattali (Ed.), *Measuring outcomes in speech-language pathology* (pp. 334-353). New York: Thieme.

Blanchet, P. G. & Snyder, G. J. (2011). Speech rate treatments for individuals with dysarthria: a tutorial. *Perceptual and Motor Skills, 110,* 965-982.

Boersma, P. & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.25) [Computer software].

Brennan, D., Georgeadis, A., & Baron, C. (2002). Telerehabilitation tools for the provision of remote speech-language treatment. *Top Stroke Rehabilitation, 8,* 71-78.

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. (2007). Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50,* 1481-1495.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12,* 462-496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246-269.

De Bodt, M. S., Hernandez-Diaz Huici, M. E., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders, 35,* 283-292.

Doremalen van, J., Cucchiarini, C., & Strik, H. (2009). Automatic detection of vowel pronunciation errors using multiple information sources. In *ASRU, IEEE* (pp. 580-585).

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81,* 162-173.

Freeware Software Inc. (2008). Audacity (Version 1.2.6) [Computer software].

Huinck, W. & Rietveld, T. (2007). The validity of a simple outcome measure to assess stuttering therapy. *Speech*

Communication, 59, 91-99.

Hustad, K. C. (2007). Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatrica et Logopaedica, 59,* 306-317.

Hustad, K. C. (2006). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica, 58,* 217-228.

Jekosch, U. (2005). *Voice and Speech Quality Perception. Assessment and Evaluation.* Berlin Heidelberg: Springer Verlag.

Kempler, D. & van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: a case study of Parkinson's Disease. *Brain and Language, 80,* 449-464.

Kennedy, M. R. T., Strand, E. A., & Yorkston, K. M. (1994). Selected acoustic changes in the verbal repairs of dysarthric speakers. *American Journal of Speech-Language Pathology, 2,* 263-279.

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: evidence from event-related potentials. *Psychological Science, 18,* 135-143.

Luyckx, K, Kloots, H., Coussé, E., & Gillis, S. (2007). Klankfrequenties in het Nederlands. Tussen taal, spelling en onderwijs : essays bij het emeritaat van Frans Daems (pp. 141–154). Academia Press.

Maier, A., Haderlein, T., Schuster, M., Nkenke, E., & Noth, E. (2007). Intelligibility is more than a single word: Quantification of speech intelligibility by ASR and prosody. In V.Matousek & P. Mautner (Eds.), *Text, Speech and Dialogue* (pp. 278-285). Berlin Heidelberg: Springer-Verlag.

Mashima, P. A. & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health, 14,* 1101-1117.

Middag, C., Martens, J. P., Van Nuffelen, G., & De Bodt, M. (2009). Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Advances in Signal Processing.*

Middag, C., Van Nuffelen, G., Martens, J. P., & De Bodt, M. (2008). Objective intelligibility assessment of pathological speakers. In *Proceedings of Interspeech 2008, Brisbane Australia* (pp. 1745-1748).

Miller, G. A. (1956). The magic number plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Mitterer, H. & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34,* 73-103.

Mitterer, H. & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. Journal of Experimental Psychology: *Human Perception and Performance, 35,* 244-263.

Müller, R. & Buttner, P. (1994). A critical discussion of intraclass coefficients. *Statistics in Medicine, 13,* 2465-2476.

Office of Technology Assessment (1978). *Assessing the efficacy and safety of medical technologies (OTA-H-75)* Washington, DC: US Government Printing Office.

Pinto, S., Ozsancak, C., Tripoliti, E., Thobois, S., Limousin-Dowsey, P., & Auzou, P. (2004). Treatments for dysarthria in Parkinson's Disease. *Lancet Neurology, 3,* 547-556.

Pluymaekers, M., Ernestus, M., & Baayen, H. R. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 118,* 2561-2569.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow up. *Journal of Neurology, Neurosurgery and Psychiatry*, 493-498.

Rietveld, T. & Chen, A. (2006). How to obtain and process perceptual judgements of intonational meaning. In S.Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schliesser (Eds.), *Methods in Empirical Prosody Research* (pp. 283-319). Berlin: Mouton De Gruyter.

Rietveld, T. & van Heuven, V. (2009). *Algemene fonetiek*. Bussum: Coutinho, in Dutch (third edition).

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37*, 401-411.

Robey, R. R. & Schultz, M. C. (1998). A model for conducting clinical-outcome research: an adaptation of the standard protocol for use in aphasiology. *Aphasiology, 12*, 787-810.

Sanders, E., Ruiter, M., Beijer, L., & Strik, H. (2002). Automatic recognition of Dutch dysarthric speech:a pilot study. In (pp. 661-664).

Scheffé, H. (1952). An analysis of variance for paired comparisons. *Journal of the American Statistical Association, 47*, 381-400.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). E-Prime (Version 2.0) [Computer software].

Strand, E. & Yorkston, K. M. (1994). Description and classification of individuals with dysarthria: a 10-year review. In J.A.Till, K. M. Yorkston, & D. R. Beukelman (Eds.), *Motor speech disorders: Advances in assessment and treatment*. (pp. 37-56). Baltimore: Paul H. Brookes.

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60*, 498-500.

Tamplin, J. (2011). A pilot study into the effect of vocal exercises and singing on dysarthric speech. *NeuroRehabilitation, 23*, 207-216.

Theodoros, D., Hill, A., Russell, T., Ward, E., & Wootton, R. (2008). Assessing acquired language disorders via the Internet. *Telemedicine and e-Health, 14*, 552-559.

Weismer, G. & Laures, J. S. (2002). Direct Magnitude Estimates of Speech Intelligibility in Dysarthria: Effects of a Chosen Standard. *Journal of Speech, Language , and Hearing Research, 45*, 421-433.

Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D.Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter.

Wertz, R. T. & Irwin, W. H. (2001). Darley and the efficacy of language rehabilitation in aphasia. *Aphasiology, 15*, 231-247.

Witt, S. M. & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication, 30*, 95-108.

Yorkston, K. M. & Beukelman, D. R. (1981). *Assessment of intelligibility in dysarthric speech*. Tigard: CC Publications.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*, 550-560.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of dysarthric speech: implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*, 55-66.

Ziegler, W. & Zierdt, A. (2008). Telediagnostic assessment of intelligibility in dysarthria: A pilot investigation of MVP-online. *Journal of Communication Disorders*, 553-577.

Zraick, R. I., Davenport, D. J., Tabbal, S. D., Hutton, T. J., Hicks, G. M., & Patterson, J. H. (2004). Reliability of speech intelligibility ratings using the Unified Huntington Disease Rating Scale. *Journal of Medical Speech-Language Pathology, 12*, 31-40.

Zraick, R. I., Dennie, T. M., Tabbal, S. D., Hutton, T. J., Hicks, G. S., & O'Sullivan, P. (2003). Reliability of speech intelligibility ratings using the Unified Parkinson Disease Rating Scale. *Journal of Medical Speech-Language Pathology, 11*, 227-240.

PART 3

PROOF OF PRINCIPAL: STUDIES INTO THE FEASIBILITY OF EST

CHAPTER 6

Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) for dysarthric patients with Parkinson's disease: a case-study

L.J. Beijer, A.C.M. Rietveld, V. Hoskam, A,C.H. Geurts, B.J.M. de Swart

Adapted from: *Telemedicine and e-Health,* 2010, 16, 732-738

Abstract

Objective: As a web application for speech training, E-learning based Speech Therapy (EST) is assumed to have potentials for neurological patients who aim at independent speech training in their home environment. This paper reports a case study of a patient with dysarthric speech due to Parkinson's Disease (PD), who enrolled in a 4-week intensive speech training through EST. The primary goal was to investigate the feasibility and the potential efficacy of EST as a web application for speech training in dysarthric patients with PD.

Materials and methods: The participant in our study used EST, following a speech training program containing parts of the Pitch Limiting Voice Treatment (PLVT) for patients with PD. The feasibility of EST for independent speech training in the home environment was verified through a questionnaire. The questionnaire addressed the participant´s individual experiences with EST as well as the extent of satisfaction with technological features of EST, the content of the speech training and the suitability of the home training environment. The potential efficacy of EST as a device to improve speech intelligibility was investigated using a repeated measures with randomized blocks design. The proportion correct orthographic transcriptions of semantically inpredictable sentences (SUS) as well as ratings of perceived intelligibility on a 10-point scale were used as measures for speech intelligibility.

Results: Outcomes of the questionnaire resulted in recommendations to enhance EST feasibility. Speech intelligibility, as measured by transcription scores, improved significantly after EST training. This improvement was maintained for two weeks after completing the EST training, whereas considerably lower scores were observed after 11 weeks without training. Subjective ratings of intelligibility did not show significant differences across time.

Conclusion: The results of this case study confirm the potential of EST for patients with PD.

6.1 Introduction

E-learning based Speech Therapy (EST) is a web application for speech training (i.e., practice), which is a specific aspect of speech therapy. EST was primarily developed for patients with dysarthric speech following acquired neurological disease such as stroke or Parkinson's Disease (PD) (Beijer et al., 2010). Hence, EST is an example of the increasing number of telehealth applications in speech and language pathology (Mashima & Doarn, 2008). A vital potential of EST is the possibility to improve speech quality in the absence of a therapist, in a place and at a time of patients' own choice. That is, EST provides the opportunity of speech training without spending time, energy and money travelling to a therapist, thus enhancing access to speech training for neurological patients who tend to be easily fatigued and less mobile due to their physical condition. This independency and easy access to speech training is expected to address the need for intensive speech training in patients with longstanding dysarthria following acquired neurological disorders (Palmer, Enderby, & Hawley, 2007). According to our clinical experience, these patients tend to suffer severely from their loss of speech quality, and are highly motivated for intensive speech training. The crucial value of intensive training for neurological patients has been confirmed by research outcomes in the field of motor rehabilitation after stroke (Kwakkel, Wagenaar, Twisk, Lankhorst, & Koetsier, 1999). Also in speech and language rehabilitation for patients with stroke and PD, intensive training has been proven to be more effective than low frequency training (Ramig et al., 2001; Teasell & Kalra, 2004). Moreover, for chronic motor stroke, there are indications that repeated therapy leads to additional improvement of motor activity and even might restore deteriorated quality of motor activity, frequently observed once periods of intensive training have been completed (Rijntjes et al., 2009). Hence, EST might be an excellent adjunct to practice, allowing patients to intensify their speech training. EST provides a suitable manner for patients with chronic dysarthria to maintain or even improve their speech quality without placing high demands on therapy resources (Palmer et al., 2007). This might benefit stroke patients as well as patients with PD who finished face-to-face therapy.

In this report, the feasibility and the potential efficacy of EST, containing the Pitch Limiting Voice Treatment (PLVT) (de Swart, Willemse, Maassen, & Horstink, 2003), is evaluated in a patient with PD. The feasibility to be evaluated in this study addresses the issue whether a patient with PD can deal successfully with the web-based speech training device EST. In addition, this case study fits in Phase I of clinical-outcome research (Robey

& Schultz, 1998). By exploring the application of this web device in clinical practice, this study contributes to insight in the potential efficacy of EST.

EST

The EST infrastructure is visualised in Figure 6.1. The keystone of this web application for speech training is a central server. This server contains audio files of target speech as well as audio files from neurological patients who upload their speech through EST. Therapists as well as patients have access to this server. Therapists are allowed to remotely design a tailor-made speech training program by selecting audio files of target speech from the server. Patients enroll in the individual speech training program in their home environment, guided by instructional text to navigate the system and carry out the exercises. By simply pushing a button they record their speech attempts and upload the audio files to the server. For an elaborate description of the EST system we refer to Beijer et al. (2010).

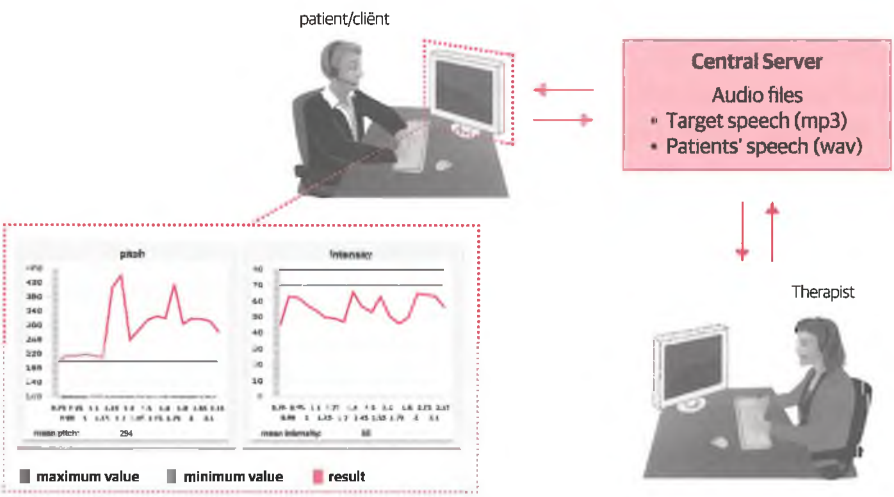


Figure 6.1:

EST infrastructure, enabling speech therapists to remotely design a tailor-made speech training, and allowing patients to remotely engage the speech training. Patients are provided with automatic visual feedback on their speech intensity and overall pitch.

Therapists are provided with a headphone to listen to audio files containing patients' speech and target speech. Patients carrying out the EST training procedure use a headset. As is shown in Figure 6.2, the training procedure for clients contains five basic steps: 1) listening to an audio example of target speech, 2) imitating the target speech, 3) comparing the target with their own speech by means of independent auditory feedback, 4) uploading their speech to the server to obtain additional automatic visual feedback about overall-pitch and intensity, and 5) deciding whether or not an adjusted speech attempt should be made to approach the target speech.

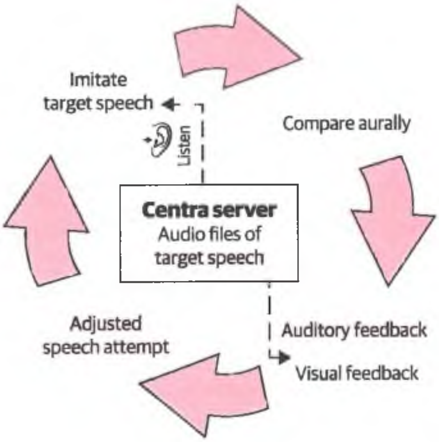


Figure 6.2:
EST training procedure

6.2 Materials and methods

Subject

A 69-year-old male with PD participated in this study. At the time of inclusion, this patient had been diagnosed with PD since eleven years. He was assessed stage 2.5 according to the Hoehn and Yahr severity grading system, which ranges from 0 to 5 (Hoehn & Yahr, 1967). This stage of disease corresponds to symptoms of Parkinsonism on both sides of the body, mild balance problems and physical independence. Since 2003 the participant had been using the medication Levodopa. No relevant comorbidity was observed.

The participant in our study had an academic degree and was retired as an enigineer. He was an active member of the Dutch Parkinson's Disease Association and was actively engaged in various social activities. Four months before our study he completed traditional face-to-face speech therapy, addressing the Pitch Limiting Voice Treatment (PLVT) for patients with PD (de Swart et al., 2003). At the time the participant enrolled in our study, he was experiencing diminished intelligibility in daily conversation.

Speech materials

For the EST speech training, a PLVT training protocol was set up. The PLVT is a Dutch treatment for PD patients which focuses on improvement of intensity (loudness) and at the same time prevents an increase in vocal pitch and, thereby, in laryngeal muscle tone and laryngeal resistance. By instructing patients to speak ´loud and low´, the PLVT aims at improvement of speech intelligibility and prevention of strained or pressed high pitched voicing (de Swart et al., 2003). The EST training program contained utterances varying in linguistic complexity from isolated vowels and monosyllabic words with CVC structure, up to 10-word sentences.

For the evaluation of speech intelligibility, a set of 30 semantically unpredictable sentences (SUS) was used. SUS sentences are characterized by a syntactically correct structure in the absence of meaningful semantic context, and are developed to test word intelligibility in sentences. For the purpose of evaluating word intelligibility, the condition of word production in sentences is ecologically more valid than words in isolation (e.g., in word lists). SUS sentences have the advantage of disrupting predictability due to semantic information (Jekosch, 2005) and are frequently used for the evaluation of synthesized speech (Benoit, Grice, & Hazan, 1996). Examples of SUS sentences are:

The mouth looks at the skin
The house stands on the tree

In our study, we applied Dutch SUS sentences for the evaluation of intelligibility in dysarthric speech. Each SUS sentence contained four key words (i.e., one verb, two nouns and one preposition). For an overview, we refer to http://www.ostt.eu/SUSSentences.aspx

Procedure

Assessing auditory discrimination and screening hearing thresholds

Before enrolling in the EST training, the participant performed an auditory discrimination test (ADT). This test has been recently developed in the Netherlands to assess auditory discrimination in speech in neurological patients with dysarthria, aiming at speech training through EST. The ADT addresses auditory discrimination of variations along five speech dimensions: segmental speech elements (articulation), intensity (loudness), overall pitch, speech rate and intonation (Beijer, Rietveld, & Van Stiphout, in preparation). The results showed a minimum score of 80% in all subtests, implying sufficient auditory discrimination abilities for independent auditory feedback required for the EST training. In addition, an audioscreening was carried out, using an Interacoustics AS 208 screening audiometer. The results showed a Fletcher Index of 52 dB for the right ear and a Fletcher Index of 13 dB for the left ear, indicating a moderate hearing loss for the right ear and a mild hearing loss for the left ear in the area relevant for speech perception.

EST training

During a 4-week EST period, the participant followed a protocolized PLVT speech training program in his home environment, using a Dell laptop, type Latitude D420, with a 12.1" screen, and a Sennheiser PC 131 headset. Based on observed intensities and overall pitch in the participant's speech prior to the EST training period, target values were assigned to these two speech dimensions (i.e., the minimum and the maximum values). The individual targets for our participant were set at mean intensities between 70 and 80 dB and an overall pitch ranging from 105 and 130 Hz. Each week, he carried out four subsequent courses containing PLVT exercises, each course being accessible for one or two days. The participant's access to each course was remotely manipulated by a speech therapist, thus optimizing a protocolized speech training. Over the training period, the participant's uploaded speech was remotely monitored by a speech therapist. The therapist and the participant had contact by telephone or by e-mail for additional instructions, when necessary (i.e., once or twice a week).

Speech recordings of SUS sentences

The participant was asked to read aloud a set of 30 SUS sentences at five times across the experimental period: five weeks before the EST training (T0), immediately before (T1) and immediately after (T2) the EST training, and finally 2 weeks (T3) and 11 weeks (T4) after completing the EST training. To avoid confounding influences of the participant's medication, to be taken daily at regular times, the recording of SUS sentences also took place at regular times (i.e., between 10.00 A.M. and 11.00 AM). The speech was recorded digitally on a Fujitsu Siemens laptop with a sampling frequency of 44.100 Hz, using Audacity, version 1.2.6.

Orthographic transcription and intelligibility ratings

For the evaluation of intelligibility, 20 untrained listeners were asked to orthographically transcribe SUS sentences recorded at different times across the experimental period (T0 to T4). Scores were expressed in the proportion correctly transcribed key words per sentence. In addition, the listeners rated the extent of perceived intelligibility for each SUS sentence on a 10-point scale, which is frequently used in the Dutch school system. The scale ranged from '1'(extremely bad intelligibility) to '10' (extremely good intelligibility). Ratings below ´6´ indicate an insufficient extent of intelligibility.

A repeated measures randomized blocks design was used with 'time' as a within subject-factor (five levels) and four homogeneous blocks of each five listeners. All blocks contained untrained listeners with an academic degree, differing in age and/or gender. This resulted in four homogeneous blocks containing, respectively, five males between 21 and 30 years old, five females between 21 and 30 years old, five males between 51 and 60 years old and five females between 51 and 60 years old. Within each block, each member was assigned randomly to a set of 30 SUS sentences recorded at one of the five assessment times The SUS-sentences were delivered binaurally by means of a Sennheiser PC131 headphone. Each sentence was presented only once. Using the E-prime software tool, version 1.2 (Schneider, Eschman, & Zuccolotto, 2002), the listeners were allowed to carry out the transcription task at their own pace.

Questionnaire

To obtain qualitative information addressing individual experiences using EST, the participant was asked to fill in a questionnaire once the EST training was completed. The questionnaire addressed the participant's experiences with technology related aspects throughout the EST training procedure, as well as the participant´s individual condition in using EST. In addition, he was asked to rate the extent to which he was satisfied with 1) the technological features of EST, 2) the content of the speech training program (PLVT) in EST and 3) the home training environment during the EST training. A 10-point scale, ranging from '1' (extremely unsatisfied) to '10' (extremely satisfied') was applied for these ratings. Ratings below '6' indicate an insufficient extent of satisfaction. Throughout the questionnaire, the participant had the opportunity to report additional personal comments. Finally, the participant was invited to recommend improvements for EST.

6.3 Data analyses

An analyses of variance for repeated measures was carried out on the orthographic transcriptions scores and the scale ratings, the within-subject factor 'time'. A 5% significance level was adopted,

6.4 Results

Orthographic transcriptions

As is shown in Figure 6.3, the orthographic transcription scores were considerably higher immediately after the EST training, whereas a lowering of scores was observed 2 weeks after completing the training period. This lowering of transcription scores, representing diminished intelligibility, proceeds in the course of time without speech training after the EST period.

The repeated measures ANOVA for orthographic transcri[tion scores yielded a significant effect of the within-subject factor 'time': $F_{2.068, 6.204}$=10.471, p=.010, Huynh-Feldt corrected. The effect size (η_p^2) was 0.777. Paired comparisons showed a significant higher transcription score after the EST training (T2), with a mean score of 0.89 (SD=0.045) compared to a mean score of 0.79 (SD=0.029) immediately before the EST training (T1). Paired comparisons between other assessments did not yield statistically significant differences. However, a tendency of lowering transcription scores is observed 2 weeks after the EST training (at T3): a slightly lower mean score occurred, implying slightly diminished intelligibility. This tendency proceeded by further deterioration of intelligibility 11 weeks after completing the EST training (T4). At this point, the mean transcription score approached the scores before the training period, at T0 and T1. Apparently, 11 weeks without training resulted in loss of the achieved improvement during the EST period.

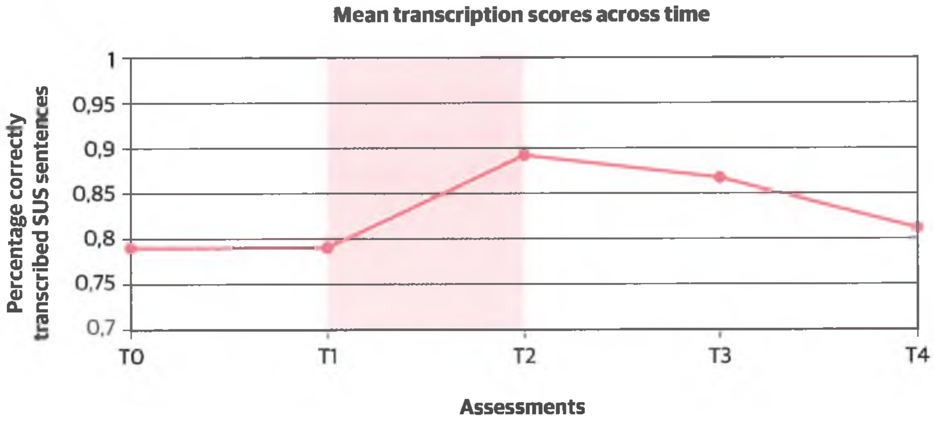


Figure 6.3:

Mean proportion of correctly transcribed SUS sentences across time. The EST training period is marked by a shaded block.

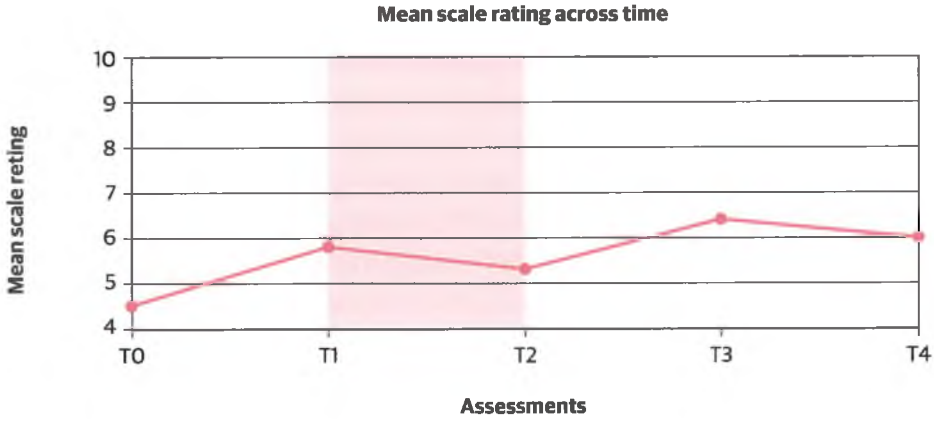


Figure 6.4:

Mean intelligibilityscale ratings across time. 1 = lowest value, 10 = highest value. Ratings below 6 represent insufficient intelligibility. The EST training period is marked by a shaded block.

