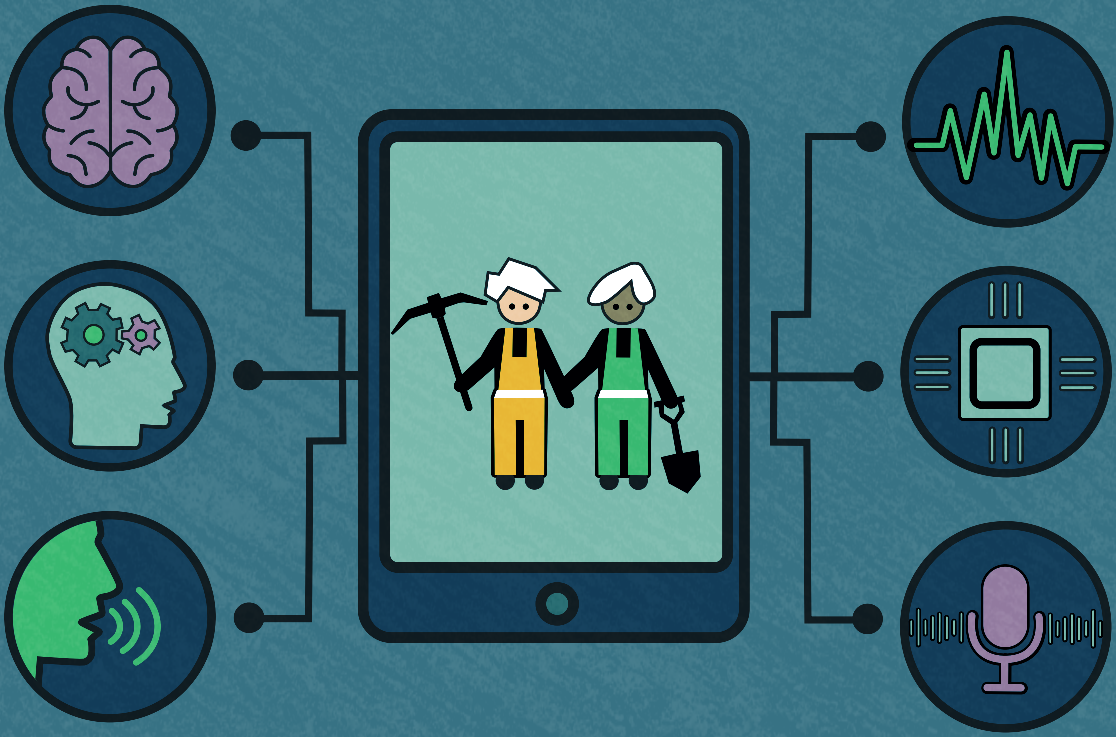
CHALLENGING SPEECH TRAINING IN NEUROLOGICAL PATIENTS BY INTERACTIVE GAMING



Centre for
Language Studies

MARIO SEBASTIAAN GANZEBOOM

RADBOUD
UNIVERSITY
PRESS

Radboud
Dissertation
Series

Challenging Speech Training in Neurological Patients by Interactive Gaming

Mario Sebastiaan Ganzeboom

The research in this thesis was funded by the Dutch Research Council (NWO) under the grant with Ref. no. 314-99-101 (CHASING).

Challenging Speech Training in Neurological Patients by Interactive Gaming
Mario Sebastiaan Ganzeboom

Radboud Dissertation Series
ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: Mario Sebastiaan Ganzeboom
Cover: Proefschrift AIO | Guntra Laivacuma
Printing: DPN Rikken/Pumbo

ISBN: 9789465152097
DOI: 10.54195/9789465152097
Free download at: https://doi.org/10.54195/9789465152097

RADBOUD
UNIVERSITY
PRESS

Challenging Speech Training in Neurological Patients by Interactive Gaming

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 30 januari 2026
om 12:30 uur precies

door
Mario Sebastiaan Ganzeboom
geboren op 5 augustus 1984
te Olst

Promotor:

Dr. W.A.J. Strik

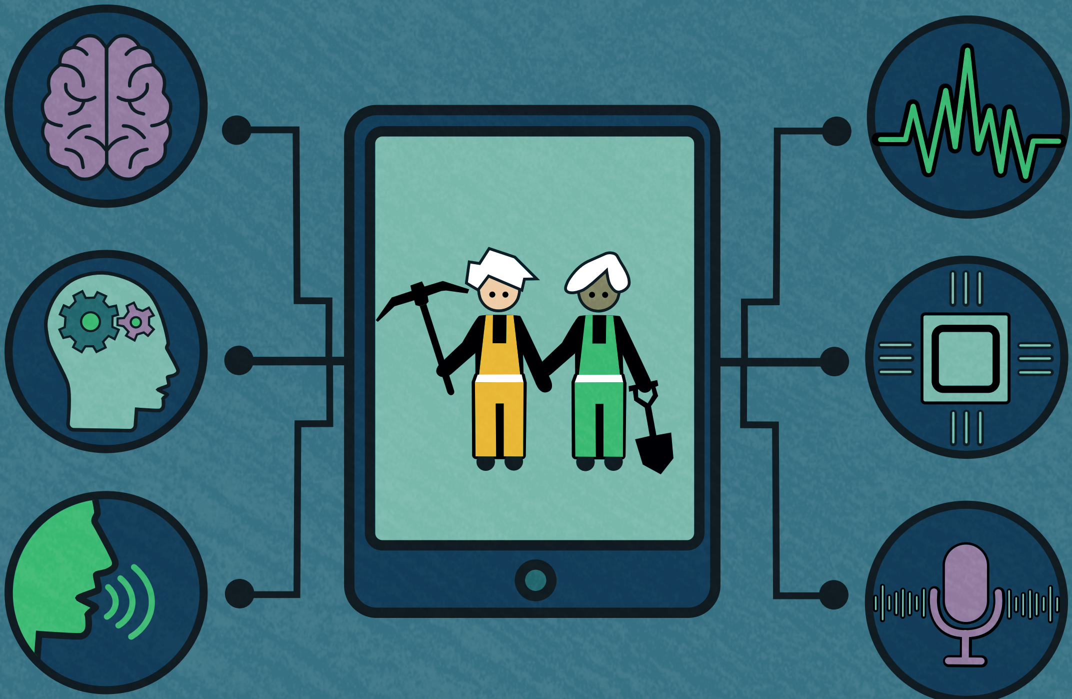
Copromotor:

Prof. dr. A.C.M. Rietveld

Manuscriptcommissie:

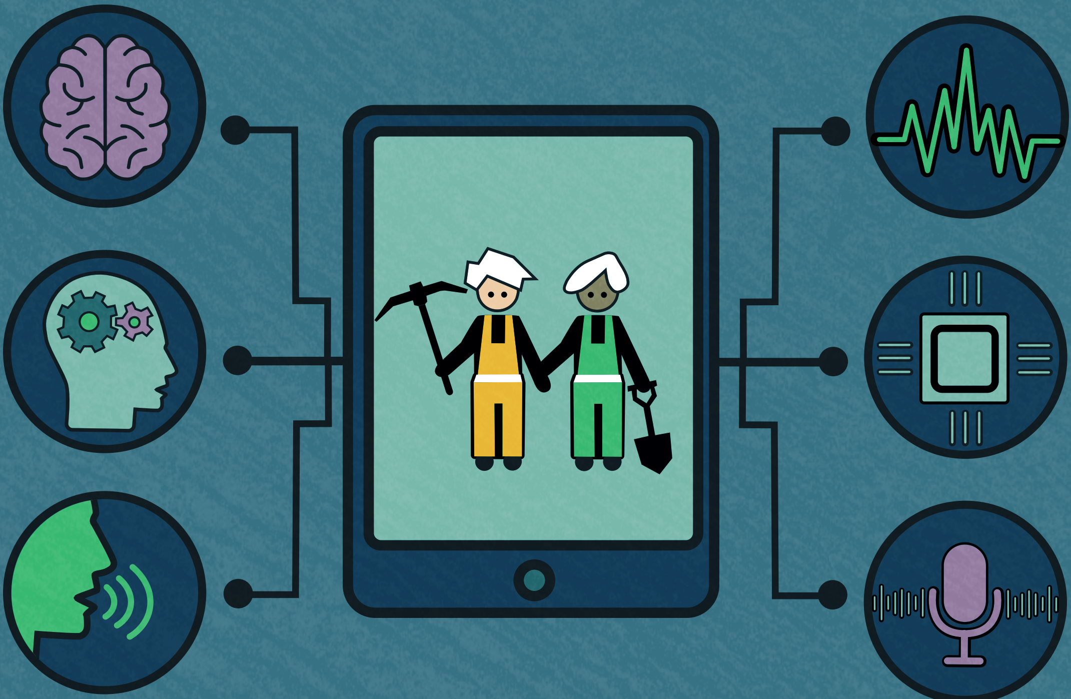
Prof. dr. H.H.J. Das
Prof. dr. H. Christensen (University of Sheffield, Verenigd Koninkrijk)
Prof. dr. ing. E. Nöth (Friedrich-Alexander-Universität Erlangen-Nürnberg, Duitsland)
Prof. dr. P. Santens (Universiteit Gent, België)
Dr. E. Janse

Contents



Part I.