Intelligibility ratings

Figure 6.4 shows the mean intelligibility ratings on a 10-point scale across time. Obviously, a slightly different pattern than for the transcription scores occurred. Whereas the orthographic transcriptions achieved the highest mean score immediately after the EST training (T2), the intelligibility ratings at this point were slightly lower than before the training. At T2 the mean rating was 5.3 (SD=1.01) and at T1 the mean was 5.7 (SD=0.55). Two weeks after the EST training, the highest mean rating was observed (M=6.4, SD=0.80).

A repeated measures ANOVA on the intelligibility scale ratings did not yield a significant within-subject-factor 'time' ($F_{2.400,7.200}$=3.026, p=.107, Huynh-Feldt corrected, observed power=0.438).

Questionnaire

With respect to technological aspects of the EST training procedure, slight technological problems in the first week were reported, mainly addressing the consequences of disruptions in the internet connection. No problems were mentioned with logging into the patient account, navigating the system, accessing the exercises in the speech training program, carrying out the training procedure or uploading speech to the server. Experiences with EST, related to the participant´s individual physical and cognitive condition, resulted in a recommendation of a larger laptop screen than 12.1" to facilitate reading the instructional text and the training items. With respect to the speech training program, restriction of warming up exercises was recommended to avoid premature fatigue during the 'genuine' speech exercises.

The results of the satisfaction ratings are presented in Table 6.1. Ratings about technological features in EST achieved a mean of 6.67 (2.35) on a 10-point scale. Items addressing the content of the speech training program (PLVT) achieved a mean rate of 7.57 (0.98), and the possibility of a home training environment was qualified with the highest mean rate of 8.33 (1.55).

From the ratings in Table 6.1, it becomes obvious that the size of the laptop screen urgently required improvement. The item 'adjustments of the speech training program' addresses the therapist's possibility to remotely vary the content of the training program. That is, after the first 3 weeks of training, our participant reported considerable fatigue carrying out all exercises per courses in the training program, impeding repeated speech attempts per item. Therefore, the therapist reduced the number of exercises per course in the fourth week, resulting in a considerably higher satisfaction rating on the item addressing the number of exercises per course. The items related to the home training environment were highly appreciated on the rating scale.

Additional to the questionnaire and the satisfaction ratings on a 10-point scale in Table 6.1, the participant was allowed to comment on his personal experiences with EST. Although he reported to find EST an attractive device for independent speech training in his home environment, he recommended additional contact once a week with a therapist by telephone or through videoconferencing. The participant also advised improvement of the user interface in order to enhance EST attractiveness for users. Furthermore, although the home training environment was highly appreciated, ergonomic requirements for using EST should be taken into account. In addition, the keyboard and the mouse should be adapted in order to meet the requirements of patients with PD, who tend to suffer from diminished motor coordination. Finally, although satisfied with the graphs for visual feedback on speech intensity and overall pitch for his personal use, our participant recommended adjustment of the visual feedback to avoid problems in patients less experienced in the interpretation of visual graphs.

Table 6.1:

Ratings of the extent of satisfaction addressing technological features of EST, the content of the speech training program and the home training environment. 1 = lowest value, 10 = highest value. Ratings below 6 represent insufficient satisfaction.

Extent of satisfaction		Rating (10-point scale)
Technological features		
1	Ease of using EST	7
2	Training procedure (manner of exercising)	8
3	Attractiveness of EST as a speech training device	7
4	Size of laptop screen	1
5	Clarity of instructions to navigate in EST	6
6	Auditory feedback	8
7	Visual feedback	9
8	Suitability of feedback as a cue to improve speech	8
9	Robustness to technical disturbances	6
Speech training program		
10	Instructional text throughout EST training	7
11	Type of speech exercises	8
12	Degree of difficulty in speech exercises	8
13	Design of the speech training program throughout the EST training period	8
14	Number of exercises per course (week 1,2,3)	6
15	Number of exercises per course (week 4)	9
16	Adjustment of the speech training program	7
Training environment		
17	Opportunity of training in home environment	9
18	Comfortability of the individual home training environment	7
19	Possibility to train independently	9

6.4 Discussion

The results of our study show that speech practice through EST has the potential to improve speech intelligibility. Apparently, the web application EST facilitates easy access to speech training, thus enhancing the opportunity to intensify training. The observed improvement of speech quality through intensive web-based training observed in our case study underscores the value of practice. This is in agreement with research findings in the field of motor rehabilitation as well as speech and language rehabilitation after stroke, showing that intensive training is effective and improves therapy results (Kwakkel et al., 1999; Rijntjes et al., 2009). As indicated by our case study, the benefit of intensive training may also apply to patients with chronic dysarthria due to PD. Indeed, the results of EST showed statistically significant higher transcription scores immediately after completing the EST training. Two weeks after completing EST, a slight decay of the achieved improvement occurred, and even dramatically lower transcription scores were observed on the long term (i.e. eleven weeks after the EST training period). The decay of skills over time underscores the transient nature of the improvement. Therefore, EST might not only improve speech intelligibility, but also help to maintain the achieved speech quality in patients with chronic dysarthria (Palmer et al., 2007).

In contrast to the orthographic transcription scores, the ratings on perceived intelligibility on a 10-point scale did not show statistically significant effects across time. With somewhat lower ratings immediately after completing the EST training, the pattern of mean ratings of perceived intelligibility throughout the experimental period actually differed from that of the orthographic transcriptions. It might be argued that, for these subjective ratings, in the absence of anchor stimuli, the interrater agreement tends to be lower due to raters' personal 'anchors' for valuation of speech intelligibility (Rietveld, in preparation). Since we used a randomized blocks design with repeated measures, the distinct pattern of perceived intelligibility ratings across the experimental period might be caused by such differences in personal 'anchors'. The observed difference between orthographic transcription scores and the intelligibility ratings, might imply that subjective rating is not a suitable measure to evaluate speech intelligibility.

In the context of patient-centered health care, qualitative information addressing the patient's experiences with EST was obtained by means of a questionnaire. The participant in our study was able to operate the EST device successfully. He highly appreciated the possibility of independent speech training in his home environment, although additional contact with a speech therapist by telephone or videoconferencing was preferred. Yet, various adaptations of the EST equipment might be required to enhance suitability for neurologically impaired patients in general. These findings are in line with previously reported advantages and disadvantages of telerehabilitation (Torsney, 2003).

Applying recommendations of the American Telemedicine Association (Krupinsky et al., 2006), future research should address technical, clinical, human and ergonomic factors of EST. In addition, economic analyses should be conducted to establish the cost-effectiveness of this web application for speech training. In the near future, we will investigate user preferences for traditional face-to-face speech training versus web-based training in patients with acquired dysarthria. Since this patient group is heterogeneous with respect to background variables such as gender, age, underlying disease, severity of dysarthria, computer experience and physical condition, it is vital to detect background profiles of potential EST users. This might avoid disappointing experiences or even premature abandonment of this web application for speech training.

For now, the results of this case study confirm the potential efficacy of EST in patients with PD. Our findings contribute to establishing sound fundamentals for research into the efficacy of EST to improve and maintain speech quality in larger groups of patients (Robey & Schultz, 1998).

Acknowledgements

The authors want to thank Bert van Esterik for participating in this case study, Marijn van Beers for his technical assistance during the experimental period and Margo van Stiphout for her assistance with the figures in this article.

References

Beijer, L.J., Rietveld, A.C.M., van Beers, M.M.A., Slangen, R.M.L., van den Heuvel, H., de Swart, B.J.M., & Geurts, A.C.H. (2010). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16* (2), 177-180.

Beijer, L.J., Rietveld, A.C.M.., & van Stiphout, A.J.L. Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an auditory discrimination test (ADT) for adult dysarthric speakers. In preparation.

Benoit, C., Grice, M., & Hazan, V. (1996). The SUS test: a method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication, 18,* 381-392.

Hoehn, M. & Yahr, M. (1967). Parkinsonism: onset, progression and mortality. *Neurology, 17,* 427-442.

Jekosch, U. (2005). *Voice and speech quality perception. Assessment and evaluation,* pp. 117 In: Berlin Heidelberg: Springer-Verlag.

Krupinsky, E., Dimmick, S., Grigsby, J., Mogel, G., Pkin, D., Speedy, S. et al. (2006). Policy: Research Recommendations for the American Telemedicine Association. *Telemedicine and e-Health, 12,* 579-589.

Kwakkel, G., Wagenaar, R. C., Twisk, J. W. R., Lankhorst, G. J., & Koetsier, J. C. (1999). Intensity of leg and arm training after primary middle-cerebral-artery-stroke: a randomised trial. *Lancet, 354,* 191-196.

Mashima PS & Doarn CR (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health, 14,* 1101-1117.

Palmer, R., Enderby, P., & Hawley, M. (2007). Addressing the needs of speakers with longstanding dysarthria: computerized and traditional therapy compared. *International Journal of Language and Communication Disorders, 42,* 61-79.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry, 71,* 493-498.

Rietveld, T. (2010). Subjective measuring in speech pathology. In preparation.

Rijntjes, M., Haevernick, K., Barzel, A., van den Bussche, H., Ketels, G., & Weiller, C. (2009). Repeat therapy for chronic motor stroke: a pilot study for feasibility and efficacy. *Neurorehabilitation and Neural Repair, 23,* 275-80.

Robey, R. R. & Schultz, M. C. (1998). A model for conducting clinical-outcome research: an adaptation of the standard protocol for use in aphasiology. *Aphasiology, 12,* 787-810.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime Reference Guide. Pittsburg.

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60,* 498-500.

Teasell, R. W. & Kalra, L. (2004). What's new in stroke rehabilitation? Advances in Stroke 2003. *Stroke, 35,* 383-385.

Torsney, K. (2003). Advantages and disadvantages of telerehabilitation for persons with neurological disabilities. *NeuroRehabilitation, 18,* 183-185.

CHAPTER 7

Exploring the effects of E-learning based Speech Therapy (EST) on speech intelligibility in dysarthric speakers with acquired neurological diseases: an exploratory effect study.

L.J. Beijer, A.C.M. Rietveld, M.B. Ruiter & A.C.H. Geurts

Submitted

Abstract

We explored the suitability of perceptual and acoustic outcome measures for testing the hypothesized potentials of E-learning based Speech Therapy (EST) to improve or maintain speech intelligibility in chronic dysarthric patients. Eight speakers with stroke (n=3), Parkinson's disease (n=4) and traumatic brain injury (n=1) participated in a repeated measures design. They followed a four-weeks period of intensive speech practice through EST. Speech samples were recorded before and after the training period. Perceptual measures for speech intelligibility were a) scale ratings for 'ease of intelligibility' and 'pleasantness' in continuous speech and b) orthographic transcription scores of semantically unpredictable sentences. Acoustic measures were c) 'intensity during closure' (ΔIDC) in the intended silent phase of /p/, /t/ and /k/, d) the movements of F1 and F2 in the articulatory-acoustic vowel space of /a/, /e/ and /o/ and e) 'the variability of the fundamental frequency' in semantically unpredictable sentences. The only consistent finding concerned the significantly increased ΔIDC after EST. This supports the hypothesis that EST affects the realization of speech. The increased instead of the expected decreased ΔIDC may be caused by increased overall speech intensity without articulatory adjustments. This implies the need for additional specific instructions to improve patients' speech quality. For other perceptual and acoustic measures, sensitivity enhancement is required for detection of subtle within-subjects changes in future effect studies.

7.1 Introduction

E-learning based Speech Therapy (EST) is a web-based application for speech training in dysarthric speakers with acquired neurological diseases, such as stroke and Parkinson's disease (PD) (Beijer et al., 2010b). EST provides dysarthric speakers in their home environment with the possibility to access a speech training program which is remotely compiled by their speech therapist. Patients can download audio examples of target speech from a server. The EST training procedure is based on the imitation of these speech audio examples, which is in line with research findings indicating the beneficial effects of using audio examples in speech training (Palmer, Enderby, & Cunningham, 2004; Rvachew, 2004). By comparing the audio example with their own speech the dysarthric speakers have to determine whether their own speech sufficiently approaches the audio example (Beijer et al., 2010b). Since there are indications for diminished auditory speech discrimination skills in neurological patients (Beijer, Rietveld, & van Stiphout, 2011), auditory feedback is supported by visual feedback on two relevant speech dimensions in the context of the Pitch Limiting Voice Treatment (PLVT) for dysarthric patients with Parkinson's disease: intensity (i.e. loudness) and fundamental frequency (i.e. pitch). This augmented visualization of speech is provided through the software package 'PRAAT' (Boersma & Weenink, 2009) once the patients' speech has been uploaded to the server.

EST is believed to have potential in neurorehabilitation for several reasons. First, neurological patients' physical conditions tend to impose considerable limitations on the dysarthric speakers' ability to visit a speech therapist. Second, the possibility to practice their speech in the home environment - independent of the physical presence of a speech therapist - allows intensified training, which is known to be effective in motor (Kwakkel, 2006) and communicative rehabilitation (Bhogal, Teasell, & Speechley, 2003; Basso, 2005). In general, patients' rehabilitation could be optimized through EST since intensified training is possible during both inpatient stay and day treatment (i.e., EST additional to face-to-face training) as well as after discharge from treatment (Winters & Winters, 2004). Third, EST meets the urgent need for innovative ways of health care delivery imposed by the current developments in health care. That is, the demographic shift towards an ageing population goes along with an increased number of neurological patients and a decreasing number of health care providers. This development not only leads to an imbalance between the availability and the need of health care capacity, but also requires a cut down on financial expenses per patient . EST may contribute to cost reduction by diminishing travel costs and personnel costs through a decrease in face-to-face therapy sessions (Hill, 2008).

The above-mentioned potential benefits of EST obviously call for clinical outcome research with regard to the effects on dysarthric speech, users' satisfaction and cost-effectiveness. By now, the technological feasibility of EST has been established along with indications for the potential efficacy in dysarthric speakers (Beijer, Rietveld, Hoskam, Geurts, & de Swart, 2010a). In addition, the results of a pilot study addressing the appreciation of EST indicated that the participating dysarthric speakers tended to prefer web-based over face-to-face training in the chronic phase (Beijer, Rietveld & Geurts, submitted). The EST studies that have been conducted until now primarily fit into Phase I and II of clinical outcome research (Robey, 2004). They address vital prerequisites for successful use of EST, such as the technological feasibility (Beijer et al., 2010a) and the assessment of neurological patients' auditory speech discrimination skills (Beijer et al., 2011). With regard to EST clinical outcome research, previous studies were confined to a first conceptualization of the treatment protocol and to initial investigations of the effect of EST on the intelligibility in dysarthric speakers (Beijer et al., 2010a; Beijer, Clapham, & Rietveld, 2012).

The current study constitutes a transition from a Phase II to a Phase III study. By further exploring and specifying the outcome measures reflecting therapeutic effects on speech intelligibility in chronic dysarthria, it aims to prepare a larger clinical trial in the longer term. In the next sections a pilot study with an experimental group (n=5) and a control group (n=3) of chronic dysarthric speakers is described. It was hypothesized that beneficial effects of EST on the intelligibility in dysarthric speakers could be established by perceptual and acoustic measures. With regard to perceptual measures the hypothesized improved intelligibility should be established by i) higher ratings on general judgements of speech quality, such as 'ease of intelligibility' and 'pleasantness', and ii) higher orthographic transcription scores of semantically unpredictable sentences. With regard to acoustic measures, iii) increased vowel space area (as measured by movements of the first and the second formants (F1 and F2 respectively) in the acoustic vowel space, iv) decreased intensity in the silent interval of voiceless plosives and v) larger standard deviations of the fundamental frequency in SUS should reflect improved speech intelligibility. It should be noticed however that unchanged outcome measures after EST might be interpreted as beneficial as well. That is, speech quality in chronic dysarthria tends to deteriorate without practice, especially in participants with degenerative neurological diseases such as Parkinson's disease (PD) (Jancovic, 2008, Beijer et al. 2010). Web-based speech training through EST may allow chronic dysarthric speakers to maintain their speech quality.

Ethical approval for this study was obtained by the Regional Committee of Medical and Health Research Ethics.

7.2 Materials and methods

Participants

In the experimental group five chronic dysarthric speakers due to stroke (n=2), PD (n=2) and traumatic brain injury (TBI) (n=1) participated. The control group consisted of three dysarthric speakers due to stroke (n=1) or PD (n=2). The participants were recruited by speech pathologists in rehabilitation centres or in health care centres for outpatient care. They were asked to select chronic dysarthric patients who had completed face-to-face speech training more than three months before. Exclusion criteria for participation in the study were aphasia, severe cognitive problems or other disabilities that could hamper speech training through EST. All participants filled in a form that was developed by the authors to collect demographic data and background information associated with their neurological disease (Table 7.1).

Since auditory discrimination of independent auditory feedback plays a vital role in the EST training procedure (Beijer et al., 2010a), the auditory speech discrimination skills in the participating dysarthric speakers were established. To this end, the participants completed an auditory speech discrimination test (ADT), consisting of five subtests. Each subtest addressed an essential dimension in the speech training provided by EST: articulation, intensity, overall pitch, speech rate and intonation. A score below 80% was considered to reflect diminished auditory discrimination on the speech dimension addressed in the subtest (Beijer et al., 2011). In addition to the qualitative hearing assessment, a quantitative pure tone test was performed to assess hearing problems. The outcomes are presented in Table 7.2.

Table 7.1:

Demographic data and background information of the participants in the experimental and the control group. Options for the thria on the participation in daily communication were 'none', 'moderate' and 'large'. Options for the amount of computer skills 'little', and 'hardly'.

Participant	Age (yrs)			Time p.o. (yrs)	Type of dysarthria	Perceived impact on daily communication	Computer skills	Dis to s the
Experimental								
1	61	male	stroke	1	spastic	moderate	considerable	≤
2	75	male	PD	3	hypokinetic	large	considerable	6-1
3	34	male	TBI	14	atactic	large	a lot	≤
4	64	male	PD	10	hypokinetic	large	hardly	≤
5	57	male	stroke	2.5	flaccid	moderate	little	6-1
Control								
6	73	male	PD	6	hypokinetic	moderate	considerable	6-1
7	71	male	PD	20	hypokinetic	large	little	≤
8	63	male	stroke	1	spastic	moderate	little	≤

Notes: PD=Parkinson's disease, TBI= traumatic brain injury

Table 7.2:

Outcomes of the subtests of the Auditory Speech Discrimination Test and the pure tone hearing test for all partic control group.

Participant	Age (yrs)			Time p.o. (yrs)	Auditory Discrimination (score percentages per su			
					Articulation	*Intensity*	*Pitch*	
Experimental								
1	61	male	stroke	1	67*	93	87	
2	75	male	PD	3	80	80	93	
3	34	male	TBI	14	93	100	93	
4	64	male	PD	10	100	93	93	
5	57	male	stroke	2.5	93	93	60*	
Control								
6	73	male	PD	6	73*	100	100	1
7	71	male	PD	20	93	100	93	1
8	63	male	stroke	1	87	67*	87	8

* = *performance below 80%*

Participant 5 had a severe bilateral hearing loss in the higher frequencies (4000 Hz and 8000 Hz) due to trauma, which was expected to cause the poor performance on the subtests ‘pitch’ and ‘intonation’.

Design

A repeated measures design was employed. A between-subject effect concerned the phase in which the groups were provided with EST. In the experimental group ET started after six weeks, whereas the control group followed EST only at the end of the experimental period (Figure 7.1).

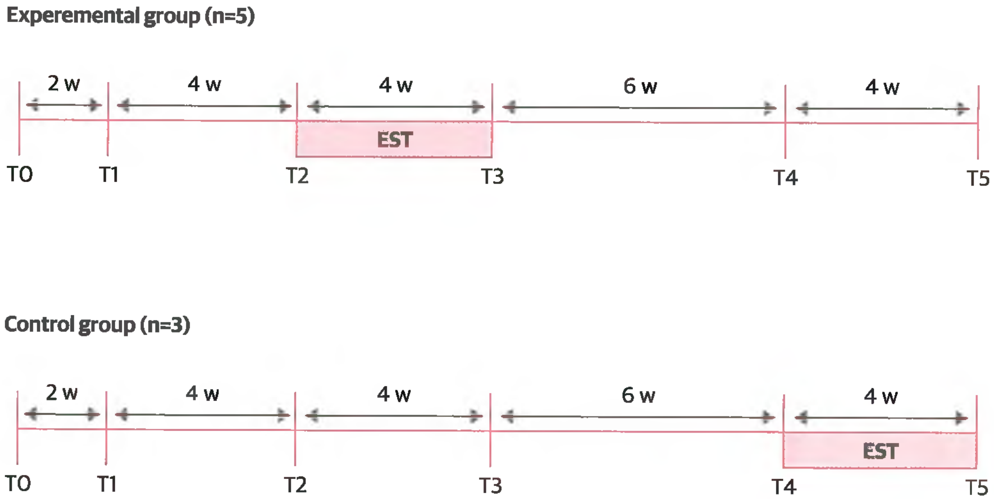


Figure 7.1:

Repeated measures (i.e. assessments TO to T5) in the experimental group (n=5) in the top trace, and in the control group (n=3) in the lower trace.

Intervention

EST

All participants in both the experimental and the control group were provided with a Sennheiser PC 131 headset that was plugged into a laptop (Dell Latitude 5500). On the laptop only those (components of) software packages were installed that were required for EST (Beijer et al., 2010b). The participants were thus provided with a dedicated laptop to avoid EST process disruptions due to potentially interfering software packages on their private laptop or desktop computer.

Both the experimental and the control group followed an intensive four-weeks period of speech training through EST in their home environment. Prior to enrolment in the four-weeks EST period, the participants received an instructional document with screenshots to guide them through a short EST tutorial. In addition, they carried out a short EST trial under supervision of the experimental leader (i.e. author LJB). During the four-weeks period the experimental leader contacted the patients by telephone once a week to evaluate the web-based speech training. In case of problems or questions, the participants were instructed to contact the experimental leader by phone or by e-mail. All participants completed the speech training through EST according to the training protocol.

Content of speech training

The content of the speech training was adjusted to general features of dysarthric speech caused by the underlying neurological disease.

For stroke patients the EST program primarily focused on articulation of plosives /p/, /t/. /k/, /b/, /d/, voiceless fricatives /f/ and /s/ and on the realization of the vowels /a/, /e/, /o/. In addition, the (suprasegmental) speech dimension 'intonation' was addressed since this speech dimension also contributes to the intelligibility in dysarthric speakers (De Bodt, Hernandez-Diaz Huici, & Van De Heyning, 2002). The training was provided in a blocked schedule, in which the patient practices a group of specific target speech movements before practicing the next target (e.g. AAA, BBB,CCC) (Maas et al., 2008). The participant with TBI followed the same speech training program as the stroke patients, since this program addressed the deviant prosodic features (i.e. monotony) as well as the reduced articulatory force and diminished coordination, causing distorted plosives and syllable reduction (Wang, Kent, Duffy, & Thomas, 2005).

The patients with hypokinetic dysarthria due to PD followed the Pitch Limiting

Voice Treatment (PLVT) (de Swart, Willemse, Maassen, & Horstink, 2003). The PLVT is an adapted form of the Lee Silverman Voice Treatment (LSVT) (Ramig et al., 2001) for patients with hypokinetic dyarthric speech due to PD. This treatment focuses on increased intensity (i.e. perceived loudness), but at the same time limits an increase in vocal pitch to prevent strained or pressed voicing. All participating PD patients had completed the PLVT face-to-face sessions more than three months before and, thus, were aware of the key issues in this training program. The EST program aimed at maintaining speech quality in these patients. For an overview of the speech training program see http://www.ostt.eu/Spraaktherapie+door+middel+van+E-learning.aspx

Data collection

Speech recordings

Patients' speech was recorded at three points in time before the EST period (T0, T1, T2) and at three points in time after the EST period (T3, T4, T5) (Figure 7.1). Thus, six assessments were completed. T0 recordings only served as feasibility tests. At T0, the appropriateness of the internet connection was established along with environmental requirements to optimize speech recordings in each individual's home environment. Environmental noise was reduced as much as possible by selecting the most quiet location (e.g. without traffic noise), turning off ticking clocks, washing machines and so forth. The speech recordings were made through the EST system and uploaded to the key server. The participant's speech audio files were thus stored in uncompressed waveform audio file (wav)-format (Beijer et al., 2010b), being available for intelligibility assessments according to the methods described below.

Speech materials

The speech materials used for the assessments, displayed on the laptop screen, were to be read aloud by the participants. The linguistic and phonetic features of the experimental targets to be assessed are listed below.

- **Semantically Unpredictable Sentences (SUS).**
  Five different sets of ten SUS, containing six words in a declarative sentence form were employed. Each sentence contained four key words (i.e. subsequently a noun, a verb, a preposition and a noun) and two determinants (i.e. the indefinite article 'a')

(e.g. 'A job went through a flight'). Although each set contained different sentences, the sets were found to be equivalent with respect to potential intelligibility and also sensitive to different degrees of intelligibility (Beijer et al., 2012).

- **Continuous speech.**
  Continuous speech, elicited by reading aloud two short texts (i.e. an oronasal text 'Papa en Marloes' and a phonetically general text 'Finland'). For each text, the selected speech samples to be judged concerned a passage containing three sentences, ranging from 7 to 12 words.

- **Voiceless plosives /p/, /t/ and /k/ in word initial position and unstressed syllables, embedded in sentences.**
  Voiceless plosives in initial unstressed syllables of multisyllabic words, embedded in sentences (e.g. 'He used to go swimming and SHOP-ping in this town'). We refer to Appendix 7A for examples in Dutch. Unstressed syllables were selected to avoid variable articulatory precision of voiceless plosives due to varying sentence accents (Ackermann & Ziegler, 1991). The sentences were read aloud after an aurally presented example.

- **Voiceless plosives /p/, /t/ and /k/ in word initial position and unstressed syllables, embedded in bisyllabic words.**
  A voiceless plosive in the second, unstressed syllable of a bisyllabic word. The words were read aloud after an aurally presented example.

- **The vowels /e/, /a/ and /o/ in unstressed syllables, embedded in sentences.**
  The correct realization of the unstressed syllable at sentence level was elicited by a question phrase.

  For example:
  Question: *'Who went to Poland?'*
  Response: *'The queen went to Poland'*

  By putting strong emphasis on 'Who' in the question phrase, a response with strong emphasis on 'queen' was elicited, thus inducing an unstressed /o/ in 'Poland'. Both the question phrase and the response phrase with adequate sentence accents were aurally presented to the speaker. The speaker was asked to repeat only the response phrase.

For a complete overview of all speech materials we refer to Appendix 7A.

Outcome measures

Both perceptual and acoustic outcome measures of speech intelligibility were employed.

1) **Perceptual measures**

a) Judgements of speech quality in continuous speech

For each dysarthric speaker, judgements of continuous speech with regard to 'pleasantness' and 'ease of intelligibility' were obtained. Both 'ease of intelligibility and 'pleasantness' were used as measures at the ICF activity dimension, providing an impression of the functional benefits of speech in daily life. The judgements were obtained by asking four trained and four untrained listeners to compare audio speech samples which were recorded at different assessments (T2 to T5). Thus, in total 64 (8X8) listeners were involved. For each participant, four listeners judged the paired speech samples in AB order (i.e. T2-T3, T2-T4, T2-T5, T3-T4, etc.), whereas the other four listeners judged the pairs in BA order (i.e. T3-T2, T4-T2, T5-T2, T4-T3, etc.). Judgements were based on scale ratings, assigned according to Scheffé's paired comparison procedure (Scheffé, 1952). Listeners had to indicate which of the samples in a pair (A or B) had a higher degree of 'pleasantness' and 'ease of intelligibility', and also had to assign a scale value to this degree of preference. In case of equal valuation of the speech samples, they assigned a '0'. In case of perceived inequality, they were asked to assign a '1' (slight preference), a '2' (moderate preference) or a '3' (strong preference) to the preferred speech sample. The listening task was computerized using the E-prime software tool (Schneider, Eschman, & Zuccolotto, 2007).

b) Orthographic transcription scores of semantically unpredictable sentences (SUS)

For each speaker, the audio recorded SUS at T1 to T5 were randomly presented to four trained and four untrained listeners, who completed an orthographic transcription task for 50 (5 X 10) SUS. Thus, 64 (8 X 8) listeners participated. The percentage of correctly transcribed key words in SUS was utilized as a measure of speech intelligibility in dysarthric speakers. Each sentence contained four key words, implying a maximum score of 40 correctly transcribed key words in 10 sentences (100%) per assessment. In a previous study, orthographic transcription scores (OTS) of different SUS sets implied equivalence with regard to potential intelligibility in dysarthric speech (Beijer et al., 2012; de Rue, 2012). In addition, the sensitivity to different degrees of intelligibility in dysarthric speakers was established for this outcome measure in the absence of a main effect for 'rater'. Hence, OTS of the different SUS sets in repeated measures were considered suitable for employment in clinical outcome studies.

2) Acoustic measures

Prior to conducting the acoustic measurements, a high-pass filter of 70 Hz was used to reduce interference of environmental noise due to the audio recordings in speakers' home environment.

a) Delta Intensity During Closure (ΔIDC) in voiceless plosives.

Intensity During Closure (IDC) is a variable that measures the acoustic quality of voiceless stop consonant production and is particularly sensitive to incompleteness of closure (undershooting) during the intended silent phase. Since the perceptual dimension 'imprecise consonants' is a hallmark of dysarthria, which applies to nearly all the perceptual dysarthria types as described by Darley (Darley, Aronson, & Brown, 1969), we consider the IDC a suitable measure of articulatory deficits which lead to incomplete occlusion of the vocal tract (Ackermann et al., 1991) in voiceless plosives /p/, /t/, /k/. That is, incomplete occlusion results in increased intensity during the intended silent phase of voiceless stop consonants. In our study, since the conditions for audio recording in patients' home environment were less ideal, we had to compensate for the environmental noise. Therefore, instead of IDC, in our study we calculated ΔIDC (delta Intensity During Closure), in which the environmental intensity was subtracted from the intensity during the intended silent interval in voiceless plosives

b) Intensity of vowels

The intensity of the vowels following the voiceless plosives in the unstressed syllables (see under acoustic measure 'a' above) was measured. An increased intensity after EST was expected, due to the explicit PLVT instructions to speak louder in PD patients, and due to general physical efforts to improve segmental and suprasegmental speech elements in patients with stroke and TBI.

c) *Movements of the vowels /a/, /e/ and /o/ in the F1-F2 space*

In our study, we investigated the vowels /a/, /e/ and /o/ in unstressed syllables. The position of vowels can be graphically presented by the articulatory-acoustic vowel space, and is based on the first formant (F1) and the second formant (F2). To a certain degree the values of F1 and F2 are associated with tongue body position: F1 is associated with high-low position of the tongue, whereas F2 corresponds to front-back position. Restricted mobility of the tongue due to paralysis or hypokinesia in neurological patients is reflected by abnormalities in the F1-F2 patterns in vowels. Centralization in vowel space belongs to the most frequently reported abnormalities of vowel production in dysarthric speakers, and is related to diminished intelligibility (Kent, Duffy, Kent, Vorperian, & Thomas, 1999; Kent & Kim, 2003). On the other hand, improvement of speech quality is associated with an increased vowel space area. The reason for selecting the vowels /a, e, o/ instead of the cardinal vowels /a, i, u/ is that /a, e, o/ allow more variability in F1 and F2. As is shown in Figure 7.2, for the target vowels in our study improved speech quality would result from an increased F1 in /a/ (F2 variability does not apply to this vowel), a decreased F1 along with an increased F2 in /e/ and a decrease of both F1 and F2 in /o/.

d) *Standard deviation of the fundamental frequency (F0) in SUS.*

The standard deviation of the fundamental frequency F0 was used as an outcome measure which reflects F0 excursion. Speakers tend to use larger F0 excursions to realize stronger sentence accents. These stronger accents tend to be associated with more pronounced articulatory movements, aimed at by the speech training programs implemented in EST (Laures & Weismer, 1999; Neger, 2010; Kim, Kent, & Weismer, 2011; Laures & Bunton, 2012).

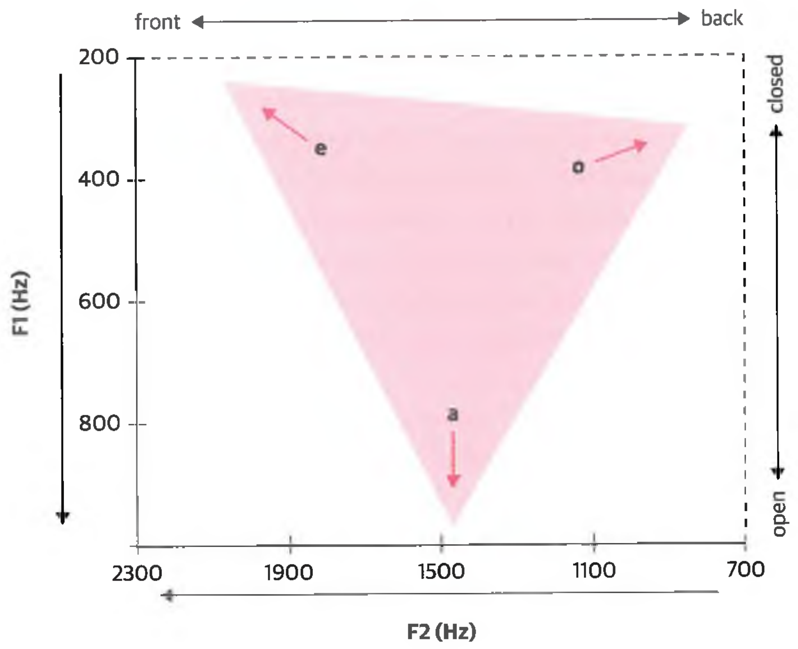


Figure 7.2:

The increase in acoustic vowel space (arrows) is caused by movements of the first formant (F1), which corresponds to high-low position of the tongue, and by movements of the second formant (F2), which corresponds to back-front position of the tongue (adapted from van Bergem, 1993).

7.3 Data analysis

For all participants in the experimental and in the control group, descriptive statistics were calculated for all outcome measures across the experimental period. The orthographic transcriptions of SUS were scored according to the scoring decision diagram as proposed by Beijer, Clapham & Rietveld (2012). For the scale ratings on 'ease of intelligibility' and 'pleasantness' in continuous speech, ANOVAs for paired comparisons (Scheffé, 1952) were carried out with dedicated software, written by A.C.M. Rietveld[1]. Scheffé's analyses yield F-ratios for main effects of time and for order effects (AB versus BA). The scale ratings on 'ease of intelligibility' and 'pleasantness' in speech recorded at T2, T3, T4 and T5, were positioned in ranked order on an interval scale. The analyses were carried out for the two texts realized for the experimental group (n=5) and for the control group (n=3), both for trained listeners and for untrained listeners. In addition, Scheffé's procedure was carried out for the two texts in each individual speaker (n=8), thus yielding 16 interval scales. Scheffé's procedure also yields a 'yardstick' to assess which pairs of stimuli have significantly different positions on the resulting scale of preferences. Repeated measures ANOVAs were conducted for SUS orthographic transcription scores, for ΔIDC in voiceless plosives and for the standard deviations of F0 in SUS. When appropriate, Huynh-Feldt adjusted p-values were used. When main effects of 'time' were established, post-hoc comparisons were carried out with adjustments for multiple comparisons according to Sidaks correction (Abdi, 2007). A statistical significance level of 5% was adopted for all analyses, unless reported otherwise.

With regard to F1 and F2 in the vowels /a/, /e/ and /o/, the movements in the expected direction between T2 and T3 and between T4 and between T5 were assigned a score of '1' and were tallied up for each vowel and for each participant. On the count data thus obtained, repeated measures ANOVAs with 'time' and 'vowel' as within-subjects variables were carried out.

The analyses for all outcome measures were conducted for the experimental and the control group as well as for all individual participants in each group. Only for the movements of F1 and F2 in the vowels no analyses at individual level were conducted, as the raw data for each condition were tallied counts.

1 Available on demand.

7.4 Results

1) **Perceptual measures**

a) Judgements of speech quality in continuous speech

Assessments of 'ease of intelligibility' or 'pleasantness' did not yield consistent results in the experimental or in the control group. No significant differences were observed between the results for trained and for untrained listeners. Although main effects of time were established in some individual participants in both groups, post-hoc comparisons were either not statistically significant or did not show consistent patterns of positive EST effects. An overview of the ANOVA outcomes for paired comparisons is provided in Appendix 7B.

b) Orthographic transcription scores (OTS) in SUS

The results of repeated measures ANOVAs for OTS in the experimental and the control group did not show clear effects of EST. Although 'time' significantly affected OTS in both groups, post-hoc comparisons did not show significant improvements after EST. A t-test for dependent samples (two-tailed, t(39)=3.19, r=.46) showed that trained listeners (M=86.7%, SE=1.2) had significantly lower scores than untrained listeners (M=88.4%, SE=1.1). Tukey's test for additivity was not significant ($F_{1,39}$=3.77, p=.06). Consequently we can assume that there was no interaction between the groups of listeners and the items.

At individual level, participants in whom 'time' significantly affected OTS, showed no significant differences between assessments according to post-hoc comparisons and showed no consistent patterns of positive EST effects on OTS. For both groups, an overview of the descriptive statistics and the ANOVA outcomes for group level and individual level, are given in Appendix 7C.

2) **Acoustic measures**

a) Intensity During Closure (ΔIDC) in voiceless plosives.

The results of repeated measures ANOVAs for ΔIDC in each group are presented in Table 7.3.

Table 7.3:

The results of the repeated measures ANOVAs (T1 to T5) of mean ΔIDC for the experimental group (n=5) and the control group (n=3). Standard deviations are reported between brackets. The outcomes were based on acoustic measurements in 45 utterances (15 utterances for /p/, /t/ and /k/ each), produced by all participants in both groups. The shaded cells represent the pre-post EST assessments in each group.

	Δ IDC (dB)							
Group	T1	T2	T3	T4	T5	ηp2	F-ratio	p
Exp	6.0 (5.2)	4.2 (5.5)	7.5 (5.9)	7.7 (5.5)	6.5 (6.4)	.08	$F_{4,897}$=18.60	.00*
Control	6.3 (5.5)	5.6 (6.5)	5.9 (5.4)	5.3 (6.1)	6.7 (6.2)	.02	$F_{4,536}$ = 0.70	.09

** = significant (p < 0.05)*

In the experimental group a main effect of time on ΔIDC was observed, although the effect size was small η_p^2 =.08. Post-hoc comparisons showed a significantly increased ΔIDC between T2 and T3 (i.e. the start and the end of EST training, respectively), along with significant increases between other pre- and post EST assessments: T1-T3, T1-T4, T2-T4, T2-T5. In the control group no main effect of 'time' on ΔIDC was found (F4,536= 2.0, η_p^2=.02).

In the experimental group, at individual level a main effect of 'time' was found in participant 2 ($F_{4,176}$=3.06, η_p^2=.07), in participant 3 ($F_{4,176}$=26.74, η_p^2=.38), in participant 4 ($F_{4,176}$=12.04, η_p^2=.22) and in participant 5 ($F_{4,176}$=3.24, η_p^2=.07). Post-hoc comparisons showed a significantly increased ΔIDC between T2 and T3 in participants 3, 4 and 5. In contrast, a decreased ΔIDC of 1.6 dB (SE=0.55), approaching statistical significance (p=.055), was observed in participant 2. ΔIDC significantly increased with 5.4 dB (SE=0.88) between T2 and T4 in participant 4. Apart from pre-post EST comparisons, significant decreases were observed between T4 and T5 in participant 3 (mean difference=3.73 dB, SE=1.0) and between T1 and T2 in participant 4 (mean difference=3.7 dB, SE= 0.786). In Figure 7.3 the ΔIDC is presented for all individual participants in the experimental group.

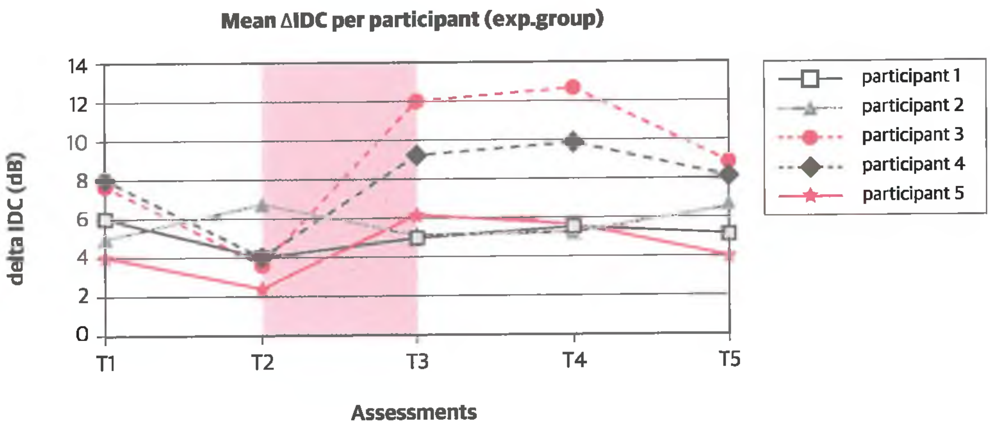


Figure 7.3:

Mean Δ IDC in the voiceless plosives /p/, /t/ and /k/ for the participants in the experimental group.

In the control group, at individual level, 'time' significantly affected ΔIDC in participant 6 ($F_{4,176}$=2.61, η_p^2= .06) with a significantly increased ΔIDC of 2 dB between T4 and T5. Although a main effect of time on ΔIDC was found in participant 8 ($F_{4,176}$=2.70, η_p^2= .06), post-hoc comparisons showed no significant differences between any assessments. No main effect of time was found in participant 7. Figure 7.4 shows the ΔIDC measured at the assessments in the controls.

The results of separate analyses for each voiceless plosive /p/, /t/ and /k/ in both the experimental and the control group, along with the outcomes for the individual participants per group are presented in Appendix 7D.

b) Intensity of vowels

The results of repeated measures ANOVAs for mean intensities in vowels in each group are presented in Table 7.4. A main effect of 'time' on intensities in vowels was established in both groups. In the experimental group intensity significantly increased between T2 and T3, according to post-hoc comparisons, whereas a significant decrease was observed between T4 and T5. In the control group however, no increase was established in the control condition (i.e., between T2 and T3) whereas a significant increase occurred between T4 and T5, that is, after EST.

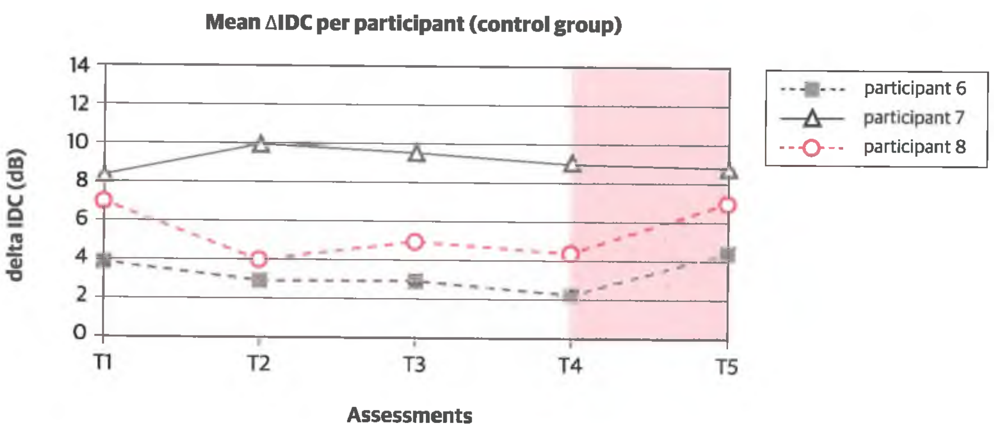


Figure 7.4:

Mean Δ IDC in voiceless plosives /p/, /t/ and /k/ for the participants in the control group.

Table 7.4:

The results of the repeated measures ANOVAs for mean intensities (dB) of the vowels following the voiceless plosives in unstressed syllables for the experimental group (n=5) and the control group (n=3). Standard deviations are reported between brackets. Post-hoc comparisons between T2 and T3, and between T4 and T5 are reported. The shaded cells represent the pre-post EST assessments in each group.

	T2	T3	T4	T5	$F_{df1,df2}$	ηp^2	p	T2-T3	T4-T5
Exp.	32.7 (3.6)	36.8 (5.4)	36.9 (7.8)	33.6 (5.6)	$F_{3,87}$=13.8	.32	.00	.00*	.00*
Control	31.4 (4.0)	31.3 (5.4)	30.0 (4.8)	34.2 (5.8)	$F_{3,51}$=19.9	.54	.00	1.00	.00*

** = significant ($p < 0.05$)*

c) *Movements of F1 and F2 in vowels.*

No significant effect of 'time' or 'vowel' on movements of F1- or F2 was observed in the experimental or in the control group.

In the experimental group, a repeated measures ANOVA showed no significant effect of 'time' ($F_{1,4}$<1, p=1.00) or 'vowel' ($F_{2,8}$=1.18, p=.36) on F1. No interaction was observed for 'time' and 'vowel'($F_{2,8}$=1.47, p=.29). The F2 was not significantly affected either by 'time' ($F_{1,4}$<1, p=1.00) or 'vowel'($F_{1,4}$<1, p=.50). No interaction was observed for 'time' and 'vowel' ($F_{1,4}$<1, p=.78).

In the control group F1 was not affected by 'time' ($F_{1,2}$<1, p=.74) or 'vowel' ($F_{2,4}$=2.33, p=.21), and no interaction between these two factors was found ($F_{2,4}$=3.13, p=.15). F2 was not affected by 'time' ($F_{1,2}$<1, p=.67) or 'vowel' ($F_{1,2}$=1.75, p=.32) in the control group, but an interaction between these factors was observed ($F_{1,2}$=30.25, η_p^2= .94, p=.03).

d) *Mean standard deviations of the fundamental frequency in SUS*

In both the experimental group and the control group, the standard deviation of F0 was significantly affected by 'time'. Post-hoc comparisons showed a statically significant increased standard deviation at T5 compared to T2 in the experimental group. In the control group the standard deviations at both T3 and T4 were significantly higher than at T2. For both groups, an overview of the descriptive statistics and the ANOVA outcomes at group level and individual level, are given in Appendix 7G.

7.5 Discussion

This paper reports an exploratory study as an attempt to identify beneficial effects of EST on speech intelligibility in dysarthric speakers and to determine the suitability of perceptual and acoustic outcome measures for this purpose. At first sight, the outcomes of our study do not provide a clear picture of positive EST effects on speech intelligibility. However, a closer look at the results for the different outcome measures may yield some beneficial insights for future effect studies.

The most consistent observation in our study concerned the ΔIDC, being the intensity during closure in the intended silent phase of voiceless plosives relative to the environmental intensity. The ΔIDC significantly increased immediately after EST in the experimental group. At the individual level an increase between assessments T2 and T3 could also be demonstrated in participants 3, 4 and 5. In contrast, in the control participants – who were not provided with EST between these assessments – no significantly increased ΔIDC was observed, neither at the group level nor in any individual participant. Control participant 6 showed a significant increase immediately after EST (i.e. between assessments T4 and T5). A closer look at the results for /p/, /t/ and /k/ (Appendix 7D) showed similar patterns of significant differences between pre-post EST assessments in the experimental group, whereas significant differences between the assessments T2 and T3 in the control condition did not occur for any voiceless plosive. ΔIDC for /k/ in the control group showed a significant increase between T4 and T5 (i.e. after EST). Overall, these observations support our hypothesis that EST affects the intensity during the silent phase in voiceless plosives, although in the opposite direction than assumed. That is, ΔIDC was expected to decrease after EST due to improved closure of the vocal tract during the intended silent phase. We hypothesized that the increased intensity after EST was caused by the speakers' efforts to speak louder (i.e. according to the 'speak loud, speak low' PLVT instruction and as a result of general physical efforts to improve speech quality), while failing to adjust articulatory settings. Thus, adequate occlusion of the vocal tract may not have been achieved. This hypothesis was supported by intensity measurements of the vowels in the unstressed syllables containing the voiceless plosives. Only after EST the mean vowel intensity significantly increased. That is, between T2 and T3 in the experimental group and between T4 and T5 in the control group. In the control condition between T2 and T3 or between T4 and T5 in the experimental group intensity did not increase.