Introduction



Chapter 1

General Introduction

1.1. Background

As populations around the world are ageing, the group of patients with acquired neurological disorders such as Parkinson's disease (PD), Cerebral Vascular Accident (CVA or stroke) and traumatic brain injury (TBI) is growing. Global, age-standardized incidence rates of 12-15, 189-218 and 331-412 individuals per 100.000 residents were reported for PD, CVA and TBI, respectively in 2016 (Feigin et al., 2020). Many of these patients have speech complaints (De Cock et al., 2020; Gandhi et al., 2020; Moya-Galé & Levy, 2019).

Neurological disorders are often complex and affect a person's life in multiple ways. The World Health Organization (WHO) has developed the International Classification of Functioning, Disability and Health (ICF) (World Health Organization, 2001), a universal framework for the description of how an individual is affected by health-related conditions. The ICF provides three components for classification: the body component (e.g. organs, limbs and their components), activities and participation (individual and societal range of functioning), and environmental factors (external factors influencing an individual's functioning).

Patients with a neurological disorder are often affected in all three of the ICF components. Among them are patients with dysarthria due to PD, CVA and TBI and other neurological disorders such as cerebral palsy (CP), amyotrophic lateral sclerosis (ALS) and multiple sclerosis (MS). Duffy (2019) defined dysarthria as a "neurologic speech disorder that reflects abnormalities in strength, speed, steadiness, tone or accuracy of movements required for the breathing, phonatory, resonatory, articulatory or prosodic aspects of speech production" (Duffy, 2019). The underlying physiological systems responsible for these aspects include: the respiratory system (breathing), larynx (phonation, prosody), velum and nasal cavities (resonance), tongue and lips (articulation). Due to the diminished functioning and coordination of muscles in these systems, speech rate, articulation, prosody and voice quality often change, the speech becomes less intelligible, and patients are thus impaired in their communication (Kent & Kim, 2003). For example, simple daily communication activities like inviting one's partner for a cup of coffee can become difficult for the partner to understand. Similarly, asking a shop assistant where to find a product in a store can require substantial effort. The patient could loose contact with friends or acquaintances, as they might not be prepared to adjust to new ways of communication. As a result, patients might enter into a depression due to isolation and loneliness. Obviously, dysarthria type and severity are of influence.

Two main fields of research have aimed at improving the communicative abilities of individuals with dysarthria: Augmentative and Alternative Communication (AAC) and speech rehabilitation. AAC research aims to augment existing communicative abilities or provide means of alternative communication. Speech rehabilitation, on the other hand, aims to improve intelligibility by restoring as much as possible of the functioning of the speech subsystems affected by dysarthria or other speech and voice disorders. The current thesis will focus on speech rehabilitation, but it is possible that the technology and the algorithms developed in this context will turn out to be useful for AAC as well.

Speech rehabilitation is normally provided by speech therapists, but the need for intensive therapy and the growing number of patients make it difficult for therapists to cope with the demand. Therefore, there is a growing attention for telerehabilitation: the delivery of healthcare services remotely using information and communication technologies (Scott et al., 2024). Attempts have been made at finding solutions to provide telerehabilitation through speech technology applications (Beijer, 2012; Beijer et al., 2014; A. E. Halpern et al., 2012). The results obtained so far are promising. Patients that participated in corresponding effect studies (Beijer, 2012; Beijer et al., 2014) described being able to follow the training program, but having increasing difficulty adhering to its repetitive 'drill-and-practice' nature and not experiencing beneficial effects in daily-life communication. The first key point of feedback indicates that the nature of the speech training exercises may not have contributed to participants' motivation to follow the training program. Additionally, the second key point of feedback indicates a poor ecological validity of the application. Exploring different approaches to remote and independent speech training potentially provides valuable insights on these topics. Interactive gaming is one such approach with the goal to leverage its entertainment and enjoyment properties for a 'serious' purpose, hence the term 'serious gaming'. For that reason, the current thesis explores the use of serious gaming for the purpose of remote and independent speech training.

To start the exploration of using serious gaming for speech training, the present chapter reviews past and current research on four main topics. The first topic, in section 1.2, provides a detailed overview of dysarthria types to give an idea of how they affect speech intelligibility and which challenges they pose to processing by means of speech technology.

In section 1.3, the second main topic of the current thesis is introduced. It provides an overview of research relevant to serious gaming and healthcare with speech rehabilitation in particular. The purpose of this section is to identify how serious gaming can be relevant to speech rehabilitation and what has already been done in this area.

After having examined the literature on dysarthria types including their effects on speech intelligibility and how serious gaming can be relevant to speech rehabilitation, the third main research topic is introduced: methods that measure the effects of a speech rehabilitation program on intelligibility. In section 1.4 an overview of such literature is provided by reviewing subjective and objective measures. The results of that review will inform the method and experiment design for measuring the effects of a speech rehabilitation program using serious gaming.

Next, the current thesis explains how speech technology has so far been employed for the purpose of speech rehabilitation (section 1.5), paying special attention to Automatic Speech Recognition (ASR) technology, the fourth main research topic in the current thesis. The introductory section provides a brief overview of approaches to ASR systems in general. Subsequently, in subsection 1.5.1 research is reviewed on how the various ASR systems have been adapted and optimized to maximally cope with the deviations in dysarthric speech identified. Research on how ASR has been employed for assessing the quality of dysarthric speech and identifying possible problematic areas is discussed in subsection 1.5.2.

This chapter then introduces the research project underlying the current thesis in section 1.6 and the last section of this chapter (section 1.7) provides the research questions the current thesis addresses as well as an outline of the remaining chapters of the current thesis.

1.2. Types of dysarthria: causes, speech characteristics and commonalities

Darley et al. (1969a, 1969b) were among the first to study how speech was affected by dysarthria with respect to its different etiologies. Their studies also analyzed to what extent these aspects correlate with speech intelligibility. They conducted a large scale study including considerable number of patients and published their results in two companion papers. In this large scale study, 32 speech samples of at least 30 patients per neurologic group were recorded. Recordings were made of seven neurologic groups: pseudobulbar palsy, bulbar palsy, amyotrophic lateral sclerosis, cerebellar lesions, parkinsonism, dystonia and choreoathetosis. The speech samples were perceptually rated by the three authors independently on 38 dimensions of speech and voice using a 7-point scale of severity. Analysis of these ratings showed that there are multiple types of dysarthria, each reflecting a different kind of abnormality in neuroanatomy or neurophysiology of the speech production system. These types of dysarthria sound different and can therefore be differentiated by the 38 dimensions of speech and voice (see appendix in Darley et al., 1969a, 1969b) and the varying degree of severity in which they occur. Seven types of dysarthria were distinguished: spastic, flaccid, mixed spastic-flaccid, ataxic, hypokinetic, hyperkinetic in chorea and hyperkinetic in dystonia. To illustrate these dysarthria types and their effects on speech, Table 1.1 provides a summary of their perceptual characteristics, the five dimensions that on average show the largest deviation and the ones that correlate strongly with speech intelligibility. For details, see Darley et al. (1969b).

Common deviating speech dimensions across dysarthrias

Reflecting on the results mentioned in the previous paragraphs, Darley et al. (1969b) found that most dysarthrias had multiple speech dimensions in common on which speech deviations were observed. Noticeable was that ‘Imprecise Consonants’ ranked first in almost all dysarthrias, except flaccid (second) and hypokinetic dysarthria (fourth). The dimensions that follow were found to be common to multiple dysarthrias and deviated with a mean of 1.5 or above, 1 representing normal (neurologically healthy) speech and 7 representing very severe deviation from normal. ‘Imprecise Consonants’, ‘Monopitch’, ‘Monoloudness’ and ‘Harsh Voice’ were found in all seven dysarthrias. This shows that all dysarthrias seem to have at least one dysfunction in the neurologic groups of articulatory inaccuracy, prosodic insufficiency and phonatory stenosis as described by Darley et al. (1969a). In five dysarthrias ‘Vowels Distorted’, ‘Phrases Short’ and ‘Reduced Stress’ were found. Dimensions occurring in four of seven dysarthrias were ‘Breathy Voice (Continuous)’, ‘Strained-Strangled Voice’, ‘Hypernasality’, ‘Slow Rate’, ‘Intervals Prolonged’, ‘Inappropriate Silences’, ‘Excess and Equal Stress’ and ‘Phonemes Prolonged’. See Darley et al. (1969b) for additional details.