To illustrate the effect of increased lung pressure - resulting in higher intensity - without changing settings of the oral musculature (like the lips), we used the software for articulatory synthesis provided by Praat (for details see Boersma, 1998: 113-131). The settings represent muscle activity as a variable between zero (at rest) and one (fully contracted). The amount of activity at a given point on the timeline is linearly interpolated between target points. The parameters simulating time-modulated actions of the muscles involved (mm. interarytenoideus, levator palatini, hyoglossus, genioglossus, masseter, cricothyroideus, orbicularis oris, styloglossus), the tongue and the lungs were set in such a way that the configuration of the vocal tract generated the vowel /↔/, with an interval of 40 ms in which some noise was generated. Two settings of the activity of the lungs were used: 0.1 and 0.3 (duration 100 ms), of which the latter setting represents higher lung pressure. The effect is a higher intensity of the vowels, but also a higher intensity of the noise of the intended silent interval (80.5 dB vs. 76.15 dB) as the supraglottal settings were not changed. The two different settings of lung activity and the corresponding intensity in the intended silent phase are depicted in Appendix 7F.

With regard to the other outcome measures in our study, insufficient sensitivity to subtle changes in our participants' speech might have played a role. That is, the outcome measures that we used are based on research studies that aimed at detection of ***between-subjects*** differences (i.e., between different dysarthric speakers or between healthy and dysarthric speakers). They are generally used to classify underlying neurological disease, as well as type and severity of dysarthria (Kim et al., 2011). This was also the case for OTS, for which sensitivity to different degrees of intelligibility in dysarthric speakers with PD and stroke was established in previous studies (Beijer, Clapham & Rietveld, 2012; de Rue, 2012), and thus concerned ***between-subjects*** differences. Measuring within-subjects differences to establish the effect of spontaneous or training-induced recovery is considerably more complicated. Confounding factors such as neurological patients' variable physical condition, learning effects, and changing environmental conditions are liable to interfere with genuine changes in speech quality. Detection of subtle within-subjects changes (i.e. small effect sizes) may therefore require higher sensitivity than is needed for the detection of ***between-subjects*** differences. In future studies, the sensitivity of either the measures themselves or of 'environmental factors' (i.e. independent of the measures themselves) should be improved. With regard to enhancing sensitivity of outcome measures, automatic pronunciation assessment would provide a solution. Rather than for example the binary OTS in our study (i.e. correct-incorrect), confidence measures such as the likelihood-ratios for the 'Goodness of Pronunciation' (GOP) (van Doremalen et al., 2009) would provide a continuous scoring scale that allows more variability in scores. With regard to improving 'environmental sensitivity', larger numbers of speech materials or larger numbers of raters would enhance the sensitivity. In addition, the type of speech materials should be suitable for detecting changes in target measures. Detection of relevant F0 variability requires a minimal length of continuous speech, for instance. The fact that general judgements on 'ease of intelligibility' or 'pleasantness', nor the standard deviation of the fundamental frequency (F0), showed consistent patterns in our study, may imply that the length of speech samples for assessments should be adjusted to the target outcome measure. The unchanged variability of F0 in most participants could be explained by the use of ten 6-words SUS in declarative form with only monosyllabic words. This possibly elicits routine intonation, inducing less F0 variability. This routine intonation would imply a learning effect, which may also be reflected by the observed tendency of diminished SUS duration between T2 and T3 in both the experimental and the control group (Appendix 7E). The initial reason for selecting SUS as speech materials for measuring the standard deviation of F0 was that longer sentences are at risk of disfluencies in dysarthric speakers. This would disrupt the original course of fundamental frequency in the utterance. In addition to adequate length of contiouous speech, it should be phonetically balanced and ecologically valid. In addition, the length and the position in the text of the selected passages should allow adequate assessment. With regard to the measurements in plosives, the elicitation of adequate sentence accents should be improved. Whereas the dysarthric participants were able to realize the target vowels in unstressed syllables by responding to a question sentence, the patients experienced more difficulties in the realization of plosives in unstressed syllables (Appendix 7A). They were asked to imitate an aurally presented, relatively long, sentence which was also displayed on the laptop screen.

Obviously, the development of suitable speech materials for speech assessments in effect studies remains a challenge. Although the enhancement of 'environmental sensitivity' calls for larger amounts of speech materials, dysarthric speakers' neurological conditions do not allow long sessions of speaking. Therefore, an adequate balance should be found between the need for enhanced sensitivity and the physical burden in neurological speakers. Future research should address the optimization of sensitivity of outcome measures and of 'environmental factors' such as the type an amount of speech materials to detect subtle changes in speech over time. Studies should be focused on acoustic measures related to segmental speech dimensions. Adequate speech materials for outcome measures such as F0 variability should be selected.

Apart from their role in effect studies, acoustic measurements may provide directions to adjustments in speech training programs. For example, the observed increased ΔIDC in the intended silent phase of voiceless plosives in combination with higher intensities in vowels imply that patients would benefit from specific instructions in speech training to achieve complete closure of the vocal tract to improve articulation of plosives along with increased speech intensity.

7.6 Conclusions

This study sought to explore the suitability of perceptual and acoustic measures to evaluate the effects of E-learning based Speech Therapy (EST) on speech intelligibility. The acoustic measure ΔIDC yielded the most consistent results in our study. Overall, the sensitivity of both acoustic and perceptual outcome measures as well as sensitivity of environmental factors such as amount of speech materials should be improved to detect small within-subjects effects in effect studies on speech intelligibility. In this respect, the development of adequate and sufficient speech materials, along with high-quality audio recordings of speech, is of vital importance.

Acknowledgement

The authors wish to thank Joop Kerkhoff, Anke van Stiphout, Gerna Scholte and Sonja Peters for their assistance in performing the experiments and the data analyses in this study.

References

Abdi, H. (2007). The Bonferroni and Sidak corrections for multiple comparisons. In N.Salkind (Ed.), *Encyclopedia of Measurement and Statistics.* Thousand Oaks, CA: Sage.

Ackermann, H. & Ziegler, W. (1991). Articuiatory deficits in parkonsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry, 54,* 1093-1098.

Basso, A. (2005). How intensive/prolonged should intensive/prolonged treatment be? *Aphasiology, 19,* 975-984.

Beijer, L. J., Clapham, R. P., & Rietveld, A. C. M. (2012). Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20,* 17-34.

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010a). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) as a web application for speech training in dysarthric patients with Parkinson's Disease: a case study. *Telemedicine and e-Health,* 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., & Geurts, A.C.H. (2010b). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16,* 177-180.

Beijer, L. J., Rietveld, A. C. M., & van Stiphout, A. (2011). Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44,* 701-718.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34,* 987-993.

Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.0.25) [computer program]. Retrieved March 23, 2009 from http://www.praat.org/.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246-269.

De Bodt, M. S., Hernandez-Diaz Huici, M. E., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders, 35,* 283-292.

Hill, A. (2008). *The assessment of acquired neurogenic communication disorders in adults using a telerehabilitation application.* PhD thesis. The University of Queensland, Brisbane.

Jancovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery and Psychiatry, 79,* 368-376.

Kent, R. D., Duffy, J., Kent, J. F., Vorperian, H. K., & Thomas, J. E. (1999). Quantification of motor speech abilities in stroke: time-energy analyses of syllable and word repetition. *Journal of Medical Speech-Language Pathology, 7,* 83-90.

Kent, R. D., Kent, J. F., Duffy, J., & Weismer, G. (1998). The dysarthrias: speech-voice profiles, related dysfunctions, and neuropathology. *Journal of Medical Speech-Language Pathology, 6,* 165-211.

Kent, R. D. & Kim, Y.-J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics and Phonetics, 17,* 427-445.

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54,* 417-429.

Kwakkel, G. (2006). Impact of intensity of practice after stroke: issues for consideration. *Disability and Rehabilitation, 28,* 823-830.

Laures, J. S. & Bunton, K. (2012). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders, 36,* 449-464.

Laures, J. S. & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research, 42,* 1148-1156.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J. et al. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17,* 277-298.

Neger, T. (2010). Acoustic measures of the automatic assessment of intelligibility in dysarthric speech: a case study. Unpublished thesis, Radboud University Nijmegen.

Palmer, R., Enderby, P., & Cunningham, S. (2004). The effect of three practice conditions on the consistency of chronic dysarthric speech. *Journal of Medical Speech-Language Pathology, 12,* 183-188.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry, 71,* 493-498.

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37,* 401-411.

Rue, N. P. W. D. de (2012). *Verstaanbaarheidsmaten voor dysartrische spraak. Orthografische transcriptiescores en schaalscores van Semantically Unpredictable Sentences (SUS) als verstaanbaarheidsmaat voor dysartrische spraak als gevolg van CVA.* Unpublished thesis, Radboud University Nijmegen.

Rvachew, S. (2004). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37,* 347-357.

Scheffé, H. (1952). An analysis of variance for paired comparisons. *Journal of the American Statistical Association, 47,* 381-400.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). E-Prime (Version 2.0) [Computer software].

Swart, B. J. M. de, Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60,* 498-500.

Van Bergem (1993). Acoustic vowel reduction. *Speech Communication, 12,* 1-23.

Wang, Y. T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2005). Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress. *Journal of Communication Disorders, 38,* 231-260.

Winters, J. M. & Winters, J. M. (2004). A telehomecare model for optimizing rehabilitation outcomes. *Telemedicine and e-Health, 10,* 200-2.

Appendix 7A

Semantically Unpredictable Sentences (declarative form, six words)

Elicitation procedure:
For each assessment (T1 to T5) the sentences of a SUS set (SUS 1 to SUS5) were read aloud by the dysarthric speakers. The speakers were asked to read the sentences carefully before reading aloud.

	SUS set number	Sentence number	Semantically Unpredictable Sentence (SUS)					
			determinant	Noun1	Verb	preposition	determinant	Noun2
		Example1	*een*	*soep*	*wreef*	*aan*	*een*	*bal*
		Example2	*een*	*doos*	*klom*	*aan*	*een*	*lijk*
T1	1	1	een	brood	hangt	voor	een	drang
		2	een	gek	vocht	door	een	land
		3	een	plan	was	na	een	lach
		4	een	stuur	zakt	bij	een	klas
		5	een	vat	blijft	uit	een	kind
		6	een	kus	schreed	aan	een	maag
		7	een	woord	loopt	langs	een	stand
		8	een	duim	stond	op	een	zak
		9	een	poes	duidt	op	een	kip
		10	een	vrees	mag	aan	een	trein
T2	2	1	een	nacht	goot	door	een	schot
		2	een	hut	komt	bij	een	man
		3	een	vlucht	sloop	langs	een	kaas
		4	een	plein	zonk	in	een	blad
		5	een	strijd	hing	naar	een	tong
		6	een	ziel	keek	op	een	lijn
		7	een	vlak	lacht	met	een	hof
		8	een	greep	blies	van	een	staat

	SUS set number	Sentence number	Semantically Unpredictable Sentence (SUS)					
			determinant	Noun1	Verb	preposition	determinant	Noun2
		9	een	bos	dreef	door	een	kamp
		10	een	paard	blaast	voor	een	ring
T3	3	1	een	straf	zoog	aan	een	haar
		2	een	zon	stoof	door	een	jas
		3	een	schaal	ligt	op	een	maat
		4	een	bed	stonk	van	een	moord
		5	een	deel	ruikt	naar	een	taal
		6	een	hoed	bleef	bij	een	graf
		7	een	stok	trok	met	een	boot
		8	een	veld	deelt	in	een	baard
		9	een	prooi	buigt	voor	een	rij
		10	een	krant	steeg	naar	een	pan
T4	4	1	een	cel	liep	langs	een	hoek
		2	een	schoot	hield	van	een	maan
		3	een	feit	grijpt	naar	een	stof
		4	een	dal	werkt	sinds	een	vuur
		5	een	film	stak	aan	een	rol
		6	een	som	duikt	in	een	boog
		7	een	droom	trekt	uit	een	mes
		8	een	kans	past	op	een	tand
		9	een	blik	kroop	langs	een	pond
		10	een	lid	zwaait	met	een	boord
T5	5	1	een	vrucht	scheen	naar	een	staf
		2	een	ster	bijt	naar	een	kop
		3	een	pot	dacht	met	een	hek
		4	een	buik	lag	aan	een	spel
		5	een	mond	zwom	voor	een	boer
		6	een	bak	stinkt	door	een	zin
		7	een	kruis	dook	op	een	recht
		8	een	toon	rookt	uit	een	hulp
		9	een	kaart	praat	bij	een	stoep
		10	een	naam	sprong	langs	een	hand

Appendix 7A (continued)

Continuous speech

Elicitation procedure:
The texts were read aloud by the dysarthric speakers.
(Selected passages for general judgements of speech quality shaded in red).

Oronasal text: 'Papa en Marloes'

Papa en Marloes staan op het station.
Ze wachten op de trein.
Eerst hebben ze een kaartje gekocht.
Er stond een hele lange rij, dus dat duurde wel even.
Nu wachten ze tot de trein eraan komt.
Het is al vijf over drie, dus het duurt nog vier minuten.
Er staan nog veel meer mensen te wachten.
Marloes kijkt naar links, in de verte ziet ze de trein al aankomen.

General text: Finland'

Er was eens een man uit Finland.
Hij had veel geld gespaard.
Dat was voor de auto van zijn dromen.
Hij nam de trein om de auto te gaan kopen.
Maar de man was bang voor dieven.
Hij bewaarde het geld in zijn onderbroek.
Hij droomde al van de eerste rit in de nieuwe wagen.
Plots moest hij naar het toilet.
De man dacht niet meer aan het geld.
Het zakje met geld viel recht in de pot.
En de man spoelde door.
Daar ging zijn fraaie plan!
Gelukkig was de politie in de buurt.
Die vond het zakje terug op de sporen

Appendix 7A (continued)

Voiceless plosive /k/ in word initial position and in unstressed syllable, embedded in sentences and bisyllabic words.

Elicitation procedure:
Imitation of the sentence audio examples, supported by the orthographically presented sentence.

/k/-sentences
(Target phonemes in capitals)

1. Het meest lach ik met mijn **broer** en met mijn beste **Kameraad**.
2. Het feest zal zijn in de prachtige **tuin** van het grote **Kasteel**.
3. In de oorlog bouwde men volop **bunkers** en **Kazematten**.
4. Op de barbecue ligt een **braadworst** en een **Karbonade**.
5. De jongen bekijkt de **vissen** en het schitterende **Koraal**.
6. In de snackbar koop ik een zak **patat** en een broodje **Kebab**.
7. Peuters worden opgevangen in een **peuterspeelzaal** of **Kinderdagverblijf**.
8. De pottenbakker gebruikt een mengsel van **klei** en **Keramiek**.
9. In Japan draagt men geen **badjas**, maar een **Kimono**.
10. Vroeger stonden er zware straffen op **stelen** en **Ketterij**.

/k/-words
(Target phonemes in capitals)

1. beKer
2. beuKen
3. duiKen
4. keuKen
5. koKer

Appendix 7A (continued)

Voiceless plosive /t/ in word initial position and in unstressed syllable, embedded in sentences and in bisyllabic words.

Elicitation procedure:

Imitation of the sentence audio examples, supported by the orthographically presented sentence.

/t/-sentences

(Target phonemes in capitals)

1. De peuter rammelt graag op een **trommel** of een **Tamboerijn**.
2. In het bushokje hangt een vieze lucht van wiet en **Tabak**.
3. Er ligt een hoop in puin na de vele **regen** en de **Tornado**.
4. Bij de visboer koop ik altijd veel **zalm**, maar geen **Tonijn**.
5. Op de salade ligt een vers **ei** en een knalrode **Tomaat**.
6. De kennis wordt getoetst met een **gesprek** en een **Tentamen**.
7. Na een dag hard werken voel ik mij **voldaan** en **Tevreden**.
8. Het nieuws is de hele dag te volgen via **radio** en **Televisie**.
9. Op de markt verkopen ze **vazen** met allerlei **Tierelantijnen**.
10. De lekkerste nagerechten zijn **ijs** en zelfgemaakte **Tiramisu**.

/t/-words

(Target phonemes in capitals)

1. boeTe
2. buiTen
3. meTer
4. raTTen
5. weTen

Appendix 7A (continued)

Voiceless plosive /p/ in word initial position and in unstressed syllable, embedded in sentences and bisyllabic words.

Elicitation procedure:
Imitation of the sentence audio examples, supported by the orthographically presented sentence.

/p/-sentences

(Target phonemes in capitals)

1. Vroeger stonden hier grote **huizen** en mooie **Paleizen**.
2. De clown loopt mee in de **optocht** en in de **Parade**.
3. De witte lijnen lopen uit **elkaar** en de zwarte **Parallel**.
4. Zijn geweldige stem maakte de man **rijk** en **Populair**.
5. De aard van deze man is **optimistisch** en **Positief**.
6. Het feest wordt gevierd voor de **bewoners** en het vele **Personeel**.
7. Heel vroeger schreef men op **klei** en later gebruikte men **Perkament**.
8. Soep is het lekkerst met verse **tomaten** en veel **Peterselie**.
9. Op school gebruikt Joep zijn **bananen** als **Pistool**.
10. De monnik schreef zijn naam op de **deur** en een **Pilaar**.

/p/-words

(Target phonemes in capitals)

1. diePe
2. doPen
3. lePel
4. paPa
5. pePer

Appendix 7A (continued)

The vowels /e/, /a/ and /o/ in unstressed syllables, embedded in sentences.

Elicitation procedure:
The examiner asked a question (between brackets) and provided the answer with the intended sentence accents. The sentence accents were represented by bold and lower case letter type. The participant responded by imitating the answer.

/a/-sentences

(Target phonemes in capitals)

1. *(Wat bouwen zij samen?)*
   SAmen bouwen zij een pa**leis**.

2. *(Wat is er waardeloos geworden?)*
   Haar pakket **aan**delen is w**AA**rdeloos geworden.

3. *(Waarop ontbreekt het zadel?)*
   Op deze **fiets** ontbreekt het zAdel.

4. *(Waardoor was de minister verlaat?)*
   De minister was verl**AA**t vanwege een **fi**le.

5. *(Wat is de kleur van een tomaat?)*
   Rood is de kleur van een tom**AA**t.

/e/- sentences

1. *(Wie deed zelf ook mee?)*
   De po**li**tieagent deed zelf ook m**EE**.

Appendix 7A (continued)

2. *(Wie gebruikt een deegrol?)*
   De **bak**ker gebruikt een d**EE**grol.

3. *(Wat doet het kleine meisje beven?)*
   De harde **knal** doet het kleine meisje b**E**ven.

4. *(Waarom werd de atleet bewonderd?)*
   De atl**EE**t werd bewonderd om zijn **in**zet.

5. *(Wie neemt het toezicht in de mediatheek over?)*
   In de mediath**EE**k neemt de por**tier** het toezicht over.

/o/-sentences

1. *(Wie brengt een bezoek aan Polen?)*
   De konin**gin** brengt een bezoek aan P**O**len

2. *(Wat heeft een olifant?)*
   Een **O**lifant heeft een lange **slurf**.

3. *(Wie houdt van lekker koken?)*
   Mijn **schoon**moeder houdt van lekker k**O**ken.

4. *(Wie verft het huis naast de boom?)*
   De **schil**der verft het huis naast de b**OO**m.

5. *(Waarin zit een atoom?)*
   Een at**OO**m zit in een mole**cuul**.

Appendix 7B

Results of ANOVAs for paired comparisons of general judgements on speech quality

Table 7B1:

Outcomes of Scheffé 's procedure for ANOVA's for paired comparisons of general judgements on perceived 'ease of intelligibility' in the experimental group.

The results for the individual speakers were based on judgements of four trained and four untrained listeners (n=8).

Experimental group												
Ease of intelligibility												
Text	speakers	listeners	df1, df2	F	order		T2-T3	T2-T4	T2-T5	T3-T4	T3-T5	T4-T5
Papa en Marloes	**group** (n=5)	n=64 (all)	3,228	0.74^{o}	5<2<3=4	0.33						
		n=32 (trained)	3,108	1.09^{o}	5<2<3<4	0.40						
		n=32 (untrained)	3,108	0.33	5<4<2<3	0.53						
	individual											
	1	n=8	3,36	0.64	2<3<5<4	0.66						
	2	n=8	3,36	0.78	5<4<2<3	0.81						
	3	n=8	3,36	6.15*^{o}	5<2<4<3	0.51	x	x		x	x	
	4	n=8	3,36	3.90*	3<4<2<5	0.75					x	
	5	n=8	3,36	5.52*^{o}	5<3<2<4	0.62				x		x
Finland	**group** (n=5)	n=64 (all)	3,228	1.73^{o}	4<5<2<3	0.31						
		n=32 (trained)	3,108	1.14^{o}	5<4<2<3	0.43						
		n=32 (untrained)	3,108	0.79	4<2<5<3	0.48						
	individual											
	1	n=8	3,36	5.39*	5<2=4<3	0.67					x	x
	2	n=8	3,36	1.50^{o}	5<4<2<3	0.60						
	3	n=8	3,36	4.19*	5<3<2<4	0.63				x		x
	4	n=8	3,36	7.64*	4<3<2<5	0.72					x	x
	5	n=8	3,36	14.72*^{o}	4<2=3<5	0.45		x		x		

***= significant main effect (p<.05), ^{o}= significant order effect p<.05), x = significant differences between assessments In the column 'order' the speech samples recorded at assessments 2 to 5 are ranked in ascending order with regard to 'ease of intelligibility'.**

Table 7B2:

Outcomes of Scheffé 's procedure for ANOVA's for paired comparisons of general judgements on perceived 'pleasantness' in the experimental group.

The results for the individual speakers were based on judgements of four trained and four untrained listeners (n=8).

Experimental group												
Pleasantness												
Text	speakers	listeners	df1, df2	F	order		T2-T3	T2-T4	T2-T5	T3-T4	T3-T5	T4-T5
Papa en Marloes	**group** (n=5)	n=64 (all)	3,228	0.53^{o}	4<5<3<2	0.34						
		n=32 (trained)	3,108	0.54	3<4<5<2	0.46						
		n=32 (untrained)	3,108	0.77	4<5<2<3	0.54						
	individual											
	1	n=8	3,36	0.60	2=4<5<3	0.67						
	2	n=8	3,36	1.13	4=5<2<3	0.70						
	3	n=8	3,36	2.82	5<2<3<4	0.76						
	4	n=8	3,36	10.0*	3<4<2<5	0.68	x		x		x	
	5	n=8	3,36	0.14	4<2<5<3	0.77						
Finland	**group** (n=5)	n=64 (all)	3,228	0.89	4<5<2<3	0.33						
		n=32 (trained)	3,108	1.08	4<3=5<2	0.45						
		n=32 (untrained)	3,108	0.54	4<2<3=5	0.50						
	individual											
	1	n=8	3,36	0.32	5<2=4<3	0.85						
	2	n=8	3,36	1.27	5<4<2<3	0.73						
	3	n=8	3,36	2.05	5<3<2<4	0.87						
	4	n=8	3,36	10.36*^{o}	4<3<2<5	0.57		x	x		x	x
	5	n=8	3,36	3.66*	4<3<2<5	0.61						x

= significant main effect (p<.05), ^{o}= significant order effect p<.05), x = significant differences between assessments

In the column 'order' the speech samples recorded at assessments 2 to 5 are ranked in ascending order with regard to 'pleasantness'.

Table 7B3:

Outcomes of Scheffé 's procedure for ANOVA's for paired comparisons of general judgements on perceived 'ease of intelligibility' in the control group.

The results for the individual speakers were based on judgements of four trained and four untrained listeners (n=8).