Dysarthria type	Causes	Symptoms	Deviating dimensions	Intelligibility correlation
Spastic (or pseudobulbar palsy)	CVA (or Stroke), TBI, CP of infancy, extensive brain tumors, encephalitis, MS	Slow and limited range of lip movement, rapid alternating movements slowed and effortful, difficulties swallowing, short outburst of laughing and crying without affect	Imprecise Consonants Reduced Stress Vowels Distorted Monopitch Monoloudness	Imprecise Consonants Monopitch Reduced Stress Harsh Voice Monoloudness ... 11. Vowels Distorted
Flaccid (or bulbar palsy; subclassification by affected cranial nerves)	TBI, viral infections, Myasthenia gravis, extensive brain tumors, encephalitis, MS	All subtypes: weakness of muscles (flaccidity)	Hypernasality Imprecise Consonants Breathy Voice (Continuous) Monopitch Nasal Emission ... 9. Monoloudness	Imprecise Consonants Vowels Distorted Hypernasality Nasal Emission Slow Rate Phrases Short
Mixed spastic-flaccid	ALS	Similar symptoms of both spastic and flaccid dysarthria	Imprecise Consonants Hypernasality Harsh Voice Slow Rate Monopitch	Imprecise Consonants Vowels Distorted Phrases Short Monoloudness Low Pitch ... 8. Monopitch
Ataxia	Brain tumours, TBI, MS, CVA, progressive degeneration, alcoholic excess, congenital conditions	Errors of varying degrees in timing, force, range and direction of movement	Imprecise Consonants Excess and Equal Stress Irregular Articulatory Breakdown Vowels Distorted' Harsh Voice'	Irregular Articulatory Breakdown Imprecise Consonants Vowels Distorted Monopitch Nasal Emission Monoloudness
Hypokinetic	PD	Rigidity of muscles, tremor at rest, movements are slow, limited in range and force of contraction, movements may become arrested, reduced range of rapid successive movements becoming progressively smaller	Monopitch Reduced Stress, Monoloudness Imprecise Consonants Inappropriate Silences	Imprecise Consonants Short Rushes Reduced Stress Variable Rate Monoloudness
Hyperkinetic in chorea	Huntington's disease, Sydenham, poststroke chorea	Quick, overactive and irregular movements being random, unpatterned and nonfunctional, uncontrollable muscle contractions	Imprecise Consonants Intervals Prolonged Variable Rate Monopitch Harsh Voice ... 7. Vowels Distorted	Vowels Distorted Imprecise Consonants Strained-Strangled Voice Phrases short Monopitch ... 10. Monoloudness
Hyperkinetic in dystonia	PD, Huntington's disease, Wilson's disease, stroke, TBI, intoxication of heavy metals or carbon monoxide	Muscle contractions slowly build up, hold there contracted posture for a while and then slowly relax	Imprecise Consonants Vowels Distorted Harsh Voice Irregular Articulatory Breakdown Strained-Strangled Voice Monopitch	Imprecise Consonants Vowels Distorted Monoloudness

Table 1.1.: *Summary of dysarthria types and their speech characteristics as described by Darley et al. (1969b). The fore last column lists the top five deviating speech dimensions and the last column the top 5 dimensions correlating with intelligibility.*

From the four deviant speech dimensions found in all dysarthrias, two or more also significantly correlate with speech intelligibility in almost all dysarthrias (flaccid dysarthria only correlates with one). Put differently, ‘Monopitch’, ‘Monoloudness’, ‘Harsh Voice’ and ‘Imprecise Consonants’ affect the speech of the studied patients in such a way that their speech becomes less intelligible for the listener. The dimensions ‘Phrases Short’, ‘Reduced Stress’ and ‘Vowels Distorted’ also have this effect, but to a lesser extent.

1.3. Serious gaming and rehabilitation

Digital games started out as a form of leisure activity providing entertaining and engaging activities. Research into digital games received increased attention in the first half of the 2000s, primarily due to the increased availability of more violent games and their negative effects on aggression and addiction (Connolly et al., 2012). Later research also found that playing games was associated with increased visual spatial abilities and eye-hand coordination (C. J. Ferguson, 2007). In combination with the motivating characteristics of digital games, new studies were started in the second half of the 2000s that researched the ability and usefulness of digital games to provide a novel method of learning (Connolly et al., 2012). At first, research focused on how commercial-off-the-shelf games could be used for learning. Later on, that focus also started to include games-based learning, games especially designed for learning. Due to this widening of focus, interest in serious games (Stanfield, 2008) and persuasive games (Bogost, 2007) was renewed. A widely accepted definition of serious games is: ’games that do not have entertainment, enjoyment, or fun as their primary purpose’ (Laamarti et al., 2014). They have been used for learning or changing behaviours in the areas of education, training and health. Persuasive games have often been used to inform and change attitudes in the areas of politics, public policy and advertising. At around the same time, the concept of gamification was introduced by the digital media industry (Al-Rayes et al., 2022). There seems to be no clear definition of gamification. However, it is widely accepted as "referring to the process of using game design elements and experience to improve individuals’ engagement and motivation in non-game processes" (Al-Rayes et al., 2022, p. 2). This immediately provides a clear distinction with serious games, which are considered to be full games that model an entire learning or behavioural change process, whereas gamification only adds game design elements (e.g. earning points, badges, or other reward system) to improve such a process. Interest in research for using serious games and gamification in health increased at the beginning of the 2010s (Lister et al., 2014; Nøhr & Aarts, 2010). For the latest developments in serious games for healthcare, Damaševičius et al. (2023) and Pakarinen and Salanterä (2020) provide comprehensive overviews. Al-Rayes et al. (2022) have done the same for gamification and healthcare in general. Koivisto and Malik (2022), focus specifically on gamification for elderly, which is relevant to the current thesis as patients affected by stroke, PD or TBI are predominant in this group (Feigin et al., 2020).

In the 2010s, researchers in the field of rehabilitation (e.g. physical, cognitive, psychological, etc.) also started researching whether serious games and gamification could be applied to increase patients' engagement and motivation with the often repetitive nature of therapy exercises (Burke et al., 2009, 2010; G. N. Lewis et al., 2011; López-Rodríguez & García-Linares, 2013). Serious games and gamification are both used to this end. However, given the available body of literature, the current thesis observed that gamification is applied moreover to games for health and well-being. They implement the idea of staying physically fit by engaging in exercise, encouraging a healthy diet, and assist in medicine or chronic disease management (Al-Rayes et al., 2022). These types of games also better suit the definition of gamification, as they do not require to model an entire learning or behavioural change process. However, that is the case in the context of rehabilitation. By definition, serious games are better suited for modelling such a process, as they can provide an immersive experience to engage and motivate the player for rehabilitation over a longer period. A large body of literature on rehabilitation that includes serious gaming focuses on training motor and cognitive functions (Ong et al., 2021). Serious gaming for the rehabilitation of elderly patients with stroke, PD or TBI, is a part of that (de Oliveira et al., 2021; Mubin et al., 2022). Specifically, for speech rehabilitation in this group, research including serious gaming is limited (Baranyi et al., 2024; Hajesmaeel-Gohari et al., 2023; Krause et al., 2013; Mühlhaus et al., 2017; Weber, 2025). However, they do show that serious games' immersive and engaging nature can be beneficial to the motivation and enjoyment of intensive speech training. As a result, these benefits can increase therapy adherence and improve dysarthric speakers' speech intelligibility.

1.4. Measurement of speech intelligibility in dysarthric speech

As the current thesis focuses on speech rehabilitation, its aim is to improve dysarthric speakers' speech intelligibility using a novel method of speech therapy. Multiple definitions of intelligibility in speech rehabilitation have been presented in the past (Duffy, 2019; Hustad, 2008; Kent, 1992; Olmstead et al., 2020). Intelligibility of speech has not only been studied in speech rehabilitation. Other fields like second language (L2) pronunciation (Levis & Silpachai, 2022), speech perception (Smiljanic, 2021), telecommunication (Tomassi et al., 2023), and speech synthesis evaluation (B. M. Halpern et al., 2021) have also studied this topic from different perspectives. A widely adopted definition of intelligibility in the field of speech rehabilitation and used in the current thesis is the one proposed by Hustad (2008): "Intelligibility refers to how well a speaker's acoustic signal can be accurately recovered by a listener". Methods to measure speech intelligibility are necessary to determine whether and how speech intelligibility improves during speech rehabilitation.

A common category of methods for the measurement of dysarthric speech intelligibility is that of subjective measurement methods. Publications in this category started in the 1960s (Darley et al., 1969b; Speaks et al., 1972; Tikofsky, 1970; Yorkston & Beukelman, 1978). Subjective measurement methods commonly rely on judgements by human raters who listen to recorded speech samples. Traditional approaches ask raters to indicate the degree of intelligibility on one or more scales. Common scales that have been employed are equal-appearing interval scales like a Likert scale (Haderlein et al., 2011; Moya-Galé et al., 2018; Yorkston & Beukelman, 1978) or a Visual Analog Scale in which a point on a horizontal line indicates the degree of intelligibility (Abur et al., 2019; Finizia et al., 1998; Rietveld, 2021; Van Nuffelen et al., 2010). Measures based on orthographic transcriptions have also been studied in the past (Beijer et al., 2012; Ishikawa et al., 2017; Yorkston & Beukelman, 1978). For these measures listeners are asked to write down what they heard. Afterwards, the percentage of correctly transcribed words is usually employed as a measure of intelligibility. However, what the historic review by Kent et al. (1989) already describes, "...intelligibility is not an absolute quantity but rather is a function of such parameters as test material, personnel, training, test procedures, and state of the speaker." Following that, the outcome of measures of intelligibility can differ when any of these parameters are altered. Thus, the context in which these measures are determined plays a pivotal role and must be kept equal over measurement sessions (e.g. when measuring the effects of speech rehabilitation). Later research has shown that the degree of intelligibility measured varies between different types of speech material (Beverly et al., 2010; Hustad, 2007; Kempler & Lancker, 2002). Severity of dysarthria was also found to play a role (Hustad, 2007), as well as the type of listeners being either expert or naive or older or younger adult listeners (Dagenais & Wilson, 2012; Pennington & Miller, 2007; Walshe et al., 2008). Testing procedures must also take listeners' familiarity with the rated speakers into account (D'Innocenzo et al., 2006; Pennington & Miller, 2007; Tjaden & Liss, 1995). This prevents skewed ratings as familiarity increases after having heard more of a particular speaker.

Criticism has been expressed towards scale-based measures as they lacked a ground truth, meaning that listeners rated dysarthric speech samples with their own impressions of speech with low and high intelligibility. This was addressed by the introduction of different methods in which one or more reference samples were presented to the listeners at the start of the listening experiment (Kent et al., 1989; Schiavetti, 1992; Weismer & Laures, 2002). Those reference samples showed what was considered intelligible and unintelligible speech. Another point of criticism is that with scale-based measures, it is not clear on what listeners base their perceptions and to what extent they include speech dimensions relevant to intelligibility. Previous research has shown that the reliability of some of these measures is too low to enable firm conclusions (Miller, 2013; Schiavetti, 1992). However, measures based on VAS have shown to be reliable for research purposes (Kent & Kim, 2011; Van Nuffelen et al., 2010). In chapter 2 of the current thesis, research has also been reported on this topic.

Different from scale-based measures, orthographic transcription-based measures rely on the amount of information listeners accurately perceived. They are considered as reliable (Bunton et al., 2001; Miller, 2013; Tjaden & Wilding, 2011). In chapter 2, the current thesis provides its results on the reliability of transcription-based measures. In contrast to scale-based measures, transcription-based measures tend to be more laborious and time-consuming. Listeners require more effort as they are asked to key in the transcription, which is prone to spelling errors and typos. It may also be more difficult to perform for less computer-literate listeners. Additionally, researchers need to process the transcriptions before being able to calculate meaningful measures. Scale-based measures do not have these disadvantages. Clicking on a scale or slider tends to be easier and provides scores that require almost no post-processing. The role of contextual information in the used speech materials also needs to be taken into account when employing transcription-based measures. This may unintentionally facilitate the listener in recognizing the words that were uttered. Previous research introduced Semantically Unpredictable Sentences to circumvent this problem (Beijer, 2012; Benoît et al., 1996). The SUS are syntactically correct, but semantically incoherent. In chapter 2 these sentences are also used in the current thesis research on intelligibility measures.

Given the advantages and disadvantages of both types of measures, it makes sense that both are used in the assessment of dysarthric speakers intelligibility. Arguably, these measures can complement each other as orthographic transcription-based measures tend to be more suitable for brief utterances (i.e. isolated sentences, word lists, and isolated words) and scale-based measures for longer utterances (i.e. passages of read text, and brief conversations). How these measures actually relate to each other is still a topic of research. Some studies found that transcription-based measures showed higher intelligibility scores for the same speech samples than scale-based ones (Hustad, 2006). Similar research was conducted and described in chapter 2 of the current thesis. Other studies found that the transcription-based measures highly correlated with VAS scores (Abur et al., 2019; Ishikawa et al., 2021; Schiavetti, 1992; Stipancic et al., 2016) and were comparable to percentage estimates (Yorkston & Beukelman, 1978).

As noted previously, the current thesis' chapter 2, adapted from a previous publication, has also contributed to the research of transcription-based measures for intelligibility by investigating them at different levels of granularity including how they relate to scale-based measures. Since then, research into these different levels, their interrater reliability, and validity has continued (Xue, Van Hout, Boogmans, et al., 2021; Xue, Van Hout, Cucchiarini, & Strik, 2021). The contribution of subword, phoneme-level transcription-based measures to the assessment of dysarthric speech intelligibility has also been researched (Xue et al., 2023). In addition to transcription-based measures, research has also focused on factors that influence listeners perceptions of dysarthric speech (Lehner & Ziegler, 2021; McAuliffe et al., 2017; Patel et al., 2014). A novel view on speech intelligibility that includes not only speaker and listener, but also their joint contributions in communication is introduced by Olmstead et al. (Olmstead et al., 2020). They argue that for advancing clinical interventions, it is important to research how speech intelligibility evolves during communicative interactions. In addition to speech intelligibility, recent research has also found that perceived listening effort plays a distinct role in the communication of dysarthric speakers (Van der Bruggen et al., 2023).

In reflection, research into subjective measures has already resulted in a large body of literature and continues in multiple directions. However, subjective measures have not only been the topic of research in measuring speech intelligibility. Publication of studies into objective measures started around the same time (i.e. the 1960s, Hardy, 1967; Lehiste et al., 1961). Objective methods rely on measures that are calculated directly from acoustic signals produced by the vocal tract. One of the reasons that researchers started studying objective methods is that subjective methods rely on scores provided by human judges. In a way, these judgements remain a subjective interpretation of a score in which judges may have used slightly different mental models for scoring. In other words, the process judges use for scoring remains a black box. Objective methods do not have these disadvantages, as measurements are made directly on the acoustics of the recorded speech signal or articulatory movements. Studies showed that acoustic measures based on pitch (Feenaughty et al., 2014; Tjaden & Wilding, 2010; Tjaden & Wilding, 2004), intensity (Bunton et al., 2000; Cannito et al., 2012; Feenaughty et al., 2014; Holmes et al., 2000; Tjaden & Wilding, 2004), and formant frequencies (Tjaden & Wilding, 2010; Tjaden & Wilding, 2004; Weismer et al., 2000) are related to speech intelligibility and that they can contribute to measuring changes therein as part of a speech rehabilitation program. Studies into the acoustics of consonant articulation have also shown relation with speech intelligibility (Ackermann & Ziegler, 1991; Johansson et al., 2023; H. Kim & Gurevich, 2023). Many of the previous studies also show that the calculation of objective measures can often be automated.

From the previous literature overview the current thesis considers that, for its purposes, subjective and objective measurement methods can both be used for detecting changes in speech intelligibility as a consequence of speech training. In particular, objective measures appear to be suited for gaming contexts where immediate automated feedback is preferred.

1.5. Speech technology for speech rehabilitation

Multiple types of treatments for speech rehabilitation exists. In addition to treatments involving intensive speech training, which is the topic of this thesis, other invasive surgical treatments followed by therapy may also be resorted to in speech rehabilitation (Swinnen et al., 2023). Other non-invasive cortical and peripheral stimulation treatments are also being researched (Balzan et al., 2022). However, recently published research shows that speech-language therapists around the world still primarily rely on the assessment of an individual's speech followed by a speech therapy program based on the result of the assessment (Balzan et al., 2023). In speech therapy, speech exercises are selected depending on the characteristics of the dysarthria to optimally target the impaired speech subsystems. Previous research has shown that intensive speech therapy can be effective in increasing voice intensity (i.e. loudness) for adequate communication (De Swart et al., 2003; Ramig et al., 2001). However, the intensiveness of the therapy and the growing number of patients make it difficult for therapists to provide sufficient treatment to everyone. Recent initiatives to address this problem by further intensifying speech therapy in the patient's home environment have proven beneficial (Beijer et al., 2014; Mendoza Ramos et al., 2021; Orozco-Arroyave et al., 2020). These initiatives have come from the field of telerehabilitation, referring to the delivery of medical rehabilitation in the home (Rosen, 1999). Telerehabilitation allows patients to train their speech at home, potentially reducing the number of visits to rehabilitation centres, which may be difficult due to possible physical disability.