Experimental group												
Ease of intelligibility												
Text	speakers	listeners	df1, df2	F	order		T2-T3	T2-T4	T2-T5	T3-T4	T3-T5	T4-T5
Papa en Marloes	**group** (n=3)	n=64 (all)	3,132	0.80^{o}	3<5<2<4	0.44						
		n=32 (trained)	3,60	0.16	4<3<2<5	0.71						
		n=32 (untrained)	3,60	2.42^{o}	3<5<2<4	0.58						
	individual											
	6	n=8	3,36	0.90	3<4<2=5	0.80						
	7	n=8	3,36	3.05*	5<2<3<4	0.88						x
	8	n=8	3,36	1.59^{o}	3<2=4<5	0.62						
Finland	**group** (n=3)	n=64 (all)	3,132	4.91*^{o}	2<5<3<4	0.34						
		n=32 (trained)	3,60	3.00*	3<2<5<4	0.44						
		n=32 (untrained)	3,60	2.92*	5<2<3<4	0.55						
	individual											
	6	n=8	3,36	2.56	3=5<2<4	0.60						
	7	n=8	3,36	2.43	5<2=4<3	0.52						
	8	n=8	3,36	10.33*^{o}	3<2<5<4	0.52		x		x		

**= significant main effect (p<.05), ^{o}= significant order effect p<.05), x = significant differences between assessments*

In the column 'order' the speech samples recorded at assessments 2 to 5 are ranked in ascending order with regard to 'ease of intelligibility'.

Table 7B4:

Outcomes of Scheffé 's procedure for ANOVA's for paired comparisons of general judgements on perceived 'pleasantness' in the control group.

The results for the individual speakers were based on judgements of four trained and four untrained listeners (n=8).

Experimental group												
Pleasantness												
Text	speakers	listeners	df1, df2	F	order		T2-T3	T2-T4	T2-T5	T3-T4	T3-T5	T4-T5
Papa en Marloes	**group** (n=3)	n=64 (all)	3,132	3.40*	5<2<3<4	0.46						
		n=32 (trained)	3,60	0.60	2=5<4<3	0.62						
		n=32 (untrained)	3,60	4.06*	5<2<3<4	0.70						
	individual											
	6	n=8	3,36	10.70*	5<2<4<3	0.69	x			x	x	
	7	n=8	3,36	0.18	5<2<3=4	0.80						
	8	n=8	3,36	4.40*	3<5<2<4	0.77		x		x		x
Finland	**group** (n=3)	n=64 (all)	3,132	1.71°	3<2<5<4	0.37						
		n=32 (trained)	3,60	0.77	3<4<2<5	0.54						
		n=32 (untrained)	3,60	2.62	3<2<5<4	0.54						
	individual											
	6	n=8	3,36	0.62	2<4<5<3	0.73						
	7	n=8	3,36	0.23	4<5<2<3	0.44						
	8	n=8	3,36	5.92*	3<2<5<4	0.70				x	x	

**= significant main effect (p<.05), °= significant order effect p<.05), x = significant differences between assessments*

In the column 'order' the speech samples recorded at assessments 2 to 5 are ranked in ascending order with regard to 'pleasantness'.

Appendix 7C

Results of repeated measures ANOVAs for orthographic transcription s

Table 7C1:

Results of ANOVAs for repeated measures for mean orthographic transcription score percentages (OTS) at T1 to ard deviations between brackets. The scores were based on orthographic transcription by 8 listeners (four traine Significant differences between assessments according to post-hoc comparisons (Sidak adjusted) are displayed.

Experi-mentals	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	pre-p		
Group (n=5)	$F_{4,156}$								T1-T2	T1-T3	T1-T4
	2.58	.06	.04*	90.2 (4.9)	89.4 (4.7)	91.9 (5.7)	89.1 (7.3)	91.2 (4.7)			
Indiv.	$F_{4,28}$										
1	4.07	.06	.01*	91.0 (3.0)	89.4 (4.4)	93.1 (2.9)	90.0 (4.4)	93.4 (3.5)			
2	6.08	.47	.00*	95.0 (4.0)	88.4 (5.7)	94.4 (2.9)	95.0 (1.3)	92.5 (3.3)			
3	1.66	.19	.19	90.6 (2.6)	91.6 (4.2)	87.8 (3.4)	90.9 (3.0)	89.4 (3.5)			
4	1.23	.15	.32	87.2 (5.2)	89.1 (4.4)	87.5 (7.2)	91.3 (5.5)	88.8 (7.1)			
5	21.26	.75	.00*	87.2 (5.4)	88.4 (5.2)	96.9 (4.6)	78.1 (6.9)	91.9 (4.4)		x	

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7C2:

Results of ANOVAs for repeated measures for mean orthographic transcription score percentages (OTS) at T1 toT5 in the contr deviations between brackets. The scores were based on orthographic transcription by 8 listeners (four trained and four untrain cant differences between assessments according to post-hoc comparisons (Sidak adjusted) are displayed.

Contr.	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	Post-hoc cor				
									pre-post EST compa				
Group (n=3)	$F_{4,32}$								T1-T2	T1-T3	T1-T4	T1-T5	T2-T3
	5.2	18	.00*	84.3(10.0)	80.3 (10.4)	84,0 (11,3)	85,0 (10,2)	79,6 (13,8)					
Indiv.	$F_{4,28}$												
6	6.43	.48	.00*	91.6 (6.0)	84.7 (8.0)	93.1 (6.4)	91.9 (6.2)	86.0 (5.7)	x				
7	9.18	.57	.00*	73.8 (8.8)	73.4 (12.8)	73.1 (10.9)	76.3 (11.6)	63.1 (10.2)					
8	2.42	.26	07	87.5 (3.8)	82.8 (6.7)	85.6 (5.5)	86.9 (5.1)	89.7 (4.7)					

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Appendix 7D

Results of repeated measures ANOVAs for IDC in voiceless plosives /k/

Table 7D1:

Results of repeated measures ANOVAs for /k/ in the experimental group. The shaded blocks regard pre-post EST-

Experi-mentals	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	pre-po		
Group (n=5)	$F_{4,176}$								T1-T2	T1-T3	T1-T4
	8.29	.20	.00*	6.5(5.5)	4.5(5.4)	8.3(5.4)	8.3(5.7)	6.3(5.6)	x		
Indiv.	$F_{4,56}$										
1	1.84	.12	.14	8.2(4.9)	4.5(7.9)	6.8(5.2)	4.9(5.5)	3.4(6.1)			
2	.92	.06	.46	5.3(2.9)	6.8(2.9)	5.1(3.5)	6.2(3.1)	5.6(2.5)			
3	11.42	.45	.00*	6.3(2.7)	3.3(4.7)	11.3(6.3)	13.7(5.0)	8.3(6.6)			x
4	8.6	.38	.00*	11.0(4.1)	4.7(4.5)	11.04(4.4)	11.06(3.2)	8.1(4.7)	x		
5	3.03	.18	.03*	2.0(7.3)	3.2(5.7)	7.2(5.1)	5.5(5.9)	6.2(6.4)			

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7D2:

Results of repeated measures ANOVAs for /k/ in the control group. The shaded blocks regard pre-post EST-assessments, where dotted lines concern the control condition without EST.

Contr.	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	Post-hoc con pre-post EST compa				
Group (n=3)	$F_{4,176}$								T1-T2	T1-T3	T1-T4	T1-T5	T2-T3
	3.84	.08	.01*	7.9(6.3)	6.1(6.7)	6.4(5.3)	5.1(7.5)	9.0(7.4)					
Indiv.	$F_{4,56}$												
6	1.73	.43	.00*	4.1(2.1)	1.8(1.6)	2.8(2.0)	1.7(2.6)	6.0(2.7)	x				
7	.89	.06	.48	10.7(6.8)	12.4(5.4)	9.8(4.9)	8.9(7.4)	11.4(6.5)					
8	1.85	.12	.13	8.7(6.9)	4.1(6.7)	6.7(6.0)	4.6(9.4)	9.5(10.3)					

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7D3:

Results of repeated measures ANOVAs for /t/ in the experimental group. The shaded blocks regard pre-post EST-

Experi-mentals	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	pre-po		
Group (n=5)	$F_{4,176}$								T1-T2	T1-T3	T1-T4
	8.29	.04	.02*	7.0(6.0)	5.5(5.9)	8.1(6.3)	8.0(5.7)	6.9(8.6)			
Indiv.	$F_{4,56}$										
1	0.25	.02	.91	5.6(5.6)	4.2(4.8)	4.5(6.4)	6.0(5.4)	4.6(7.0)			
2	3.30	.19	.02*	4.8(2.1)	8.3(3.2)	5.6(2.3)	4.4(2.6)	8.7(9.3)	x		
3	9.14	.40	.00*	10.0(6.6)	4.8(3.7)	14.6(6.3)	13.6(3.5)	10.3(7.8)	x		
4	2.21	.14	.08	7.7(4.3)	5.7(5.3)	8.8(5.9)	9.5(3.9)	9.3(4.6)			
5	2.06	.13	.01*	6.8(8.5)	4.3(9.8)	6.9(4.7)	6.4(6.9)	1.6(10.6)			

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7D4:

Results of repeated measures ANOVAs for /t/ in the control group. The shaded blocks regard pre-post EST-assessments, where dotted lines concern the control condition without EST.

Contr.	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	Post-hoc co… pre-post EST compa…				
Group (n=3)	$F_{4,176}$								T1-T2	T1-T3	T1-T4	T1-T5	T2-T3
	1.85	.04	.12	6.5(5.6)	5.1(6.0)	6.3(5.5)	5.8(5.9)	7.4(5.6)					
Indiv.	$F_{4,56}$												
6	1.66	.11	.17	3.3(3.3)	3.5(2.8)	3.4(2.6)	3.2(3.2)	5.4(4.2)					
7	358	.03	84	9.7(5.9)	8.8(5.9)	10.6(5.2)	9.9(6.6)	9.0(5.3)					
8	2.28	.14	.07	6.4(5.6)	3.1(6.9)	4.9(5.6)	4.3(5.4)	7.7(6.6)					

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7D5:

Results of repeated measures ANOVAs for /p/ in the experimental group. The shaded blocks regard pre-post EST-

Experimentals	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	pre-po		
Group (n=5)	$F_{4,176}$								T1-T2	T1-T3	T1-T4
	11.44	.13	.00*	4.6(4.9)	2.7(4.9)	6.1(5.7)	6.7(5.1)	6.3(4.2)	x		x
Indiv.	$F_{4,56}$										
1	2.16	.13	.09	4.3(3.1)	3.1(5.2)	3.9(4.4)	5.9(5.1)	7.1(4.1)			
2	.28	.02	.89	4.5(2.8)	5.1(2.5)	4.7(3.6)	4.1(1.7)	4.7(2.6)			
3	6.74	.33	.00*	5.9(3.7)	2.3(2.7)	10.0(8.2)	10.2(5.5)	7.7(4.4)	x		
4	4.49	.24	.00*	5.2(3.1)	2.4(6.6)	7.6(5.4)	8.2(3.4)	7.5(4.1)			
5	2.87	.17	.03*	3.0(4.5)	.36(5.3)	4.5(3.8)	5.0(6.2)	4.6(4.7)			

Table 7D6:

Results of repeated measures ANOVAs for /p/ in the control group. The shaded blocks regard pre-post EST-assessments, where… dotted lines concern the control condition without EST.

Contr.	$F_{df1,df2}$	η_p^2	p	T1	T2	T3	T4	T5	Post-hoc com…					
									pre-post EST compar…					
Group (n=3)	$F_{4,176}$								T1-T2	T1-T3	T1-T4	T1-T5	T2-T3	T…
	1.07	.02	.37	4.6(3.5)	2.7(4.9)	6.1(5.7)	6.7(5.1)	6.3(4.2)						
Indiv.	$F_{4,56}$													
6	97	07	43	3.5(3.4)	3.6(5.2)	3.0(3.1)	2.1(3.2)	1.4(4.4)						
7	4.60	25	00*	4.1(2.7)	7.8(4.3)	8.1(4.8)	8.1(4.1)	5.9(3.3)		x	x			
8	.51	.04	.73	6.1(5.0)	5.2(9.7)	3.7(6.0)	4.8(4.3)	3.9(3.8)						

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Appendix 7E

Results of repeated measures ANOVAs dor SUS duration

Table 7E

The results of the repeated measures ANOVAs (T1 to T5) of SUS duration (ms) for the experimental group (n=5) and the control group (n=3). Standard deviations are reported between brackets. The means were based on 10 SUS, produced by all participants in both groups. The shaded cells represent the pre-post EST assessments in each group.

	T1	T2	T3	T4	T5	ηp^2	$F_{df1,df2}$	p
Exp.	2253 (601)	2337 (683)	2212 (552)	2298 (732)	2408 (773)	.07	$F_{4,196}=3.8$	.005*
Control	2031 (604)	2001 (607)	1993 (567)	2056 (693)	1887 (487)	.053	$F_{4,104}=1.49$	.224

* = *significant* ($p < 0.05$)

Appendix 7F

Illustration of two different settings of lung activity and the corresponding intensity in the intend voiceless plosive /p/.

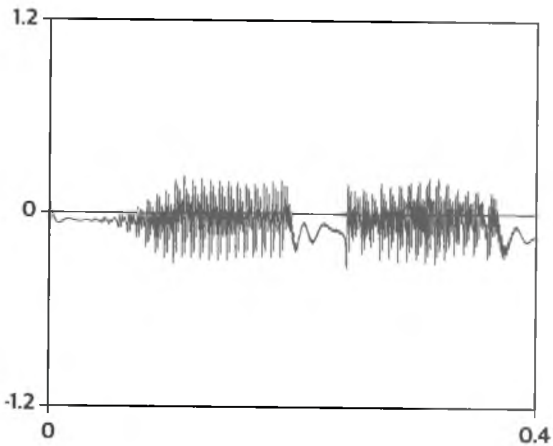


Figure F.7a:

Result of articulatory synthesis of /↔p↔/ with PRAAT: parameter for lung pressure set at 0.1 for 100 ms.

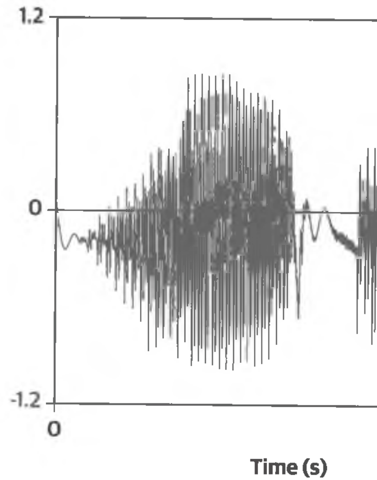


Figure F.7b:

Result of articulatory synthesis of /↔p↔ parameter for lung pressure set at 0.3 fo

Appendix 7G

Results of repeated measures ANOVAs for the standard deviation of F0

Table 7G1:

Results of repeated measures ANOVAs for the standard deviation of F0 (Hz) at T1 toT5 in the experimental grou brackets. Significant differences between assessments according to post-hoc comparisons (Sidak adjusted) are d acoustic measurements in 10 SUS per participant. The shaded blocks regard pre-post EST assessments.

Experi-mentals	$F_{df1,df2}$	η^2	p	T1	T2	T3	T4	T5	pre-po		
Group (n=5)	$F_{4,196}$								T1-T2	T1-T3	T1-T4
	4.34	08	00*	17.6 (4,8)	17.0 (5.4)	18.5 (5.2)	18.5 (4.6)	19.3 (6.9)			
Indiv.	$F_{4,36}$										
1	4.67	34	*	13.4 (2.6)	11.7 (1.7)	11.0 (2.8)	11.2 (1.3)	14.0 (1.1)			
2	1.49	14		17.2 (2.2)	17.5 (2.1)	18.8 (2.5)	18.7 (2.8)	16.8 (1.4)			
3	5.16	37	*	16.7 (1.8)	15.2 (2.5)	17.6 (2.5)	19.5 (2.5)	16.9 (2.0)			
4	8.52	49	*	15.4 (2.2)	14.7 (3.4)	19.4 (2.5)	19.9 (2.8)	17.2 (1.9)			
5	8.74	49	*	25.1 (3.7)	25.9 (2.3)	25.5 (2.3)	22.9 (2.7)	31.7 (5.0)			

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

Table 7G2:

Results of ANOVAs for repeated measures for the standard deviation of F0 (Hz) at T1 toT5 in the control group (n=3). Standard brackets. Significant differences between assessments according to post-hoc comparisons (Sidak adjusted) are displayed. The acoustic measurements in 10 SUS per participant. The shaded blocks regard pre-post EST assessments, whereas the shaded b concern the control condition without EST.

Contr.	$F_{df1,df2}$	η^2	p	T1	T2	T3	T4	T5	Post-hoc con				
										pre-post EST compar			
Group (n=3)	$F_{4,104}$								T1-T2	T1-T3	T1-T4	T1-T5	T2-T3
	2.79	.01	.03*	18,5 (5,8)	16,6 (3,1)	19,2 (3,7)	19,3 (4,9)	18,9 (4,0)					x
Indiv.	$F_{4,104}$												
6	1.33	.18		20.3 (5.2)	16.7 (2.7)	18.3 (3.7)	17.9 (3.5)	17.3 (5.0)					
7	7.44	.45	*	22.7 (3.5)	18.4 (2.3)	21.8 (2.2)	23.1 (2.3)	20.1 (3.4)					x
8	3.90	.30	*	12.8 (2.6)	14.0 (2.6)	17.3 (3.7)	16.4 (5.4)	19.9 (3.9)				x	

** = significant main effect (p< .05), x = significant differences between assessments(p< .05).*

CHAPTER 8

Exploring dysarthric speakers' appreciation of E-learning based Speech Therapy (EST) compared to face-to-face speech training

Beijer, L.J., Rietveld, A.C.M., & Geurts, A.C.H.

Submitted

Abstract

Purpose: A pilot study was conducted to explore dysarthric speakers' appreciation of E-learning based Speech Therapy (EST).

Procedure: The participants followed a four-weeks intensive, protocolized speech training through EST. After this telepractice period, the participants completed a paired comparisons task, addressing the preference for telepractice relative to face-to-face training on a scale from 0 to 3. A trade-off between the extent of intelligibility improvement and the alleged benefits of telepractice was involved. In addition, the participants filled in a short questionnaire, addressing their satisfaction with EST. Once the experimental training period was completed, it was registered whether the participants requested to continue EST on their own initiative .

Results: Overall, the dysarthric speakers tended to prefer telepractice over face-to-face training. However, when facing the trade off between telepractice and the extent of speech intelligibility improvement, they judged an optimal speech training outcome more important than the type of training. The completed questionnaires pointed out that the participants were satisfied with EST. Improvements for interface and visual feedback on pitch were recommended. Half of the participants spontaneously requested to continue EST after the study.

Conclusion: The positive results warrant further research with more homogeneous groups of dysarthric speakers.

8.1 Introduction

Telehealth applications are currently gaining considerable interest in the field of neurological rehabilitation. They are considered potentially beneficial for patients with motor, cognitive and communicative disability due to acquired neurological diseases such as stroke, Parkinson's disease (PD) and traumatic brain injury (Piron et al., 2006; Mashima & Doarn, 2008; Garcia-Molina et al., 2010; Johansson & Wild, 2011). Given the vital benefits of reducing physical efforts, time and costs that are usually involved with visiting a therapist, it is obvious that telehealth applications are increasingly subjected to usability research. Usability refers to the extent to which a product can be implemented to the satisfaction of specific users to achieve individual goals in an effective and efficient manner, to the satisfaction of these users (Schumacher & Lowry, 2010). In the context of the need to cope with current demographic shifts and the foreseen imbalance between supply and demand of healthcare, usability studies for telerehabilitation in patients with motor, communicative and cognitive disabilities are increasingly being reported (Brennan, Georgeadis, & Baron, 2002; Mashima et al., 2008; Brennan & Barker, 2008; Garcia-Molina et al., 2010; Beijer, Rietveld, Hoskam, Geurts, & de Swart, 2010a; Johansson et al., 2011).

In the Netherlands, the usability of a web-based speech training application for dysarthric speakers with acquired neurological diseases (Beijer et al., 2010a; Beijer et al., 2010b; Beijer, Rietveld, & van Stiphout, 2011) is currently being investigated. This telehealth application, called E-learning based Speech Therapy (EST), provides neurological patients with feedback on their dysarthric speech with the possibility to practice speech in the home environment in order to improve speech quality (i.e. speech intelligibility). Obviously, researchers' main interest lies in the efficacy of innovative devices for health care delivery such as EST. With the technological feasibility being warranted (Beijer et al., 2010a), the primary question to be answered is whether this tele-application affects dysarthric speakers' intelligibility at least as much as obtained by face-to-face training. Therefore, clinical outcome research is currently underway. Thus far, the positive result of a case study, showing improved intelligibility in a patient with Parkinson's disease (PD) after EST (Beijer et al., 2010a) gave rise to a small-sized effect study (n=8) (in preparation).

In addition to the efficacy issue, users' satisfaction should also be investigated, since it is one of the factors that contribute to the usability of a telehealth application (Schumacher et al., 2010). Therefore, Brennan and Barker (2008) have addressed the relevance of taking into account the features of target users when conducting telerehabilitation. That is, factors such as age, education and computer skills, but also potential impairments in these neurological patients may affect their satisfaction. Although studies with large numbers of participants are required for conclusive implications for adjustment of telerehabilitation application such as EST, there are preliminary data supporting the efficacy of EST. In this perspective, we explored the users' appreciation of EST relative to traditional face-to-face speech training in a variety of neurological patients. The following research questions were addressed:

1) How do dysarthric speakers' appreciate EST versus traditional face-to-face speech training?
2) Will dysarthric speakers' request to continue EST on their own initiative?

8.2 Materials and methods

Participants

Eight dysarthric speakers with various underlying neurological diseases (stroke, PD, traumatic brain injury) were included. The participants were recruited by speech therapists who had completed face-to-face training with these dysarthric speakers at least three months before. With regard to PD patients, only participants who had completed face-to-face Pitch Limiting Voice Treatment (PLVT) (Swart de, Willemse, Maassen, & Horstink, 2003), a Dutch adaptation of the Lee Silverman Voice Treatment (Ramig et al., 2001), were included. Dysarthric speakers with additional aphasia or severe cognitive or physical disability were excluded from participation.

Personal factors, such as the extent of computer skills and the degree of mobility limitations, were registered along with the experienced impact of the dysarthria on personal communication and the geographical remoteness from the speech therapist. Participants scored the experienced impact of dysarthria on their communication on a 3 point scale ('large', 'moderate', 'none'). Computer skills were scored on a 5 point scale ('none', 'hardly', 'little', 'considerable', 'a lot') and the degree of mobility limitations on a 3 point scale ('no limitations', 'moderate limitations', 'severe limitations'). In Table 8.1 the demographic data are presented.

Table 8.1:

Demographic data of participants (n=8).

	Age (yrs)			Time p.o. (yrs)	Impact on		Distance to speech therapist	Mobility
1	61	male	Stroke	1	large		≤ 5km	no constraints
2	75	male	PD	3	large		6-10 km	no constraints
3	34	male	TBI	14	moderate	a lot	≤ 5km	no constraints
4	64	male	PD	10	large	hardly	≤ 5km	severe constraints
5	57	male	Stroke	2.5	moderate	little	6-10 km	no constraints
6	73	male	PD	6	moderate		6-10 km	no constraints
7	71	male	PD	20	large	little	≤ 5km	moderate constraints
8	63	male	Stroke	1	moderate	little	≤ 5km	moderate constraints

Note: Part. = participant; TBI = traumatic brain injury; PD = Parkinson's disease

Procedure

Speech training through EST

The participants were equipped with a dedicated laptop (Dell Latitude E5500), a headset (Sennheiser PC 131) and a computer mouse. They received an instructional text provided with screenshots of the training procedure and they followed a test trial to get used to speech training through EST. During the experimental period, they were provided with a four-weeks intensive telepractice program to be carried out in their home environment. The EST training procedure consisted of successively listening to audio examples of target speech, imitating the speech, and comparing the own speech realization with the audio example. Auditory feedback on intensity and pitch was supported by additional visual feedback that was received after uploading the audio files containing the patients' speech (Beijer et al., 2010b). Each training item had to be realized twice, enabling patients to approach the target utterance more closely based on the feedback on the first speech attempt.

During the training period, patients were provided with remote access to each course for nearly one-and-a-half day. The speech pathologist monitored the patients' proceedings, encouraged them to finish the exercises of each course in time, and remotely provided access to the subsequent course in the training program. Thus, patients followed an intensive, protocolized training period during four weeks. The content of the speech training programs was tailored to stroke and PD. That is, for stroke patients the training intervention was directed at both the segmental speech dimension (i.e. articulation) and a suprasegmental speech dimension (i.e. intonation). For PD patients, the training mainly focussed on suprasegmental speech dimensions (i.e. intensity and pitch), according to the PLVT (Swart de et al., 2003). The number of exercises in a course was based on implications for this dimension, obtained in a former case study with a PD patient (Beijer et al., 2010a). The dysarthric speakers with TBI followed the same program as the patients with stroke. Once a week, the speech pathologist contacted the participant by telephone to guide the EST training and to answer questions if needed.