Speech analysis and speech technology, i.e. Automatic Speech Recognition (ASR), can be applied to provide speech telerehabilitation. The advantage of using ASR and related algorithms in addition to speech analysis is that the content of the utterances produced by the patients can be controlled for to make sure that analyses are carried out on the relevant aspects of the patient's speech and that feedback is provided on the features that need to be improved.

In speech rehabilitation, ASR technology can be applied for assessing the patient's speech and for detecting deviations on individual utterances, words and speech sounds with a view to providing feedback and therapy. In principle, both tasks can be carried out offline by first analysing samples of the patient's speech and then providing delayed feedback either in the form of a score in the case of assessment or as information about the utterances, words and speech sounds that need to be improved in the case of therapy. In the latter case, however, it is preferable to provide instantaneous, online feedback on individual utterances, words or problematic realisations of a speech sound when these are still available in working memory and the patient can more easily try to improve them. As will become clear in subsection 1.5.1, this poses considerable challenges to ASR technology.

In the following subsection (subsection 1.5.1), the first paragraphs review research on the different types of ASR systems that have been developed in recent history. Subsequent paragraphs and subsubsections describe how these types of ASR systems have been adapted and optimized to maximally cope with the deviations in dysarthric speech identified in section 1.2. Finally, subsection 1.5.2 will introduce research on ASR-based evaluation of dysarthric speech.

1.5.1. ASR for dysarthric speech

Research into the first major approach to ASR systems started in the 1970s (Baker, 1975; Reddy et al., 1973) and remained state-of-the-art until the end of the 2000s.

These, now classical ASR systems, are generally composed of three ‘knowledge sources’: a language model, lexicon and acoustic models. The language model contains probabilities of words and strings of words from which the prior probability is calculated on large text corpora. Acoustic models contain probabilities on the way sounds (phones) are pronounced. Many classical ASR systems use Gaussian Mixture Model-based Hidden Markov Models (GMM-HMMs) as acoustic models. The lexicon contains a list of words that maps the way they are written (strings of graphemes), and the way they are pronounced (strings of phones). Variation in pronunciation due to speaker characteristics (Strik & Cucchiarini, 1999), but also due to neurological disorders (Mengistu & Rudzicz, 2011) makes the ASR process additionally challenging. As classical ASR systems cannot perform speech recognition on the raw audio signal well, the extraction of features that represent the speech in the audio signal is required in an audio preprocessing stage. For more details on the basics of classical ASR, see Rabiner (1989) and Jurafsky and Martin (2024).

The focus gradually moved to deep learning technology in the early 2010s (Dahl et al., 2012; Hinton et al., 2012; Mohamed et al., 2009; Sainath et al., 2013; Seide et al., 2011). In non-dysarthric ASR research, deep learning networks were introduced in a so-called hybrid ASR approach (Dahl et al., 2011, 2012; Seide et al., 2011; Zavaliagkos et al., 1993). In this approach, Deep learning Neural Networks (DNNs) are interfaced with traditional HMMs (i.e. DNN-HMMs), replacing the classic Gaussian mixture distributions. The hybrid approach became widely adopted as a consequence of that the discriminative power of densely connected networks with many hidden layers was now available due to advances in efficiently training those networks (Dahl et al., 2011; Hinton et al., 2006). Similar to the classical approach, the hybrid approach still required the three knowledge sources. For an extensive overview of research into DNN-HMMs and its variants, see O’Shaughnessy (2024) and Jurafsky and Martin (2024).

After having recognized deep learning’s potential, research into the end-to-end (E2E) ASR approach was introduced to simplify hybrid non-dysarthric ASR systems and reduce the often laborious work of obtaining the three knowledge sources they require. This approach follows the idea of sequence-to-sequence processing using a single, integrated ASR model, in which an input sequence (e.g. acoustic speech signal) is directly mapped into an output sequence (e.g. string of words, Perero-Codosero et al., 2022). The integrated model removes the requirement for the three separate knowledge sources from the hybrid approach. It also supports generalisation of the entire training process (Prabhavalkar et al., 2023), making it more efficient. However, at present, E2E systems tend to need larger amounts of training data than Hybrid DNN-HMM systems before reaching acceptable levels of performance Prabhavalkar et al. (2023). For more on the details of E2E technology and its advantages/disadvantages, see J. Li (2022) and Prabhavalkar et al. (2023).

Trends in dysarthric ASR research often follow those in non-dysarthric ASR, as the remainder of this section will show. However, some previous studies have deviated from these trends. For example, Raghavendra et al. (2001) studied the performance of dysarthric speakers using two commercial ASR systems. They found that systems could not perform well for severely dysarthric speakers and recognition scores of dysarthric speech were generally lower than those of non-dysarthric speech. A similar, more recent evaluation incorporating ASR based on deep learning technology claimed that recognition improvement was only limited in the best-case scenario (Jaddoh et al., 2023).

Mustafa et al. (2015) pointed out, that the trends in dysarthric ASR research show strong resemblances with the research for ASR in general. Similar to non-dysarthric ASR, research focused on more general factors like speech mode (e.g. isolated words, continuous speech), speaker mode (e.g. dependent, independent, adaptive), and vocabulary size. The next two paragraphs continue by describing research on these factors. Afterwards, this subsection continues by describing research on topics supporting the recognition improvement.

Acoustic modelling technology

Following the field of non-dysarthric ASR, research into acoustic modelling of dysarthric speech started in the 1990s with discrete Hidden Markov Models (dHMMs, Deller Jr. et al., 1991) and Artificial Neural Networks (ANNs F. Chen and Kostov, 1997; Jayaram and Abdelhamied, 1995). Later research studied the application of continuous (density) Hidden Markov Models using a mixture of Gaussian distributions (i.e. GMM-HMM Green et al., 2003; Hawley, Enderby, Green, Cunningham, Brownsell, et al., 2006; Hawley, Enderby, Green, Cunningham, and Palmer, 2006; Hawley et al., 2005; Sharma et al., 2009). Dysarthric ASR researchers recognized the potential of this, now classical, GMM-HMM approach which became the state-of-the-art in the field. All the previous cited research primarily describes speaker-dependent, isolated-word recognition with whole-word models. Arguments for using word-level acoustic models in dysarthric speech are that reliable phone-level units cannot be defined due to the high abnormality of speech at the phonetic level, which is influenced by the type of dysarthria and its severity. An additional argument is the application of the dysarthric ASR system. For example, Hawley, Enderby, Green, Cunningham, Brownsell, et al. (2006) and Mulfari et al. (2022) researched a dysarthric ASR system for assistive purposes, enabling control of devices in the home environment via speech. Such an application only requires the recognition of a single word or command string. A disadvantage of word-level acoustic models is that they result in less flexible ASR systems, because an acoustic model needs to be trained for every word before it can be recognized. Systems with phone-level acoustic models can generally recognize any word provided that the word and its phone transcription are added to the lexicon. Multiple studies researching phone-level acoustic models using the classical GMM-HMMs have been conducted in the past (Rudzicz, 2011; E. Sanders et al., 2002; Sharma et al., 2009, ch. 5.3.2). In addition, some other acoustic modelling technologies for recognizing dysarthric speech have been studied (Rudzicz, 2011; Shahamiri & Salim, 2014; Wan & Carmichael, 2005).

Following the introduction of deep learning technologies, one of the first studies that researched their effects on dysarthric ASR recognition accuracy was Christensen et al. (2013). Later research included multiple variants of hybrid DNN-HMM approaches (España-Bonet & Fonollosa, 2016; Hahm et al., 2015; Hermann & Magimai-Doss, 2020; Hu et al., 2022; Joy & Umesh, 2018; Joy et al., 2017; Nakashika et al., 2014; Sidi Yakoub et al., 2020; J. Yu et al., 2018; Zaidi et al., 2021). All these studies included comparisons to classical GMM-HMM approaches and showed that hybrid deep learning acoustic models provided significant improvements in recognition performance. Research on this topic has also been conducted for the current thesis in chapter 3 and chapter 4. For more details on the advancements in deep learning technologies for dysarthric ASR, see Qian et al. (2023) and Pophale and Chavan (2023).

Recently, researchers started to investigate end-to-end (E2E) ASR systems for the recognition of dysarthric speech (Takashima, Takiguchi, & Ariki, 2019). As described in the introductory section, the main challenge is to obtain a large amount of training data to reach acceptable performance levels. Collecting large amounts of dysarthric speech data is complex due to the highly variable nature of dysarthric speech and its heterogeneous group of speakers. For that reason, research has been conducted mostly on methods that reach acceptable levels with only a limited amount of data. For example, restructuring the layers of the trained acoustic DNN (Lin, Wang, Dang, et al., 2020), making more efficient use of the available data (Lin, Wang, Li, et al., 2020), or combining multiple methods using novel E2E modelling, data augmentation, and transfer learning (Almadhor et al., 2023; Shahamiri et al., 2023).

Training data

Most of the previous technologies require segmented speech recordings to train the acoustic models. The amount of data (i.e. speech recordings) required depends mainly on: the type of speech to be recognized (e.g. isolated words or continuous speech), the type of acoustic unit used (e.g. words, phones, or subphones), the size of the vocabulary (e.g. 50, 5000, or 25000 words), and the speaker mode (e.g. speaker-dependent, independent or adaptive). For example, the context that requires the most data is speaker-independent recognition of large-vocabulary continuous speech by using subphones. This is because many subphones need to be trained that must generalize over the variation in speech of many speakers. Research into dysarthric ASR is commonly conducted on dysarthric speech databases. These databases are not abundantly available, as it is complex to record dysarthric speakers due to their physical and cognitive impairments. For the same reason, it is often not possible to record large amounts of speech from a single dysarthric speaker. Consequently, the dysarthric speech databases that have become available remain small in comparison to those of non-dysarthric speech. For example, commonly used databases for English dysarthric speech are Nemours (approx. 3 hours, Menéndez-Pidal et al., 1996), UASpeech (47.8 hours, H. Kim et al., 2008), and TORGO (approx. 15 hours, Rudzicz et al., 2012). The COPAS database (3 hours, Middag, 2012) has been made available containing dysarthric speech for the Flemish variety of Dutch (in Belgium). For the northern variety of Dutch (in The Netherlands), The Netherlandic Dutch dysarthric speech database (6 hours, Yılmaz et al., 2016) has been recorded as part of the current thesis' project. This is a large difference when comparing to the hundreds or even thousands of hours in the English (Cieri et al., 2004) and Dutch speech databases (Oostdijk et al., 2002) for non-dysarthric speech. Consequently, other methods to improve recognition accuracy and compensate for the lack of training data had to be found.

Data augmentation

One of the methods that can compensate for the lack of training data, is data augmentation in which variations of the original dysarthric speech are created using signal processing techniques. Speed, tempo, and vocal tract length perturbation are common manipulations and have been found to be effective in increasing recognition accuracy (Vachhani et al., 2018; Xiong et al., 2019; Geng et al., 2020). Soleymanpour et al. (2021) have found a data augmentation technique tailored to dysarthric speech using prosodic transformation and masking. Other data augmentation techniques that have been found effective are based on deep learning technology to modify non-dysarthric speech spectra into those closer to dysarthric speech (Jin et al., 2021; Bhat et al., 2022; Zheng et al., 2023; C.-J. Li et al., 2025) or text-to-speech (TTS) synthesis (Soleymanpour et al., 2022, 2024; Wagner et al., 2025) in which a TTS system is trained to synthesize dysarthric speech for acoustic model training.

Acoustic feature representation

Another method for improving dysarthric speech recognition accuracy is found in researching the representation of acoustic features. Before the advent of deep learning technologies, this was often part of the signal preprocessing and feature extraction stage of an ASR system. Different acoustic features may represent dysarthric speech more effectively. In the past, studies included the analysis of deviant dimensions like speaking rate, frequencies of pauses, and articulation Magnuson and Blomberg (2000). They also investigated correlations between the parameters and measured fluency, intelligibility and articulation. Blaney and Wilson (2000) studied the acoustic differences between non-dysarthric, mild and moderate degrees of ataxic dysarthria, and attempted to analyse what the impact is on acoustic features and their variability on ASR. More successful feature sets have been proposed in literature of which Mel Frequency Cepstral Coefficients (MFCCs), Perceptual Linear Prediction (PLP) Coefficients, and mel-scale filterbank features are commonly used (M. Kim et al., 2016; Mathew et al., 2018). Recent efforts that included raw source and filter components for acoustic modelling have also shown notable recognition improvements (Yue, Loweimi, Christensen, et al., 2022). Including features from speech modalities other than acoustics like articulatory features from Electromagnetic Articulography (EMA) measurements (Rudzicz, 2011; Yılmaz et al., 2018; Yue, Loweimi, Cvetkovic, et al., 2022) or video recordings (Liu et al., 2019; Miyamoto et al., 2010) have also been found to improve recognition accuracy. The introduction of deep learning technologies saw the advent of representation learning. In this field deep learning technologies are used to automatically learn and extract useful speech features (Qian et al., 2023; Takashima et al., 2015, subsection 3.2.7).

Transfer learning

In chapter 3 of the current thesis, the effects of increasing RBM pre-trained DNN-HMMs training data by combining non-pathological data of different language varieties are studied. This research is continued in chapter 4 by adding retraining of the DNN-HMM model to the training process, which includes retraining on dysarthric data only and in combination with elderly speech. These chapters are considered early examples of transfer learning research applied to dysarthric ASR. Generally, in this type of research, the knowledge acquired from a source task or dataset is utilized to benefit a related target task or dataset. For dysarthric ASR this typically involves transferring knowledge from the acoustic space of neurologically healthy speech to the acoustic space of dysarthric speech. The following three very different examples of transfer learning research show its diverse application in dysarthric speech feature enhancement (Vachhani et al., 2017), the inclusion of dysarthric speech from a non-target language into pre-training on healthy speech from the target language (Takashima, Takashima, et al., 2019), and an utterance-based data selection strategy to more accurately select potentially beneficial out-of-domain data for retraining (Xiong et al., 2020).

Finally, a more traditional method for improving the accuracy of dysarthric speech recognition that the current thesis would like to note is: lexicon adaptation (Caballero-Morales & Trujillo-Romero, 2014; Mengistu & Rudzicz, 2011; Sriranjani et al., 2015).

1.5.2. ASR-based evaluation of dysarthric speech

In the previous subsection, research was reviewed on how to optimize the various ASR components to improve recognition of dysarthric speech. In order to use it for speech rehabilitation purposes, however, in many cases the target of ASR technology should not only be speech recognition, but should also be able to evaluate the quality of dysarthric speech and identify problematic areas. Research so far has mainly addressed the evaluation aspect providing assessment instruments that appear to correlate strongly with human ratings of dysarthric speech (Janbakhshi et al., 2019; Joshy & Rajan, 2022; Middag et al., 2009, 2011, 2014). Although such instruments are very valuable for speech therapy, more advanced technology that provides more detailed information about the problematic aspects of dysarthric speech is needed. This can then be used as a basis for providing immediate, automatic feedback on the speech of individuals with dysarthria. Application of such technology in speech training in a dysarthric individual's home environment is also among the possibilities. However, such technology has, to our knowledge, not yet been developed. In the past, studies were conducted on mispronunciation detection in children (Bunnell et al., 2000; Yin et al., 2009), but these were limited to isolated words spoken by children with developmental disabilities. Pronunciation error detection metrics have received more attention in language learning research (Van Doremalen et al., 2013; Laborde et al., 2016; Cucchiarini & Strik, 2017; García et al., 2020; Bharati et al., 2023; Thi-Nhu Ngo et al., 2024; Muzakki Bashori & Cucchiarini, 2024; Lounis et al., 2024). Attempts have been made at investigating the application of one of these metrics, Goodness of Pronunciation (GoP) (Witt, 1999), to mispronunciation detection in dysarthric speech (Laborde et al., 2016; Pellegrini et al., 2014, 2015). The results were promising, but only recently a new study was published that researched GoP using deep learning neural networks (Yeo et al., 2023). Other recent publications included the detection of mispronunciations at an articulatory level (Lin, Wang, Dang, et al., 2020) and used utterance verification (Fritsch & Magimai-Doss, 2021). More research is still needed before these metrics can be used in a speech therapy application.

1.6. CHASING: CHAllenging Speech training In Neurological patients by interactive Gaming

1.6.1. CHASING rationale

In the previous sections of this chapter, several research areas were described. The introduction and first section described how dysarthria due to Parkinson's disease, stroke, and traumatic brain injury affects patients' speech and, consequently, their lives. It also addressed the growing incidence of these neurological disorders, which places a high burden on healthcare. Existing healthcare resources may become overstretched as there may not be enough speech therapists to match the increasing demand.

Additionally, continuation of intensive speech training after initial rehabilitation has shown to have positive effects on speech intelligibility (Beijer et al., 2014). For progressive neurological disorders, like Parkinson's disease it may prevent or at least delay the decline in speech intelligibility. From this, it could be argued that patients may need intensive speech training throughout their lives and that training programs should perhaps be designed for lifelong provision. Consequently, this will place an even higher burden on healthcare.