Paired comparisons preference task

Once the training period was completed, users' appreciation of EST versus traditional face-to-face training was investigated by comparing four hypothetical speech training scenarios. The scenarios differed with regard to 1) the type of speech training (web-based versus face-to-face training) and with regard to 2) the hypothesized speech outcome (i.e. 'slightly improved' versus 'strongly improved' speech intelligibility). An overview of the scenarios is presented in Table 2.

Table 8.2:

Overview of the speech training scenarios with type of speech training and hypothesized outcome as factors.

Scenario	Type of speech training	Hypothesized outcome (symbols)
1	Face-to-face (FtfF)	Slightly improved speech intelligibility (+ -)
2	Face-to-face (FtfF)	Strongly improved speech intelligibility (+ +)
3	Web-based (EST)	Slightly improved speech intelligibility (+ -)
4	Web-based (EST)	Strongly improved speech intelligibility (+ +)

The four hypothetical scenarios were presented visually, and were supported by concomitant verbal explanations (Figures 8.1 to 8.4).

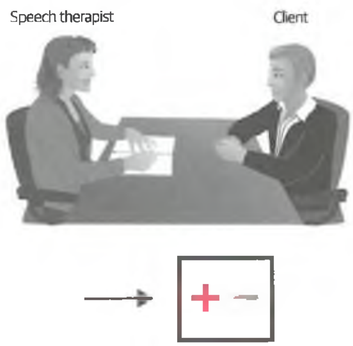


Figure 8.1:

The client visits his speech therapist twice a week. Speech intelligibility slightly improves (Scenario 1).

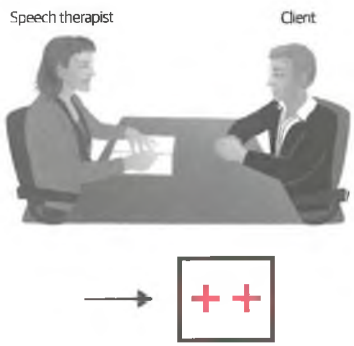


Figure 8. 2:

The client visits his speech therapist twice a week. Speech intelligibility strongly improves (Scenario 2).

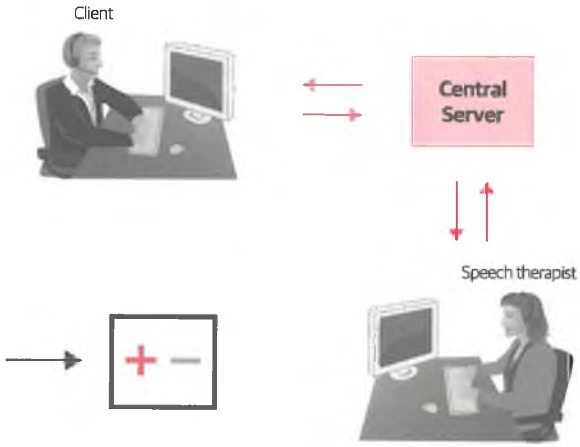


Figure 8.3:

The client practices his speech in his own environment through EST under remote supervision of his speech therapist. Once a month the therapist contacts the client by telephone. Speech intelligibility slightly improves (Scenario 3).

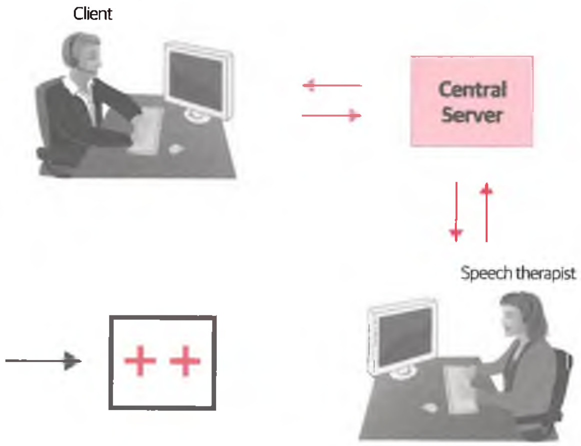


Figure 8.4:

The client practices his speech in his own environment through EST under remote supervision of his speech therapist. Once a month the therapist contacts the client by telephone. Speech intelligibility strongly improves (Scenario 4)..

The depicted speech training scenarios were presented pair-wise to the patients on a laptop screen. Although combination of the four scenarios yields six pairs ((4x3)/2 = 6), only four pairs were presented. The pairs containing the scenarios 3 and 4, and the scenarios 1 and 2 were excluded from comparison since this would boil down to comparison of strongly improved versus slightly improved speech intelligibility under similar speech training conditions. The pairs are presented in Table 8.3 (version A).

Table 8.3:

Combination of the four presented scenario pairs.

Pair	Scenarios (stimulus presented on left side	_	stimulus presented on right side)
2	Scenario 3	_	Scenario 1
3	Scenario 1	_	Scenario 4
4	Scenario 3	_	Scenario 2
5	Scenario 2	_	Scenario 4

The left versus the right location of the scenarios on the laptop screen were varied across the presented pairs in order to keep clients alert. For each pair, clients were asked to indicate their preference for one of the scenarios and also to rate the extent to which they would prefer that particular scenario compared to the alternatives. In the case of perceived equal appreciation (i.e. no preference), clients were asked to assign a '0'. In case of inequality, the clients were asked to assign a scale value to the preferred scenario: '1' (slight preference), '2' (moderate preference) or '3' (strong preference). An example is depicted in Figure 8.5.

Using the software package for stimulus presentation 'E-prime 2.0' (Schneider, Eschman, & Zuccolotto, 2007), the test procedure was conducted on a laptop (Dell Latitude E5500(Schneider, Eschman, & Zuccolotto, 2007). Clients were provided with an instructional text on the laptop screen, followed by two test items. These items contained therapy scenarios that were related to speech training scenarios in general, but differed from the experimental items (e.g. speech training in a group). Using the software package E-prime 2.0, the test procedure was standardized and an e-data file was automatically generated.

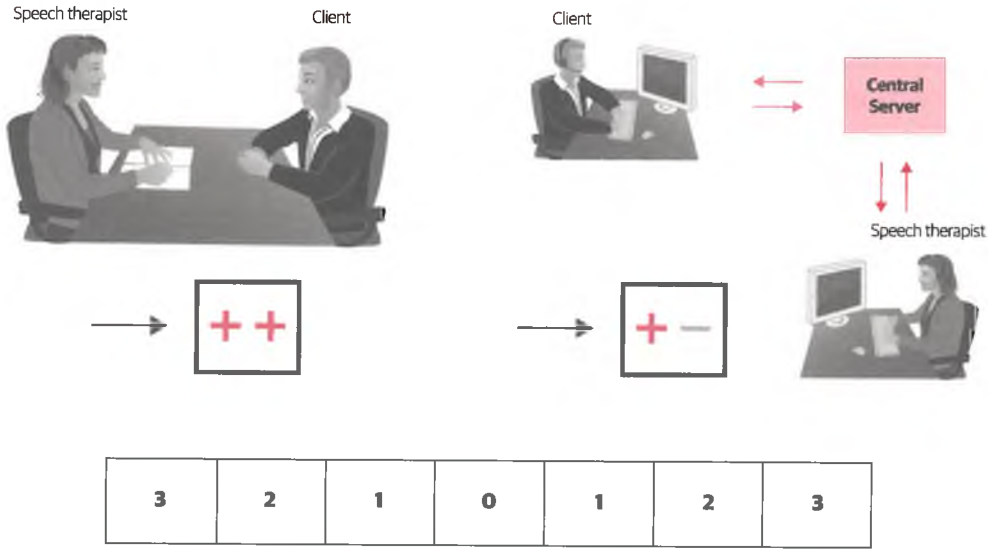


Figure 8.5:
Example of pair-wise depicted speech training scenarios (scenario 2 on the left side, scenario 3 on the right side).

Registration of continuation of EST

After completion of the experimental period, it was registered whether or not the participants asked for continuation of EST on their own initiative.

Questionnaire

Participants were asked to fill in a questionnaire addressing the satisfaction on their experience with EST. They were asked to rate on a 10-point scale, ranging from ‘1’ (extremely unsatisfied) to ‘10’ (extremely satisfied). Ratings below ‘6’ indicated an insufficient level of satisfaction. Dutch people are familiar with this scale since it is commonly used in the Dutch school system. The questions are presented in Table 8.4.

Table 8.4:

Questions to be rated on a 10-point scale.

	Questions
1	How satisfied are you with the user interface of EST?
2	How satisfied are you with the ease of use of EST?
3	How satisfied are you with the attractiveness of EST?
4	How is your overall satisfaction with EST as a web-based speech training device?

8.3 Data collection and data analyses

The data were collected from the e-data files that were automatically generated for all participants. For each pair of speech training scenarios, the score for the preferred scenario was registered. For each participant the assigned scale scores for the preferred scenario in a paired comparison task were cumulated. Subsequently, the sum of scale scores per scenario for all participants was calculated. In addition, it was registered whether or not the participants on their own initiative requested to continue EST. The assigned scale scores for the questionnaire were registered by hand.

The scale scores for each scenario were cumulated for all participants, and the mean scale score for each scenario was calculated. Adopting a significance level of 5%, the data were analyzed using the GLM Repeated Measures procedure of SPSS, with 'scenario' as a within-subject factor. In addition, Sidak adjusted p-values for post-hoc comparisons were computed.

8.4 Results

Paired comparisons of preference

The cumulative scale scores per scenario for all participants are presented in Appendix 8A. The mean scores for each scenario for all participants are presented in Figure 8.6.

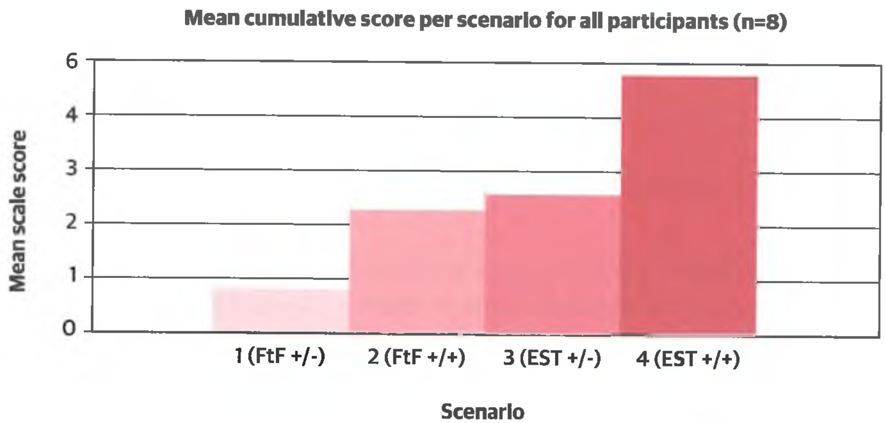


Figure 8.6:

Mean cumulative scores for the scenarios 1, 2, 3 and 4 for all participants. The abbreviations and the symbols between brackets represent the type of speech training and the hypothesized outcome.

Scenario 4 (M=4.75, SD = 2.19) obtained the highest score, subsequently followed by scenario 3 (M=2.50, SD = 1.41), scenario 2 (M=2.25, SD = 1.91) and scenario 1 (M=0.75, SD=2.12). The GLM Repeated Measures procedure showed a significant overall effect of 'scenario' ($F_{1.361,9.528}$ = 4,836 and p = 0.046, Huyn-Feldt adjusted). Post-hoc comparisons yielded a significant difference between scenario 3 and 4 (p=.015), whereas differences between other scenario pairs did not exceed the significance level ($p \geq .05$).

User satisfaction

Overall, the participants were very satisfied with web-based speech training through EST. Table 5 shows the ratings for four questions about user satisfaction.

Table 8.5:

Ratings on a 10-point scale (1 to 10) for four questionnaire about user satisfaction as well as expressed wish to continue EST.

Participant number	Questions				Requested to continue EST
	interface	ease	attractiveness	overall	
1	7	8	8	8	no
2	7	8	7	8	no
3	8	7	8	8	yes
4	7	8	8	9	yes
5	8	8	8	9	yes
6	7	8	8	8	no
7	8	9	9	9	yes
8	8	8	9	9	no
Mean	7.50	8.00	8.13	8.50	

Participants' non-formatted remarks that were added to the scale ratings mainly concerned suggestions for improvement of the interface. Adaptations of font and lay-out were recommended in order to improve visibility and clarity of instructional and training text. The visual feedback on speech intensity was found to be informative, whereas the visual feedback on pitch was found to be confusing. The opportunity to practice in the home environment was very much appreciated.

8.5 Discussion and conclusion

The goal of this study was to compare the appreciation of EST with traditional face-to-face training in a variety of neurological patients. Since we included only eight dysarthric speakers with either stroke, PD or TBI, conclusions should be drawn with caution. Nevertheless, the results of this pilot study indicate that web-based speech training may be a suitable alternative for traditional face-to-face speech training sessions in chronic dysarthric speakers. The outcomes of the paired comparisons task pointed out that web-based speech training was preferred over face-to-face training by the majority of the dysarthric speakers in our study. The scenario of EST resulting in strongly improved speech intelligibility obtained the highest mean cumulative score in our participants, whereas face-to-face training with moderately improved intelligibility obtained the lowest mean cumulative score.

A closer look at the results of the paired comparisons task showed that, when comparing EST and face-to-face training under the condition of equal speech improvement outcomes, seven out of eight participants would prefer EST. Under the condition of unequal speech improvement however, the decision making between EST and face-to-face training in fact involves an interesting trade-off. That is, dysarthric speakers have to decide at what expense they would prefer the benefits of web-based speech training. Given the observed overall preference for web-based training, comparison of scenario 2 (i.e., face-to-face training resulting in strongly improved speech intelligibility) and scenario 3 (i.e., web-based speech training through EST with moderately improved speech intelligibility) provides vital information on this issue. For the seven dysarthric speakers in our study who would prefer EST over face-to-face training, the outcome of the speech training tends to be decisive: five out of seven patients would prefer optimal speech intelligibility outcomes, even if this required face-to-face training. In other words, these patients would prefer web-based speech training over face-to-face training, but not at the cost of the effect of training. The remaining two patients (i.e., participants 4 and 7), however, would categorically prefer EST over face-to-face training even if this implied less training effect. It is noteworthy that these patients had indicated to be severely (participant 4) and moderately (participant 7) constrained in their mobility (Table 8.1). These participants were two out of four dysarthric speakers in our study who requested to continue EST on their own initiative.

The categorical behaviour in participants 4 and 7 is an interesting observation, which in fact was also shown by participant 8 in favour of face-to-face training. This person preferred face-to-face training regardless of speech intelligibility outcome. Apparently, in some patients personnel factors strongly determine their categorical preference for one type of training over the other, irrespective of the efficacy of the training. This implies the relevance of investigating background variables in dysarthric speakers. Insight into these variables, such as the extent of mobility limitations or patients' social environment, possibly leads to the identification of patient profiles that might predict dysarthric speakers' preference for either web-based or face-to-face speech training.

Whether or not participants requested to continue EST on their own initiative, obviously may depend on more variables than the extent of appreciation of EST alone. That is, although patients' request to continue EST on their own initiative implies considerable appreciation of web-based training, not requesting to continue EST does not necessarily indicate patients' low appreciation of EST. These participants might have been satisfied with the speech outcome after the four-weeks-training and therefore decided to complete speech practice for some time. Although the high scale scores on the questionnaire addressing satisfaction with EST are in line with this possibility, socially desirable behaviour might have inflated these scores. This might have been the case with participant 8, who categorically preferred face-to-face training instead of EST, but assigned high scale scores to EST in the questionnaire. It will be clear that the outcomes with regard to participants' initiative to continue EST should be interpreted with care and needs further research. The use of qualitative methods such as interviews should be considered to obtain useful additional information regarding users' appreciation of EST.

The results of this exploratory study provide directions for future studies on the usability of EST. First, larger groups of neurologically impaired dysarthric speakers should be included to better identify personal factors that determine the individual preference for EST versus face-to-face training. Determining factors for both categorical preferences and for trade-off decisions might thus be identified. Second, the nature and the extent of concomitant disabilities should be registered by means of suitable scales. That is, the data should allow conclusions about statistical correlations between functional limitations (e.g., in the motor, cognitive, or communicative domains) on the one hand and the extent of EST user satisfaction on the other hand. Third, a valid and reliable questionnaire should be developed that systematically investigates user satisfaction with distinct dimensions of EST (e.g., technological features, content of speech training, users' conditions). User interviews may add valuable insight into factors that affect users' appreciaton. Thus, along with research outcomes addressing the efficacy of EST (Beijer et al., 2010a), it will be possible to gain more insight into the appreciation of EST as a web-based form of speech training in dysarthric patients.

References

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010a). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) as a web application for speech training in dysarthric patients with Parkinson's Disease: a case study. *Telemedicine and e-Health*, 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., & Geurts, A.C.H. (2010b). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16*, 177-180.

Beijer, L. J., Rietveld, A. C. M., & van Stiphout, A. (2011). Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44*, 701-718.

Brennan, D., Georgeadis, A., & Baron, C. (2002). Telerehabilitation tools for the provision of remote speech-language treatment. *Topics in Stroke Rehabilitation, 8*, 71-78.

Brennan, D. M. & Barker, L. M. (2008). Human factors in the development and implementation of telerehabilitation systems. *Journal of Telemedicine and Telecare, 14*, 55-58.

Garcia-Molina, A., Rodriguez Rajo, P., Sanchez Carrion, R., Gomez Pulido, A., Ensenyat, A., Garcia Rudolph, A. et al. (2010). Clinical program of cognitive tele-rehabilitation for traumatic brain injury. *Proceedings of E-Challenges, Warsaw*, pp. 1-10.

Johansson, T. & Wild, C. (2011). Telerehabilitation in stroke care - a systematic review. *Journal of Telemedicine and Telecare, 17*.

Mashima, P. A. & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health, 14*, 1101-1117.

Piron, L., Tonin, P., Cortese, F., Zampolini, M., Piccione, F., Agostini, M. et al. (2006). Post-stroke arm motor telerehabilitation web-based. *Virtual Rehabilitation*, 145-148.

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry, 71*, 493-498.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). E-Prime, Psychology Software Tools Inc. (Version 2.0) [Computer software].

Schumacher, R. M. & Lowry, S. Z. (2010). *NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records* (Rep. No. NISTIR 7741).

Swart de, B. J. M., Willemse, S. C., Maassen, B. A. M., & Horstink, M. W. I. M. (2003). Improving of voicing in patients with Parkinson's disease by speech therapy. *Neurology, 60*, 498-500.

Appendix 8A

Cumulative scale scores for each scenario per participant.

Participant	Gender	Cumulative scores			
		Scenario 1 FtF +/-	Scenario 2 FtF +/+	Scenario 3 EST +/-	Scenario 4 EST +/+
1	1	0	2	3	6
2	1	0	3	3	6
3	1	0	2	2	6
4	1	0	0	2	3
5	1	0	3	3	6
6	1	0	2	2	5
7	1	0	0	5	6
8	1	6	6	0	0

CHAPTER 9

General Discussion

With the current tendency to promote outpatient healthcare to reduce the costs of inpatient care, telehealth provides alternatives to the conventional modes of outpatient and home visit services. Stakeholders in healthcare show increasing interest in telehealth applications for the rehabilitation of neurological patients. However, the beneficial effects of telerehabilitation should be proven before implementation is warranted.

This thesis addresses initial investigations into the potentials of E-learning based Speech Therapy (EST) in neurological patients with dysarthric speech. The experiments that were conducted are typically Phase II studies or constitute a transition from Phase II to Phase III outcome research. The results provide vital information to adequately prepare larger clinical trials on the effects of EST (Robey, 2004). In the general discussion below, the main findings will be reported along with clinical implications for web-based training. In addition, limitations of the studies and directions for future research will be discussed.

Auditory speech discrimination skills

A key issue in this thesis concerns dysarthric patients' auditory speech discrimination skills as a prerequisite for adequately carrying out the training procedure implemented in EST. In general, it was argued that neurological users' capacities for web-based speech training should be taken into account. For example, sufficient motor abilities are required to operate a keyboard and navigate through the EST system. Furthermore, patients' general physical condition should allow independent and intensive speech practice. In Chapter 4 particular attention was paid to the assessment of auditory speech discrimination abilities, since it is frequently reported that neurological patients show deficits in this respect. Indeed, compared to healthy control subjects, the dysarthric speakers in our study scored lower on the five speech dimensions addressed in the auditory discrimination test (ADT) that was specifically developed for this purpose (Beijer, Rietveld, & van Stiphout, 2011). The lower scores in the dysarthric group reached statistical significance for the subtests 'segmental elements', 'overall pitch', 'speech rate' and 'intonation', but not for 'loudness'. Although these results point into the direction of diminished auditory speech discrimination in neurological patients, they should be interpreted with care. That is, the neurological participants' lower scores may not have been caused by diminished auditory speech discrimination per se. The greater hearing loss in neurological patients appeared to considerably affect the lower scores in this group. In addition, it remains unclear whether genuine hearing loss alone or also general cognitive deficits affect the performance on both the auditory discrimination test and the pure tone hearing test. Although this is an interesting issue, it is questionable whether an in-depth analysis of the underlying causes of diminished auditory speech discrimination in dysarthric speakers is clinically relevant. Whether based on hearing loss, cognitive deficits or impaired auditory discrimination per se, the diminished auditory discrimination observed in neurological patients call for augmented feedback on speech to facilitate web-based speech training. Therefore, visualizing speech may enhance the sensory feedback to dysarthric speakers. It may help neurological patients to estimate whether or not their speech approaches the audio examples provided. Visualizing speech is in line with the emphasis on self monitoring in motor rehabilitation. This is especially useful in treatment for patients with PD, in whom impaired high level executive functions - such as sensorimotor processing and the ability to internally scale, regulate and cue movement parameters - tend to negatively affect speech quality (Ramig et al., 2001).

Although visualization of the suprasegmental speech dimensions 'intensity' (i.e. perceived loudness) and 'fundamental frequency' (i.e. perceived pitch) is implemented in the current version of EST, the type of visual feedback is post hoc, static and abstract. Dysarthric speakers who participated in EST studies indicated that this type of visualization is not appealing and needs improvement (Beijer, Rietveld, Hoskam, Geurts, & de Swart, 2010a). Therefore, the development of psychologically valid and intuitive visualizations of pitch and loudness is currently underway in the VEST (Visualizing E-learning based Speech therapy) project. This project aims at providing information on the type and form of visualizations that are appealing to neurological patients. In addition, given the major contribution of articulation to speech intelligibility (De Bodt, Hernandez-Diaz Huici, & Van De Heyning, 2002; Kim, Kent, & Weismer, 2011), research efforts are being made to provide augmented feedback on the realization of the segmental speech elements in the Dutch language (Strik, van Doremalen, ten Bosch, Rietveld, & Beijer, 2012; Middag, Van Nuffelen, Martens, & De Bodt, 2008; Middag, Martens, Van Nuffelen, & De Bodt, 2009).

Outcome measures for speech intelligibility

In the context of preparing clinical trials for EST effect studies, the suitability of several outcome measures for speech intelligibility was explored. In Chapter 5, orthographic transcription scores (OTS) and intelligibility scale ratings of semantically unpredictable sentences (SUS) were introduced for perceptual assessment of dysarthric speech (Beijer, Clapham, & Rietveld, 2012). Although Hustad already reported the use of SUS for this purpose (Hustad, 2007), these unusual sentences had not yet been developed for the Dutch language. Unlike intelligibility scores on a 10-point scale, OTS of 6-words SUS in declarative form (SUS D6) were found to meet the pre-established requirements for use in clinical effect studies. That is, according to OTS, the different sets of SUS D6 were equivalent with regard to potential speech intelligibility and were also sensitive to variable degrees of intelligibility, regardless of the factor 'rater' (Beijer et al., 2012). Orthographic transcriptions of SUS have the benefits of being both less time-consuming than acoustic measurements and more objective than speech intelligibility scale scores.