The CHASING project tried to determine whether there is a method which could reduce that burden by providing intensive speech training in dysarthric speakers' home environment. This way, such therapy is easy accessible as it does not require travelling to a rehabilitation centre. It is provided in a, for the dysarthric speaker, comfortable environment, but still under remote supervision of the speech therapist.

The remote nature of such an application requires some form of feedback on dysarthric speakers' speech other than that a speech therapist provides in face-to-face sessions. Dysarthric speakers' partners or family members can provide feedback, but a more objective method is preferred. Automatic feedback provided by speech analysis and recognition algorithms as described in section 1.4 and section 1.5 respectively, are suitable to perform this task.

Rehabilitation exercises often have a repetitive, 'drill-and-practice' nature, which does not contribute to a joyful experience when doing them. Results from the E-learning based speech therapy (EST) project (Beijer, 2012; Beijer et al., 2014) that included research into dysarthric speakers' appreciation for web-based speech training, indicated that the training program should be made more attractive and motivating. They also indicated that it was difficult to transfer the skills learned in the exercises to speech in daily situations. As described in section 1.3, making use of interactive (or serious) gaming can improve patients' motivation for training. Also, games have a greater potential for simulating speech in daily situations due to their design flexibility.

For the previous reasons, the current thesis goal is to combine intensive training with automatic feedback and serious gaming in one device, enabling its potential to improve dysarthric speakers' speech intelligibility in their home environment in a challenging and motivating way.

1.6.2. Research in CHASING

Multiple research disciplines should be involved in developing device applications that provide intensive speech training from a home environment by integrating serious gaming and automatic feedback on speech. For that reason, the CHASING program should include research into unanswered questions in many of those research disciplines.

One of these questions should address research into the characteristics of the target group, as obviously, they are the ones who will be the end users. This includes general characteristics like age, personal skills, and hobbies, but also physical as well as cognitive capabilities and especially, game preferences.

Important in this research is also the remote nature of the training device. Dysarthric speakers must be able to play the serious game independently to a large extent. Considering that many dysarthric speakers tend to experience motor, communicative and cognitive disabilities, research into the prerequisites that dysarthric speakers should meet in order to adequately play the game is warranted.

Another question addresses research into the properties of speech exercises. It involves the identification of properties that contribute to increasing speech intelligibility in patients' with dysarthria due to Parkinson's disease, stroke, and traumatic brain injury. These properties should then be used to design speech exercises that integrate properly with game concepts and stories. Together, the exercises should form a training program. To measure the efficacy of such a program, research into suitable outcome measures is required. Given their expertise on these topics, the St. Maartenskliniek rehabilitation centre[1] joined as a partner in the CHASING program.

In addition to the properties of speech exercises, research into how automatic feedback can be included in these exercises should also be part of the program. It should focus on identifying speech dimensions suitable for automatic feedback to improve speech intelligibility. This research should also include how such feedback can technically be provided and visualized such that dysarthric speakers are able to process it.

The research described in the previous paragraphs will result in criteria and constraints for the design of a serious game. Given those, how can a motivating serious game be designed and developed? To address this question, the potentials of applying a user-centered design methodology combined with game concept design research should be investigated. Waag Society's Creative Care Lab [2] partnered in the program to provide their expertise and cooperation on these topics.

As the previous paragraphs already imply, research in the CHASING program must be interdisciplinary. Figure 1.1 provides a visual overview of the multiple disciplines combined in the current thesis.

[1] The Sint Maartenskliniek rehabilitation centre, Ubbergen, The Netherlands, http://www.maartenskliniek.nl, last accessed on October 28, 2024.

[2] Waag Society's Creative Care Lab, Amsterdam, The Netherlands, https://www.waag.org/nl/project/chasing, last accessed on October 28, 2024.

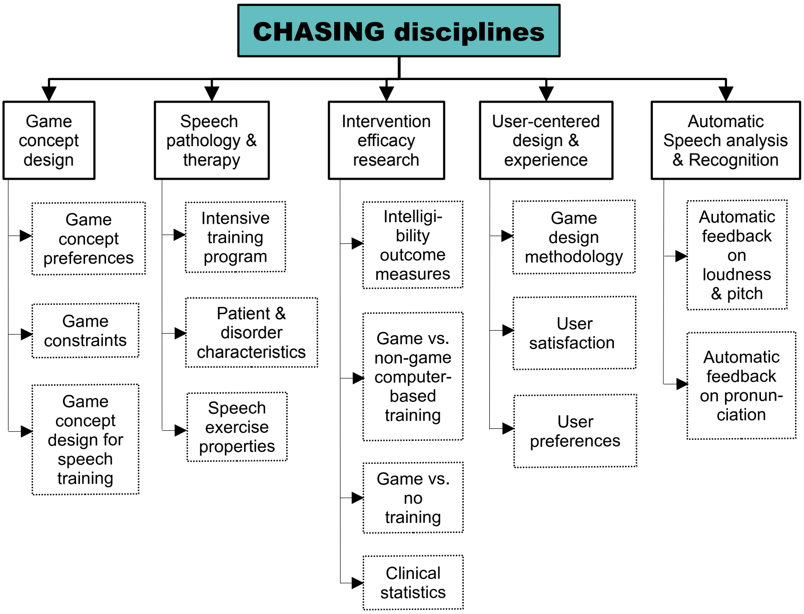

CHASING disciplines
Game concept design
Game concept preferences
Game constraints
Game concept design for speech training
Speech pathology & therapy
Intensive training program
Patient & disorder characteristics
Speech exercise properties
Intervention efficacy research
Intelligibility outcome measures
Game vs. non-game computer-based training
Game vs. no training
Clinical statistics
User-centered design & experience
Game design methodology
User satisfaction
User preferences
Automatic Speech analysis & Recognition
Automatic feedback on loudness & pitch
Automatic feedback on pronunciation

Figure 1.1.: Overview of the disciplines involved in the CHASING research program (uninterrupted boxes). The dotted boxes represent the research issues addressed in the current thesis.

1.7. Thesis outline

The current thesis addresses the use of serious gaming for speech training in patients with dysarthria due to Parkinson's disease, stroke or traumatic brain injury. Investigating its potentials for providing automatic feedback on speech is included. Previous research on serious gaming for rehabilitation showed that serious gaming could contribute to a more engaging and motivating physical therapy. In a similar way, serious games could do the same for speech therapy and provide an intensive speech training that also positively affects speech intelligibility. In this perspective, the following research questions are addressed in the current thesis:

1. Can speech intelligibility measures be obtained that provide evaluations at multiple levels of detail for the outcomes of different types of therapy?
2. What are ways to effectively improve the automatic recognition of dysarthric speech due to Parkinson's disease, stroke, and traumatic brain injury?
3. How can a serious game that provides automatic feedback on speech be designed for a speech intervention at dysarthric speakers' home environment?
4. How does game-based speech training compare to non-game computer-based

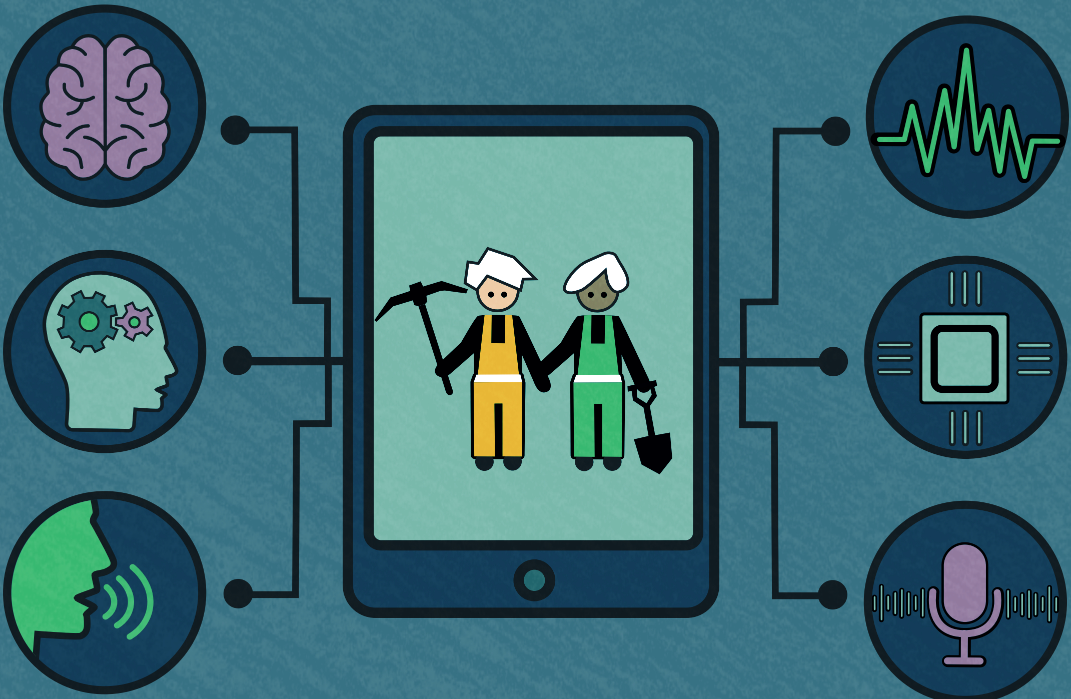
speech training with respect to speech intelligibility outcomes and patient satisfaction?