The sensitivity of OTS to changes in speech intelligibility was confirmed in the case study reported in Chapter 6, which showed significantly improved OTS of SUS in a dysarthric speaker after EST (Beijer et al., 2010a). However, the explorative study reported in Chapter 7 (Beijer, Rietveld, Ruiter & Geurts, submitted) did not show consistent pre-post EST differences. In this context it should be noticed that, at the time of the case study, the equivalence of the different SUS D6 (Chapter 5) had not yet been established. Since, for this reason, employment of SUS sets in repeated measures was not warranted, a randomized block design for the repeated measures in the case study was employed: 30 SUS per assessment, containing three different syntactic types of ten sentences, were transcribed by a homogeneous group of four listeners. This procedure yielded 120 SUS transcriptions per assessment, compared to 80 in the pilot study reported in Chapter 7 (i.e. 10 SUS transcribed by 8 listeners per assessment). Hence, at the individual level, statistical power in the case study exceeded that in the explorative study, resulting in higher sensitivity to changes in speech intelligibility. However, comparisons of statistical power should be made with care, since the employed SUS in the pilot effect study in Chapter 7 only concerned SUS D6, whereas in the case study (Chapter 6) different syntactic forms and sentence lengths were used. That is, in addition to SUS D6, ten question sentences (SUS Q6)and ten 13-words declarative sentences (SUS D13) were employed. On the other hand, it is possible hat the genuine speech improvement of the participant in the case study exceeded those in the explorative study. Nevertheless, although its sensitivity to between-subjects differences was established in Chapter 5, it is possible that OTS of SUS D6 are not sensitive to subtle within-subject changes in speech intelligibility.

From a linguistic perspective, it could be argued that using SUS as an outcome measure is not ecologically valid. Indeed, the distracting semantic content of these sentences puts demands on listeners that exceed those in common verbal communication. Why then artificially complicating the listeners' task and not benefit of the linguistic context to enhance comprehensibility of dysarthric speakers? A main reason is that intensive speech training through EST focuses on speakers' capacities rather than their daily life performance, the latter being partially determined by listeners' linguistic abilities to comprehend the message. In contrast, speakers' capacities are determined solely by the acoustic signal and should therefore correlate with acoustic measures of speech quality. A disadvantage of using SUS may be the risk of disfluencies in dysarthric speakers, caused by the unpredictable nature of these speech materials. To limit the chance of distorted fluency, we used sentences with only six words in a common (i.e. declarative) syntactic form (SUS D6).

The prominent role of articulation in speech quality is in agreement with the observation that the most consistent outcome of EST was the ΔIDC, being the intensity during the intended silent phase of voiceless plosives relative to the environmental intensity. This measure applies to the segmental speech dimension of voiceless plosives, and appeared to increase after EST. This was an unexpected observation since the intensity during the intended silent phase in voiceless plosives was hypothesized to decrease due to enhanced closure of the vocal tract. This may have been caused by speakers' failure to adjust their articulatory settings along with the increased speech intensity as instructed during the training. As a result, the intended improvement of articulatory precision by intensive training of vocal functions (Ramig et al., 2001; Mahler, Ramig, & Fox, 2009; Sapir, Spielman, Ramig, Story, & Fox, 2007) may not have occurred. Again, visualizing the segmental – in addition to suprasegmental – speech dimensions may enhance dysarthric speakers' awareness of the need to adjust articulatory settings. A benefit of visualizing speech dimensions based on acoustic measurements is the objective information that is obtained on patients' speech. In the long term this could provide valuable information regarding the content of speech training programs.

In the explorative study, judgements of speech continuous speech quality, movements of the first and second formants (i.e. F1 and F2) in the acoustic vowel space and the standard deviations of the fundamental frequency did not show consistent changes over time. A key issue in the suitability of outcome measures of speech intelligibility remains the speech materials employed for perceptual and acoustic assessments. A sound balance should be found between both speakers' capacities (physical and cognitive abilities) and listeners' capacities (e.g. short-term memory for aurally presented speech) on the one hand and criteria for linguistic and phonetic features of the speech materials on the other hand. Thus, it is a true challenge to develop adequate speech materials for outcome studies of EST or other types of speech training.

Appreciation of EST

Apart from patients' capacities, the motivation of patients and speech therapists to use web-based speech training is vital for successful implementation of EST. A study into dysarthric speakers' appreciation of EST (Beijer, Rietveld, & Geurts, 2012) indicated an overall tendency in favour of EST compared to face-to-face training, although some individual patients had a clear preference for either web-based or face-to-face-training, regardless of outcomes. To investigate and optimize patients' motivation for and compliance with EST, background variables such as age, gender, education level, and mobility should be registered to gain insight in the underlying determinants of their preferences. Large numbers of patients are required to attain this goal. With regard to therapists' motivation for and compliance with EST, a change in healthcare culture is needed to warrant a shift from face-to-face towards web-based training. That is, web-based speech training is not simply an alternative way of speech training, but requires a fundamental innovation in healthcare processes with far reaching consequences for both patients and speech therapists. It is possible that the current role of a speech therapist as a physically present trainer will be supplemented with that of a remotely available coach in the near future. Although information and communication technologies will allow a therapeutical relationship, its nature is liable to change due to the physical distance between patients and therapists. In addition, web-based training will increasingly induce the employment of objective outcome measures to establish therapy effects. This may enhance objective assessments, which may lead to adjustments of speech training interventions to enhance positive effects and contribute to improvement of cost-effectiveness of EST.

Elaborating the EST research program

Apart from further investigations into suitable outcome measures for clinical EST effect studies, also the relevance of the research issues in the dashed boxes in Figure 1.4 has become obvious. The development of automatic feedback on pronunciation is believed to support dysarthric speakers' efforts to improve their articulation along with vocal efforts to increase speech intensity. In this context, the Pronunciation Error Detection in Dysarthric Speech (PEDDS) project was recently completed (Strik et al., 2012). In PEDDS, written and graphic feedback on continuous speech is displayed on a computer screen for dysarthric patients who want to improve their articulation by means of a computerized speech training program. To genuinely support dysarthric speakers' training efforts, the visualization of feedback should be appealing. In this perspective, a project aiming at the development of intuitive visualization of intensity and overall-pitch for neurological users is currently underway. With regard to automatic pronunciation assessment for the Dutch language, research efforts are being made in Belgium (Middag et al., 2009; Van Nuffelen, Middag, De Bodt, & Martens, 2009) and in the Netherlands (Doremalen van, Cucchiarini, & Strik, 2009)

Considerations for improvement of EST

Apart from the study of perceptual and acoustic measures of speech intelligibility and the initial exploration of users' appreciation of EST, this thesis yielded indications for improvements regarding the technology and man-computer interaction in EST. These are required to enhance its usability, a concept which refers to the extent to which a product can be used by specified users to achieve goals in an effective, efficient and satisfying manner (Winters & Winters, 2004). The relevance of EST's usability is clear, since it determines its accessibility for various target groups.

Technology

Technological robustness

The feasibility of EST is determined by the robustness of the information and communication technology involved. This robustness is commonly ensured when patients use EST in an unprotected digital environment, such as their home environment. However, when using EST in an inpatient (e.g. hospital) environment, the information exchange will probably be distorted because many clinical institutes impose serious restrictions with regard to network transmissions through the use of an extended 'firewall'. These restrictions may impede exchange of audio files of speech samples with an external server. Therefore, to ensure the feasibility of EST in a clinical setting, the transmission rules for firewalls with respect to e-health applications should (continuously) be adjusted. Rehabilitation centers should put this issue on the agenda in times of an emerging need of e-health.

Software

In experimental EST trials, the dysarthric speakers use dedicated laptops with components of Adobe shockwave and Adobe flash player, since these software packages are required to apply the EST program (Beijer et al., 2010b). The laptops are exclusively utilized for the EST experiments, thus avoiding the potential interference by other software. The implementation of EST in daily practice, however, would require the use of patients' private laptops or desktop computers which tend to be used for numerous other purposes. According to our experience EST is at risk of disruptions caused by interfering software packages installed on the same system. From this perspective, the current EST program should be upgraded to enhance robustness.

Helpdesk

A helpdesk should become available to monitor the functionality of EST, to detect disturbances, and to support patients who experience practical difficulties when using EST. Only with the presence of an online helpdesk, the technological feasibility of EST can be warranted. A helpdesk might also play a crucial role in maintaining patients' and therapists' compliance with EST in the long term.

Data base of dysarthric speech

To optimize the benefits of automatically generating a data base of dysarthric speech by uploading speech samples, additional functionality should be developed. That is, the current system allows therapists only to search speech audio files of their own patients. Although this facilitates the therapist in comparing an individual's speech over time (e.g., before and after EST), the potential benefits of a large data base of audio speech samples cannot be utilized without orthographic, linguistic and phonetic annotations being (anonymously) made available to all users (Beijer & Rietveld, 2011). Only then, well described queries for specific speech characteristics will yield significant results for fundamental and applied research in the field of speech pathology, speech technology, and speech rehabilitation (Beijer et al., 2010b). The data base should allow a systematic search of specific speech characteristics among well defined groups of dysarthric speakers. Given the enormous amount of work needed to build such a corpus of dysarthric speech, substantial funding must be found to support this work. A well accessible and large data base of dysarthric speech is vital for the development of human language applications for patients with communicative disabilities (Ruiter et al., 2012) as well as for domotic applications based on automatic speech recognition (e.g. remote domotic control for neurological patients with severe physical disabilities). In fact, with the generally complicated recruitment of large neurological patient populations - due to their physical disabilities - in mind (Winters, 2002), the possibility of remote data collection through EST will advance rehabilitation research.

Man-technology interaction

The man-technology interface plays a crucial role in the interaction between computers and users. The focus for optimizing man-technology interfaces is the study of human performance and product usability. Since EST was primarily developed for neurological patients, it should be well adjusted to the possibilities of this target group (Hoppestad, 2006). Brennan and Barker (2004) already emphasized to take into account the functional limitations due to neurological diseases when developing innovative technologies for specific patient groups. In the case of EST, even dysarthric speakers with various concomitant disabilities must be able to proceed with the web-based speech training. It is, therefore, vital that a customized user interface will be designed to facilitate a diversity of patients in the successful use of EST (Winters, 2002). Until now, visual feedback was added to the aural feedback to compensate for the diminished auditory speech discrimination skills in neurological patients (Beijer et al., 2011). Based on patients' experiences in the EST experiments and on a case-study (Beijer et al., 2010a), additional improvements have been proposed with regard to an adjustable letter font, a (minimal) size of the computer screen, and an alternative type of visual feedback (i.e., comprehensible graphic form). As mentioned in Chapter 7, psychologically valid and intuitive forms of speech visualization are currently developed in the VEST project, which is a 'spin-off' project of EST.

Conclusions

Web-based speech training through EST is promising given its potential benefits. The enhanced accessibility allows intensification and prolongation of speech training in the home environment of dysarthric speakers, which is alleged to yield positive results. Thus, EST may contribute to solving the increasing healthcare capacity problem and to optimal cost-effectiveness ratio.

This thesis aimed at detecting improvement or maintenance of speech intelligibility by means of EST in chronic dysarthric patients. The suitability and sensitivity of outcome measures were investigated along with prerequisites for web-based speech training. Auditory speech discrimination skills in dysarthric patients were diminished compared to healthy controls. Therefore, augmentative visualization of speech may support the auditory speech discrimination task in the EST procedure. The outcome measures for speech intelligibility need further research. Although orthographic transcription scores of SUS were sensitive to *between-subjects* differences of intelligibility in dysarthric speakers, subtle *within-subjects* changes were not detected in the exploratory EST effect study. Only ΔIDC in voiceless plosives showed consistent results, implying EST effects on speech. Future research should address the enhancement of sensitivity of perceptual and acoustic outcome measures. To this end, adequate types and larger amounts of speech materials are required to detect small effect sizes. These arrangements will only yield results if the validity and the reliability of acoustic measures is warranted by improving the quality of web-based audio recordings. With regard to users' valuation of web-based speech training, our initial explorations suggest good appreciation of EST after – or as an adjunct to – traditional face-to-face therapy in our chronic dysarthric participants.

Once the beneficial effects of EST have been established, practical obstacles should be overcome to meet the requirements for successful implementation of EST in daily healthcare. Apart from improving augmented (e.g. visual) speech feedback, the applied technology should be sufficiently robust to guarantee continuous availability of EST regardless of location. In addition, the man-technology interface should be adjusted in order to 'tune' the application to the neurological target group.

References

Beijer, L. & Rietveld, T. (2011). Potentials of Telehealth Devices for Speech Therapy in Parkinson's Disease. in *Parkinson's Disease, Book 6,* Chapter 18, 379-402. InTech Open Access Publisher.

Beijer, L. J., Clapham, R. P., & Rietveld, A. C. M. (2012). Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20,* 17-34.

Beijer, L. J., Rietveld, A. C. M., & Geurts, A. C. H. (2012). Exploring dysarthric speakers' appreciation of E-learning based Speech Therapy (EST) relative to face-to-face speech training. *International Journal of Telemedicine and Applications.*

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010a). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) as a web application for speech training in dysarthric patients with Parkinson's Disease: a case study. *Telemedicine and eHealth,* 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., & Geurts, A.C.H. (2010b). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and eHealth, 16,* 177-180.

Beijer, L. J., Rietveld, A. C. M., & van Stiphout, A. (2011). Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44,* 701-718.

De Bodt, M. S., Hernandez-Diaz Huici, M. E., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders, 35,* 283-292.

Doremalen van, J., Cucchiarini, C., & Strik, H. (2009). Automatic detection of vowel pronunciation errors using multiple information sources. In *ASRU, IEEE* (pp. 580-585).

Hoppestad, B. S. (2006). Essential elements for assessment of persons with severe neurological impairments for computer access utilizing assistive technology devices: A Delphi Study. *Disability and Rehabilitation: Assistive Technology, 1,* 3-16.

Hustad, K. C. (2007). Contribution of two sources of listener knowledge to intelligibility of speakers with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 50,* 1228-1240.

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54,* 417-429.

Mahler, L. A., Ramig, L. O., & Fox, C. (2009). Intensive voice treatment for dysarthria secondary to stroke. *Journal of Medical Speech-Language Pathology.*

Middag, C., Martens, J. P., Van Nuffelen, G., & De Bodt, M. (2009). Automated intelligibility assessment of pathological speech using phonological features. *EURASIP Journal on Advances in Signal Processing.*

Middag, C., Van Nuffelen, G., Martens, J. P., & De Bodt, M. (2008). Objective intelligibility assessment of pathological speakers. In *Interspeech 2008* (pp. 1745-1748).

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M. et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry, 71,* 493-498.

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37,* 401-411.

Ruiter, M. B., Beijer, L. J., Cucchiarini, C., Krahmer, E. J., Rietveld, T. C. M., Strik, H. et al. (2012). Human language technology and communicative disabilities: requirements and possibilities for the future. *Language Resources and Evaluation, 46,* 143-151.

Sapir, S., Spielman, J., Ramig, L. O., Story, B., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT] on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *Journal of Speech, Language and Hearing Research, 50,* 899-912.

Strik, H., van Doremalen, J., ten Bosch, L., Rietveld, A., & Beijer, L. (2012). Pronunciation Error Detection for Dysarthric Speech. Available at: http://www.youtube.com/watch?v=Bm27GEQIGXc

Van Nuffelen, G., Middag, C., De Bodt, M., & Martens, J. P. (2009). Speech-technology-based assessment of phoneme intelligibility in dysarthria. *International journal of Language and Communication Disorders, 44,* 716-730.

Winters, J. M. (2002). Telerehabilitation research: emerging opportunities. *Annual Review of Biomedical Engineering, 4,* 287-320.

Winters, J. M. & Winters, J. M. (2004). A telehomecare model for optimizing rehabilitation outcomes. *Telemedicine and e-Health, 10,* 200-2.

SUMMARY

Samenvatting

Summary

This thesis explores the potentials of e-health for adult neurological patients with acquired dysarthric speech. The general introduction in **Chapter 1** sheds some light on the background of increasingly emerging telehealth applications. Due to the current demographic developments, the group of chronic neurological patients who experience limitations in their daily life activities is rapidly growing. Along with the decreasing therapeutic and financial resources these developments cause a healthcare capacity problem, which calls for innovations in healthcare. Telerehabilitation might contribute to a solution of this need-and-capacity problem. In this context a web application for speech training in neurological patients with dysarthric speech is introduced: E-learning based Speech Therapy (EST). The EST research program and the research questions addressed in this dissertation are presented.

In **Chapter 2**, potentials of telehealth devices for patients with neurological diseases, particularly for patients with Parkinson's disease, are described. Attention is paid to the technical requirements for telehealth and its benefits for both therapeutic purposes and scientific research in the field of speech pathology and speech technology. In this perspective, the possibility of collecting speech samples for a data base of dysarthric speech by means of EST is highlighted. The potentials of speech technology for pronunciation training are elaborated.

In **Chapter 3** EST is presented. The infrastructure, the application requirements, and the architecture of EST training programs are elucidated. In addition, the speech training procedure, the target users and the potentials for therapeutic and research purposes are described in more detail.

Chapter 4 focuses on the assessment of auditory speech discrimination skills in dysarthric speakers. The development and the evaluation of an Auditory Discrimination Test (ADT) are described. The test contains five subtests that each concern one speech dimension which is addressed in EST: segmental speech elements (articulation), intensity (loudness), overall-pitch, speech rate and intonation. Test items that met the predetermined criteria for sensitivity to diminished auditory discrimination were used to compare 14 dysarthric speakers and 14 control subjects who were matched for age, gender and education. The outcomes support the hypothesized diminished auditory speech discrimination skills in dysarthric speakers and hence, call for augmented (e.g. visual) feedback on speech in the training procedure implemented in EST.

Chapter 5 addresses the suitability of orthographic transcription scores and intelligibility scale ratings of semantically unpredictable sentences (SUS) as a perceptual outcome measure for intelligibility in dysarthric speech. Five sets of SUS were developed. Each set contained ten declarative 6-word sentences (SUS D6), ten declarative 13-word sentences (SUS D13) and ten 6-word question sentences (SUS Q6). Five speakers with PD and different degrees of dysarthria read aloud the sets. Ten trained listeners with sufficient interrater reliability were provided with the sentences realized by the five speakers in an orthogonal array (i.e. latin square design) to avoid learning effects, and conducted the orthographic transcription task and the scale ratings task. Only orthographic transcription scores of declarative 6-words SUS met the pre-established criteria for employment in clinical outcome research: no main effects of the factors 'SUS set' or 'listener' were found, but the factor 'speaker' did affect the orthographic transcription scores.

In **Chapter 6** a case study of a male patient with PD is reported. Since the study reported in Chapter 5 had not been completed at the time of the case study, intelligibility scale ratings and orthographic transcription scores of 30 SUS (10 D6, 10 D13, 10Q6) were used as outcome measures in an adjusted design. A repeated measures randomized blocks design was used with four homogeneous blocks of listeners. The results pointed out a significantly increased orthographic transcription score after EST, whereas the scale ratings did not show this pattern. This latter observation was explained by the lack of 'anchor points' for intelligibility in scale ratings, causing a somewhat poor interrater reliability. The participant filled in a questionnaire about the level of satisfaction they experienced about the technological features of EST, about the content of the training program, and about the appreciation of a home-based training environment.

In **Chapter 7** an exploratory EST effect study is reported that typically fits into a transition from Phase II to Phase III clinical outcome research. In preparation of an efficacy study with a larger number of dysarthric speakers, the suitability of various perceptual and acoustic outcome measures for speech intelligibility in clinical outcome research was investigated. Perceptual measures were general judgements of speech quality and orthographic transcription scores of SUS. Acoustic measures concerned the intensity in the intended silent phase of voiceless plosives in unstressed syllables (ΔIDC), the standard deviation of the fundamental frequency in SUS, movements of the unaccented vowels /a/, /e/ and /o/ in the articulatory-acoustic vowel space as defined by the first and second formants in vowels (F1-F2). Consistent results were found only for ΔIDC. This measure was affected by EST, albeit in the opposite direction than expected. In general, the outcomes shed some light on the complexity of employing outcome measures to detect subtle changes in dysarthric speech. The home environment for audio recordings of speech, the type and amount of speech materials used for speech assessments as well as the variability between and within dysarthric speakers were considered factors that should be addressed in future effect studies. On the other hand, the relevance of acoustic measures in speech quality assessment was underscored. That is, the analyses of the acoustic speech signal provides vital information on the way how training may affect speech and, thus, provide indications for adjustments of speech training programs.

Chapter 8 addresses a pilot study (n=8) of dysarthric speakers' appreciation of EST as a web-application for speech training in a home environment. Neurological participants assigned scale ratings of preference to pairwise (visually) presented speech training scenarios. The distinctive scenarios were based on different manners of speech training delivery (i.e. face-to-face versus EST) and different outcomes of the training (i.e. moderate versus strong improvement of speech quality). Overall, the participants tended to prefer web-based training through EST over face-to-face training. However, when facing the trade-off between telepractice and the extent of speech intelligibility improvement, they judged an optimal speech training outcome more important than the type of training. Still, one participant categorically preferred face-to-face-training, regardless of the results of the speech training. This finding supports the importance of taking into account background variables such as age, education level and mobility in future research addressing users' appreciation of EST. These factors are believed to play a vital role in the preferences of dysarthric speakers for a particular type of speech training. The participants also suggested improvements with regard to the interface used in EST.

Chapter 9 contains the general discussion of this thesis. Attention is paid to the auditory discrimination skills in dysarthric speakers as a prerequisite for adequate use of EST. The role of visual feedback on speech in the EST training procedure to support diminished auditory discrimination is discussed. Another issue for consideration concern the sensitivity of outcome measures for detecting subtle within-subjects changes in speech intelligibility. In this context, environmental conditions for speech audio recordings, the type and the amount of speech materials for speech quality assessments are discussed and provide directions for future EST outcome research. Attention is paid to the relevance of acoustic measures in EST effect studies for adjustments of speech training programs. Finally, suggestions for improvement of EST are provided, and prerequisites for successful implementation of web-based speech training in healthcare are discussed.

Some critical points are made with regard to the web-based generation of a data base of dysarthric speech for the purpose of scientific research in the field of speech pathology and speech technology.

Samenvatting

In dit proefschrift worden de mogelijkheden geëxploreerd van e-health voor volwassen patiënten met dysartrische spraak als gevolg van verworven neurologisch letsel. De algemene introductie in **Hoofdstuk 1** schetst de achtergond van de opkomst van telehealth applicaties. Als gevolg van de actuele demografische ontwikkelingen, groeit de groep chronische neurologische patiënten die beperkingen ervaren in hun dagelijkse leven in hoog tempo. In combinatie met de kleiner wordende groep therapeuten en de afnemende financiële middelen leidt dit tot een capaciteitsprobleem, dat om innovatie in de gezondheidszorg vraagt. Telerevalidatie biedt mogelijk een oplossing voor dit vraag-en-aanbod probleem. In deze context wordt een webapplicatie voor spraaktraining geïntroduceerd voor neurologische patiënten met dysartrische spraak: E-learning gestuurde SpraakTherapie (EST).

In **Hoofdstuk 2** worden de potenties van telehealth voor patiënten met neurologisch letsel, in het bijzonder Parkinsonpatiënten, beschreven. Er wordt aandacht besteed aan technische eisen voor telehealth en de voordelen voor zowel therapeutische doeleinden als wetenschappelijk onderzoek op het gebied van spraakpathologie en spraaktechnologie. In dit perspectief worden de mogelijkheden belicht voor verzameling van spraaksamples ten behoeve van een databank van dysartrische spraak door middel van EST. De potenties van spraaktechnologie voor uitspraaktraining worden verder uitgediept.

In **Hoofdstuk 3** wordt EST gepresenteerd. De infrastructuur, de applicatie-eisen en de architectuur van het EST trainingsprogramma worden toegelicht. Verder worden de spraaktrainingsprocedure, de gebruikersdoelgroepen en de potenties voor therapeutische en onderzoeksdoeleinden in meer detail beschreven.

Hoofdstuk 4 richt zich op de beoordeling van het vermogen tot auditieve spraakdiscriminatie bij dysartrische sprekers. De ontwikkeling en de evaluatie van een Auditieve Discriminatie Test (ADT) worden beschreven. De test bevat vijf subtests, die ieder een spraakdimensie betreffen die in EST aan de orde komt: segmentele spraakelementen (articulatie), intensiteit (luidheid), toonhoogte, spreektempo en intonatie. Testitems die voldeden aan de vooraf vastgestelde criteria voor sensitiviteit voor verminderde auditieve discriminatie werden gebruikt bij de vergelijking van 14 dysartrische sprekers met 14 gezonde controlepersonen die gematcht waren op leeftijd, geslacht en opleidingsniveau. De uitkomsten ondersteunen de hypothese dat dysartrische sprekers een verminderd vermogen tot auditieve spraakdiscriminatie hebben, en onderbouwen aldus de noodzaak voor ondersteunende (visuele) feedback over spraak in de trainingsprocedure in EST.

In **Hoofdstuk 5** wordt de geschiktheid van orthografische transcriptiescores en schaalscores voor verstaanbaarheid van semantisch onvoorspelbare zinnen ('semantically unpredictable sentences : SUS) als perceptieve uitkomstmaat voor verstaanbaarheid in dysartrische spraak aan de orde gesteld. Vijf SUS sets werden ontwikkeld. Elke set bevatte tien declaratieve 6-woordzinnen (SUS D6), tien declaratieve 13-woordzinnen (SUS D13) en tien vraagzinnen bestaande uit 6 woorden (SUS Q6). Vijf sprekers met PD en verschillende graden van dysartrie lazen de sets hardop. De door de sprekers gerealiseerde zinnen werden auditief aangeboden aan tien getrainde luisteraars met voldoende interbeoordeelaarsbetrouwbaarheid. Dit gebeurde volgens een latin square design om leereffecten te voorkomen. De luisteraars voerden de orthografische transcriptaak uit en scoorden alle zinnen op een 10-puntschaal. Uitsluitend de declaratieve 6-woordzinnen (SUS D6) voldeden aan de vooraf vastgestelde criteria voor gebruik ten behoeve van klinische effectstudies: er werden geen hoofdeffecten gevonden voor de factoren 'SUS-set' of 'luisteraar', terwijl de factor 'spreker' van invloed was op de orthografische transcriptiescores.

In **Hoofdstuk 6** wordt een case study beschreven van een mannelijk patiënt met de ziekte van Parkinson. Aangezien de studie in Hoofdstuk 5 nog niet afgerond was ten tijde van deze studie, werden de schaalscores voor verstaanbaarheid en de orthografische transcriptietaak voor 30 SUS (10 D6, 10 D13, 10 Q6) gebruikt in een aangepast design: een repeated masures gerandomiseerd block design met vier homogene blocks van luisteraars. De resultaten lieten een significante toename van orthografische transcriptiescore zien na EST, terwijl dit voor de schaalscores niet het geval was. Mogelijk kan deze laatste bevinding verklaard worden door het ontbreken van ankerpunten voor schaalcores met mogelijk een lagere interbeoordeelaarsbetrouwbaarheid als gevolg. De patiënt vulde een vragenlijst in met betrekking tot de tevredenheid over de technologische kenmerken van EST, de inhoud van het trainingsprogramma en over de waardering van oefenen in de eigen leefomgeving.

In **Hoofdstuk 7** wordt verslag gedaan van een explorerend EST-effectonderzoek dat past in de overgang van Fase II naar Fase III in klinische effectstudies. In voorbereiding op een 'efficacy-study' met een groter aantal dysartrische sprekers, wordt de geschiktheid onderzocht van enkele perceptieve en akoestische uitkomstmaten voor verstaanbaarheid van spraak in klinische effectstudies. Perceptieve maten waren algemene oordelen van spraakkwaliteit en orthografisch transcriptiescores van SUS. Akoestische maten betroffen de intensiteit in de beoogd stille fase van stemloze plosieven in onbeklemtoonde syllaben

(ΔIDC), de standaarddeviatie van de grondfrekwentie (F0) in SUS, beweging van de onbeklemtoonde klinkers /a/, /e/ en /o/ in de articulatorische-akoestische klinkerruimte, zoals gedefinieerd door de eerste en tweede klinkerformanten (F1 en F2). Consistente resultaten werden uitsluitend gevonden voor ΔIDC. De waarde van deze uitkomstmaat na het volgen van EST nam toe en niet - zoals verwacht - af . In het algemeen boden de uitkomsten van deze studie enig zicht op de complexiteit van het gebruik van uitkomstmaten voor de detectie van subtiel veranderingen in dysartrsiche spraak. De thuisomgeving voor audiorecordings van spraak, het type en de hoeveelheid van spraakmateriaal dat gebruikt werd voor de assessments van spraak, en de variabiliteit tussen en binnen dysartrische sprekers werden beschouwd als factoren waaraan aandacht moet worden besteed in toekomstige effectstudies. Tevens werd het belang van akoestische maten onderstreept. De analyse van het akoestische spraaksignaal biedt immers vitale informatie over de wijze waarop training de spraak beinvloedt, en geeft indicaties voor aanpassingen in spraaktrainingprogramma's.

Hoofdstuk 8 is gewijd aan een pilot studie (n=8) naar de waardering door dysartrische sprekers van EST als een webapplicatie voor spraaktraining in de thuisomgeving. Neurologische patiënten gaven schaalscores aan hun voorkeur voor paarsgewijs aangeboden afbeeldingen van spraaktrainingsscenario's. De scenario's onderscheidden zich op basis van de wijze waarop spraaktraining werd aangeboden (face-to-face versus EST) en de verschillende resultaten van de training (lichte mate van verbetering versus sterke mate van verbetering). Over het algemeen gaven de deelnemers de voorkeur aan de webgebaseerde versie door middel van EST boven de face-to-face training. Echter, wanneer een keuze moest worden gemaakt tussen de wijze van training en het resultaat, werd het resultaat van de training het belangrijkst gevonden. Een deelnemer koos categorisch voor de face-to-face versie van de spraaktraining, ongeacht het resultaat. Deze bevinding ondersteunt het belang van het betrekken van achtergrondvariabelen zoals leeftijd, opleidingsniveau en mobiliteit in vervolgonderzoek naar de waardering van EST door gebruikers. Deze factoren worden geacht een belangrijke rol te spelen bij de voorkeur van dysartrische sprekers voor een bepaalde vorm van training. De deelnemers aan deze studie deden op verzoek aanbevelingen voor verbeteringen van de interface van EST.

Hoofdstuk 9 bevat de algemene discussie van dit proefschrift. De auditieve discriminatievaardigheden van dysartrsiche sprekers wordt besproken als een voorwaarde voor adequaat gebruik van EST. De rol van visuele feedback op spraak in de EST procedure als ondersteuning voor de verminderde auditieve discriminatie wordt besproken. Een ander punt van discussie betreft de sensitiviteit van uitkomstmaten voor de detectie van subtiele veranderingen in dysartrische spraak. In deze context worden ook de condities voor spraakopnames, het type en de hoeveelheid spraakmateriaal voor beoordelingen van spraakkwaliteit bediscussieerd en geven richting aan toekomstig EST effect onderzoek. Tevens wordt aandacht besteed aan de relevantie van akoestische maten voor aanpassingen van spraaktrainingsprogramma's. Tenslotte worden suggesties gedaan voor verbetering van EST en worden enkele voorwaarden besproken voor succesvolle implementatie van webgebaseerde spraaktraining in de gezondheidszorg. Een kritische blik wordt gegeven op de webgebaseerde generatie van een databank van dysartrische spraak ten behoeve van wetenschappelijk onderzoek op het gebied van spraakpathologie en spraaktechnologie.

List of publications

Beijer, L.J., Clapham, R.P., & Rietveld, A.C.M. (2012). Evaluating the suitability of orthographic transcription and intelligibility scale rating of semantically unpredictable sentences (SUS) for speech training efficacy research in dysarthric speakers with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20*, 2, 17-34.

Beijer, L. and Rietveld, T. (2011). Potentials of Telehealth Devices for Speech Therapy in Parkinson's Disease In: Parkinson's Disease, Book 6, Chapter 18, 379 - 402. ISBN: 979-953-307-247-9.

Ruiter, M.B., Beijer, L.J., Cucchiarini, C., Krahmer, E.J., Rietveld, A.C.M., Strik, H., & Van hamme, H. (2011). Human Language Technology and communicative disabilities: Requirements and possibilities for the future. *Language Resources and Evaluation*. DOI 10.1007/s10579-011-9173-9.

Beijer, L.J., Rietveld, A.C.M. & van Stiphout, A.J. (2011). Auditory discrimination as a condition for E-learning based Speech Therapy: A proposal for an Auditory Discrimination Test (ADT) for adult dysarthric speakers. *Journal of Communication Disorders, 44*, 701-718.

Beijer, L. J., Rietveld, A. C. M., Hoskam, V., Geurts, A. C. H., & de Swart, B. J. M. (2010). Evaluating the feasibility and the potential efficacy of E-learning based Speech Therapy (EST) for dysarthric patients with Parkinson's disease: a case-study. *Telemedicine and e-Health, 16*, 732-738.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., & Geurts, A.C.H. (2010). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine and e-Health, 16*, 177-180.

Dankwoord

Het schrijven van het dankwoord is een prettige bezigheid. Het proefschrift is af en eindelijk kan ik iedereen bedanken die daaraan direct of indirect een bijdrage heeft geleverd.

Allereerst bedank ik mijn eerste promotor, professor dr. Rietveld. Toni, we kennen elkaar al ruim twintig jaar en hebben gedurende de laatste vijf à zes jaar zeer nauw samengewerkt. Dat was een arbeidsintensieve periode, waarin je mij rondleidde in de wereld van de methodologie en de statistiek, zonder de jou vertrouwde motto's zoals 'Het leven is fonetiek' uit het oog te verliezen. We hebben hard gewerkt, maar ook veel plezier gehad. Ik bewonder de wijze waarop je mij op basis van jouw expertise begeleidde bij de opzet en uitvoering van experimenten en bij het schrijven van wetenschappelijke artikelen, waarbij je mij steeds als een bevriende collega bejegende.

Graag bedank ik ook mijn tweede promotor, professor Geurts. Sander, ook jou ken ik al bijna twintig jaar. Onze kennismaking dateert uit mijn begintijd op de Sint Maartenskliniek waar je als revalidatiearts een interdisciplinair team aanstuurde. Toen al hadden de multidisciplinaire overleggen met jou doorgaans een educatief karakter en wist je complexe zaken tot hanteerbare proporties terug te brengen. Dat is een kwaliteit die je ook gedurende mijn promotietraject aan de dag legde, en die mijn gedachtengang hielp structureren.

Toni en Sander, jullie boden mij niet alleen een uitstekende combinatie van wetenschappelijke en vakinhoudelijke expertise. Het persoonlijke contact met jullie beiden was en is voor mij heel waardevol. Jullie toonden jullie betrokken in een bijzonder moeilijke fase van mijn leven, en hebben met mij meegedacht over een zinvolle en haalbare invulling van mijn promotietraject. Voor de praktische uitvoering kreeg ik veel ruimte. Ik heb genoten van het contact met jullie en hoop dat we elkaar in de toekomst nog vaak zullen ontmoeten.

Dit proefschrift is mede tot stand gekomen dankzij de Sint Maartenskliniek, die mij gefaciliteerd heeft om dit onderzoek uit te voeren. Twee mensen wil ik graag in het bijzonder bedanken.

Pétri Holtus, het is wellicht wat ongebruikelijk om je voormalig leidinggevende als paranimf te kiezen. Ik kon echter geen beter symbolisch gebaar bedenken om je te bedanken voor je initiatieven en de geïnspireerde wijze waarop je mij – en vele anderen – faciliteerde om een bijdrage te leveren aan de academisering van de logopedie. Jij realiseert je daarbij terdege dat de logopedische zorg daadwerkelijk de vruchten moet kunnen plukken van academische activiteiten, en hebt altijd inspanningen verricht om de onderwerpen van de onderzoeksactiviteiten te selecteren op basis van klinisch en strategisch belang. Jouw handelen wordt doorgaans ingegeven door een scherpe geest, maar zeker ook door een warm kloppend hart. Dat heb ik gedurende de laatste jaren vaak mogen ervaren.

Jacques van Limbeek, meerdere malen heb je je niet alleen een visionair en een scherp denker, maar ook een warm mens getoond in de voorbije jaren. Dank voor de ruimte die je mij hebt gegeven om dat te kunnen doen in mijn werk wat ik voor ogen had.

De leden van de begeleidingscommissie ben ik zeer erkentelijk voor het mede monitoren en bijsturen van het EST-project. Naast Toni Rietveld, Sander Geurts, Pétri Holtus en Jacques van Limbeek, hebben Arjan van Hessen, Helmer Strik en Bert de Swart een waardevolle bijdrage geleverd tijdens de projectbijeenkomsten.

De medewerkers van het UCI hebben een belangrijke bijdrage geleverd aan het EST-project. Marijn van Beers, jij stond samen met Toni en mij aan de wieg van EST, en samen hebben we het winnen dan de Nationale Zorgvernieuwingsprijs gevierd. Vele sessies hebben we overlegd over een bruikbare versie van EST ten behoeve van het onderzoek. Je geduld was bijzonder groot. Je kwam zoveel mogelijk tegemoet aan de wensen die wij hadden vanuit gebruikersperspectief. Ook jouw collega Robert Slangen heeft daaraan een belangrijke bijdrage geleverd. Dank voor dat alles, want jullie werkzaamheden boden een belangrijke basis voor het EST-project.

Tineke Mulder, Leo van den Broek en Ineke van de Craats, bedankt voor jullie inspanningen tijdens het inspreken van het audiomateriaal voor de experimenten. We waren niet snel tevreden, maar jullie lieten je niet uit het veld slaan.

Joop Kerkhoff, als het gaat om technische ondersteuning bij het bouwen en uitvoeren van experimenten, en dat met veel geduld en optimisme, dan komt jouw naam automatisch boven. Ongelooflijk, dat een mens zo opgeruimd kan blijven terwijl zijn eigen werkzaamheden te pas en te onpas onderbroken worden door vragen van promovendi, studenten en collega's. Hartelijk dank voor alles!

Collega's logopedie van de Sint Maartenskliniek, jullie hebben mij trouw geholpen in deze promotieperiode. Niet alleen door jullie medewerking te verlenen aan luisterexperimenten en aan trials met EST, maar ook door de gezelligheid binnen de vakgroep Logopedie. Ik ben zelf door alle werkzaamheden de laatste tijd een ontrouw bezoeker van de koffie- en lunchpauzes geweest, maar ik voel me nog altijd thuis bij jullie. Ook dank aan de vele collega-logopedisten in de eerste lijn, het UMCN, de ZZG Zorgroep en de HAN, die ondanks de drukke werkzaamheden hun medewerking hebben verleend aan de luisterexperimenten. Jullie belangstelling voor mijn promotieonderzoek was heel motiverend.

Behalve logopedisten, hebben ook veel niet-logopedisten meegewerkt aan het EST-project. Zowel gezonde proefpersonen als proefpersonen met een neurologische aandoening hebben deelgenomen. Ik heb dat enorm gewaardeerd. Met name de mensen die gedurende 20 weken een experimenteel traject met E-learning gestuurde SpraakTherapie (EST) doorlopen hebben, ben ik zeer erkentelijk. Het kan niet vaak genoeg gezegd dat medewerking van deze groepen onmisbaar is voor klinische effectstudies.

Mijn naaste onderzoekscollega's binnen het OSTT, Marina Ruiter en Ans Kilkens, dank voor de 'gesprekken tussendoor' over het wel en wee in het leven van een onderzoeker en over nog veel meer zaken daarbuiten. Ik hoop dat we nog heel lang zullen samenwerken.

Collega's van wat lange tijd de RD&E afdeling van de Maartenskliniek was, dank voor alle gezelligheid en collegialiteit. Bart Nienhuis, je belangstelling voor mijn onderzoek en je inspanningen als ik - vast ook wel eens ongelegen - om technische ondersteuning kwam vragen, heb ik enorm gewaardeerd. Brenda Groen, Petra Heesterbeek en Bianca Nijhuis, jullie 'tips and tricks' bij het vervaardigen van het manuscript van dit proefschrift waren heel waardevol. Ook dank aan Mariëlle Gradussen, Jorien van Beek, Karin Derksen en Lisette Beelen voor de ondersteuning. Eerst een kop koffie pakken bij jullie vóór het begin van de werkdag was altijd gezellig.

De scriptiestudenten en student-assistenten, Anke van Stiphout, Renee Clapham, Vera Hoskam, Nadine de Rue, Gerna Scholte, Sonja Peters, Linda van den Oever, Carlijn Bleumer en Alex Teulings, dank voor jullie meet- en regelwerkzaamheden, en voor het meedenken over de experimenten. Margo van Stiphout, je grafische ondersteuning bij het aanleveren van figuren in het voor publicatie vereiste format in de juiste resolutie betekende voor mij een enorme verlichting in tijden van naderende deadlines. Dank daarvoor!

Gaetano Ambrosino, goede vriend, zonder jou was dit boekje niet zo mooi geworden. Niemand anders had deze omslag kunnen maken. Dank voor je inspanningen, je vriendschap en je geduld bij de grafische vormgeving van dit proefschrift.

Lianne Remijn, dank voor je pragmatische maar o zo lieve steun in zware tijden. Ik bewonder je vermogen om je rollen als collega en vriendin te scheiden, en ik ben heel blij dat je mijn paranimf wilde zijn.

Een promotietraject neemt doorgaans een aantal levensjaren in beslag. Dat vraagt naast professionele support ook om een informeel vangnet gedurende een lange periode. De niet aflatende belangstelling voor mijn werk en de steun van familie en vrienden – veraf en dichtbij –was hartverwarmend. De verleiding is bijzonder groot om namen te noemen, maar de angst om toch enkele intens lieve mensen te vergeten, weerhoudt mij daarvan. Ik hoop dat jullie je aangesproken voelen en weten dat ik enorm waardeer wat jullie voor mijn gezin en mij betekenen.

Enkele persoonlijke woorden wil ik graag richten tot de mensen die mij heel dierbaar zijn.

Door dit promotietraject heb ik mij kunnen ontwikkelen in vele opzichten. Rein en Agnes Beijer, mijn ouders en trouwe fans vanaf het eerste uur, jullie hebben daarvoor al lang geleden de basis gelegd. Dank jullie wel voor de veilige start die jullie mij boden in ons gezin en voor de nooit aflatende aanmoediging om mijzelf te ontwikkelen. Jullie hebben mij altijd gesteund in alles wat ik deed, en waren vooral in de laatste anderhalf jaar een rots in de branding voor mij en de kinderen. Jullie liefde en warmte is onbetaalbaar.

Lieve Huib, deze dag had jij er bij moeten zijn. Samen hadden we hier naar uitgekeken. Altijd heb je mij gestimuleerd en gesteund in dit promotietraject. Nooit was iets jou teveel. Het is nog steeds onvoorstelbaar dat je er niet meer bent, maar in mijn gedachten ben jij altijd bij ons. Samen hebben we onze kinderen mede gevormd tot de mensen die ze zijn: levenslustige en lieve meiden met oog voor de mensen om hen heen. Lieve Merel, Ruth en Pien, ik sta elke dag weer versteld van zoveel kracht en liefde in de jonge mensen die jullie zijn. Ik heb diepe bewondering voor het geduld en het begrip dat jullie in een zo moeilijke periode hebben opgebracht voor je promoverende moeder. Ik ben supertrots op jullie!

Curriculum Vitae

Lilian Beijer werd geboren op 23 maart 1965 in Heerlen. Het grootste deel van haar jeugd bracht zij door in het naburige Landgraaf, waar ze in 1983 haar VWO-diploma behaalde aan het Eijkhagencollege. In 1987 studeerde zij af aan de Opleiding Logopedie van de Hogeschool Zuyd te Heerlen, en vond aansluitend een baan als logopedist in het Speciaal Onderwijs in Haarlem. De drang om grenzen te verleggen won het echter al snel van de zekerheid van een vaste baan, en in mei 1988 werd ze door Stichting Horizon Holland voor een jaar uitgezonden naar het eiland São Miguel op de Azoren. Naast de ervaring die ze daar opdeed in haar adviserende rol inzake logopedische vraagstukken over kinderen met meervoudige beperkingen, vormde het omgaan met een onbekende cultuur en andere levensomstandigheden op een klein eiland een persoonlijke uitdaging.

Enkele levenslessen rijker bij terugkomst in Nederland in de zomer van 1989, besloot ze een jaar de tijd te nemen om zich verder op de toekomst te oriënteren. In dat jaar verrichtte ze werkzaamheden als waarnemend logopedist en behaalde enkele modules Organisatiekunde aan de Open Universiteit. Uiteindelijk bleek de studie Spraak- en Taalpathologie aan de toenmalige Katholieke Universiteit Nijmegen en de stap naar het Nijmeegse studentenleven te voorzien in de behoefte om kennis en ervaring te verrijken. In 1993 studeerde ze cum laude af, en startte in datzelfde jaar haar werkzaamheden als logopedist op een afdeling voor niet-aangeboren hersenletsel van de Sint Maartenskliniek. Met veel plezier was zij elf jaar lang actief in een multidisciplinair revalidatieteam, waarbij revalidatiebrede vraagstukken steeds haar belangstelling hadden. In 2004 en 2005 was zij werkzaam in het operationele management.

Eind 2005 werd ze in de gelegenheid gesteld om een medior- c.q. onderzoeksfunctie te vervullen binnen het Ontwikkelcentrum voor Spraak- en Taaltechnologie (OSTT). De Sint Maartenskliniek werkt in het kader van dit expertisecentrum nauw samen met de Radboud Universiteit (Faculteit Letteren) en de afdeling Revalidatie van het UMCN. Gefaciliteerd door de Sint Maartenskliniek startte zij in 2007 haar promotieonderzoek over E-learning gestuurde SpraakTherapie (EST). Zij werd daarbij begeleid door professor Rietveld van de Faculteit der Letteren aan de Radboud Universiteit, en door professor Geurts van de afdeling Revalidatie in het UMCN. Het winnen van de Nationale Zorgvernieuwingsprijs 2008 met het EST-project was een grote stimulans om ook andere e-health projecten met verschillende partners te initiëren. De unieke zorg- en academische setting voor toegepast onderzoek in Nijmegen biedt daarvoor een uitstekend klimaat. Binnen de verschillende projecten wordt het implementeren van bestaande taal- en spraaktechnologie ten behoeve van mensen met communicatieve beperkingen nagestreefd. Een voorbeeld is het ComPoli-project (2011-2013), dat wordt uitgevoerd in het kader van het Innovatieprogramma Revalidatie. Dit project beoogt digicommunicatie van neurologische patiënten met zorgprofessionals te faciliteren.

Stellingen

1. In de huidige beeldcultuur blijft spreken als communicatiemiddel onverminderd van belang en blijft spraaktraining noodzakelijk voor Parkinsonpatiënten. Hierbij speelt E-learning een essentiële rol (Ir. L.G. (Bert) van Esterik, participant EST-onderzoek).

2. Voor het succesvol doorlopen van de EST-procedure door neurologische patiënten is het van belang om de verminderde auditieve discriminatie van de eigen spraak met de doelspraak te ondersteunen aan de hand van automatische, gevisualiseerde feedback (dit proefschrift).

3. Voor het meten van subtiele verschillen in verstaanbaarheid op verschillende tijdstippen binnen sprekers is veelal een hogere sensitiviteit van uitkomstmaten vereist dan voor het meten van verschillen in verstaanbaarheid tussen sprekers (dit proefschrift).

4. Het vinden van een optimale balans tussen verschillende typen spraakmateriaal enerzijds en de maximale hoeveelheid spraakmateriaal voor patiënten met dysartrische spraak anderzijds, is een essentieel onderdeel in de voorbereidende fasen van klinisch effectonderzoek (dit proefschrift).

5. Dankzij spraaktechnologie biedt een middels e-health gegenereerde digitale databank van dysartrische spraak op termijn ook mogelijkheden voor telerevalidatie op functiedomeinen buiten de communicatie (dit proefschrift).

6. Het is van belang om achtergrondvariabelen van patiënten, zoals mobiliteit en ervaring met computergebruik, te betrekken in onderzoek naar de waardering van EST door gebruikers (dit proefschrift).

7. Neurologische patiënten van 65 jaar en ouder maken verrassend vaker gebruik van internet dan verwacht (EST-researchprogramma).

8. Het sluiten van vriendschap met de data vormt veelal de grootste uitdaging in wetenschappelijk onderzoek.

9. Het effect van lippenstift op intermenselijke interacties zou wetenschappelijk onderzocht moeten worden.

10. De betekenis van het vijfde wiel aan de wagen wordt doorgaans onderschat.