5. Can a game-based speech training positively affect speech intelligibility in patients with dysarthria?
6. Do patients with dysarthria prefer game-based speech training to traditional face-to-face training?

The current thesis consists of four main parts. The first part contains the current chapter, as a general introduction to the current thesis. Part two (chapters 2 to 4), includes the studies that address research and development of technology required for realizing a serious game that can provide and monitor a speech intervention remotely. In chapter 2, addressing research question 1, research is reported on objective measurements for the intelligibility of dysarthric speech. This chapter discusses the need for more detailed measures in addition to the existing ones at speaker and utterance level. Measures that include intelligibility ratings at three levels of granularity are investigated. In chapter 3, addressing research question 2, results are reported on a method to improve automatic recognition of dysarthric speech in low-resourced languages by including non-disordered speech data of a different language variety. It is hypothesized that the acoustic models trained on the increased variety of the target language will be beneficial to recognizing dysarthric speech which also has a large variability due to the nature and severity of the disorders. In chapter 4, also addressing research question 2, this research is continued by reporting on an investigation into a multi-stage scheme for training the acoustic models. First, a background model is trained on non-disordered speech with or without data from language varieties and secondly, it is retrained with target dysarthric speech data and data that potentially resembles it. The research into improving dysarthric speech recognition is important to validly provide automatic feedback.

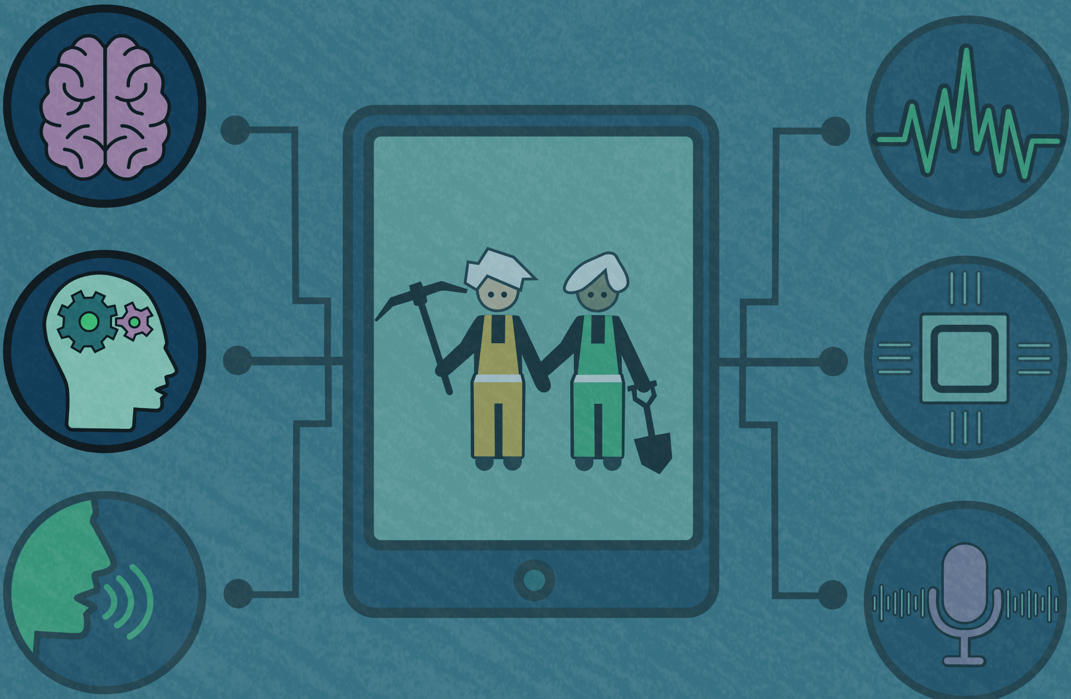
In the third part of the current thesis (chapters 5 to 7), research on the design and development of serious game concepts for speech training is addressed. Also, the efficacy of game-based speech training is explored. In chapter 5, that addresses research question 3, the design of the game is described, the process that was followed, and the lessons that were learned about designing games for speech training including the target group. Subsequently, in chapter 6, the first version of the serious game 'Treasure Hunters' is part of a within-subjects experiment to compare the serious game to a non-game computer-based speech training system on speech intelligibility, user satisfaction, and user preference. The therapeutic efficacy and patient satisfaction of a second, improved version of the serious game 'Treasure Hunters' are explored in chapter 7. Both chapters 6 and 7 address research questions 4 to 6.

The fourth and last part of the current thesis, consisting of chapter 8, discusses the results of the studies in the previous chapters with respect to the above-mentioned research questions. In addition, the limitations of the current thesis' research are discussed and suggestions are given to resolve them. To conclude, directions for future research are described.



Part II.

Developing speech technology in a serious game for speech training



Chapter 2

Intelligibility of Disordered Speech: Global and Detailed Scores

This chapter has been reformatted and slightly modified from:
*Ganzeboom, M., Bakker, M., Cucchiarini, C., Strik, H. (2016)
Intelligibility of Disordered Speech: Global and Detailed Scores.
Proceedings of Interspeech 2016, 2503-2507,
doi: 10.21437/Interspeech.2016-1448.*

2.1. Introduction

In the clinical practice of speech therapy it is often necessary to establish to what degree a patient's speech is intelligible. According to Hustad (2008) "Intelligibility refers to how well a speaker's acoustic signal can be accurately recovered by a listener". Assessments of intelligibility can be used for diagnostic purposes, but also to determine the degree of progress a patient has made. Similarly, in many lines of research on pathological speech it is necessary to assess patients' speech intelligibility, for instance to gauge the effectiveness of a specific treatment. Speech intelligibility has been studied not only in speech pathology research, but also in many other fields, such as second language (L2) pronunciation (Munro & Derwing, 1995), speech synthesis evaluation (Benoît et al., 1996; Gibbon et al., 1998) and speech perception in adverse conditions (Cooke et al., 2013; Cutler et al., 2008). In spite of the considerable attention intelligibility scoring has received, many aspects are still unclear. In this paper we try to gain more insight into speech intelligibility scoring by investigating measures with different degrees of granularity.

Speech intelligibility is usually measured by collecting subjective judgements by human raters. Because these are by definition subjective, they should preferably be collected from multiple raters, after which average ratings and reliability measures are calculated. Subjective ratings of intelligibility can take many different forms (Barreto & Ortiz, 2008; Miller, 2013; Yorkston & Beukelman, 1978). A common practice is to ask raters to indicate the degree of intelligibility on a scale, such as an equal-appearing interval scale (or Likert scale; for example in Yorkston and Beukelman (1978)), or a visual analogue scale (VAS), (placing a point on a horizontal line to indicate the degree of intelligibility; for example in Finizia et al. (1998)). Although this procedure may provide reliable ratings, it is not clear to what extent these ratings are valid representations of intelligibility, because there is no ground truth. Second, scale ratings are generally collected at the speaker or utterance level, and thus provide relatively broad measures of intelligibility.

An alternative, and in a sense more valid procedure to measure intelligibility, consists in asking subjects to make orthographic transcriptions, i.e. to listen to speech fragments and write down what they hear (Hustad, 2006; Laures & Weismer, 1999; Munro & Derwing, 1995). For this form of intelligibility measurement, different types of speech material can be used, including isolated words or pseudowords, whole sentences, and Semantically Unpredictable Sentences (SUS) Benoît et al. (1996). All these types of materials have their own pros and cons. The advantage of using isolated words and pseudowords is that in this case the effect of context can be minimized. Isolated words and pseudowords have also been used to obtain more detailed scores at the word and even the phoneme level, e.g. by having experts write down or select the phoneme that was heard in a specific position in a certain (pseudo)word (Kent et al., 1989; Van Nuffelen et al., 2009; Ziegler & Zierdt, 2008). However, isolated words and pseudowords constitute a rather unnatural context, and it is unclear how the identification of specific phonemes in isolated (pseudo)words relates to speech intelligibility in a more natural context. In a sense, ratings of phonemes in isolated (pseudo)words are comparable to phonemic or phonetic transcriptions, where expert transcribers are supposed to indicate, as much as possible, how speech sounds have been realized, thus approximating an articulatory description of the sounds. However, it is questionable whether discrepancies observed between such phonetic transcriptions of the realized utterances and the corresponding canonical or reference transcriptions can be taken as measures of speech intelligibility, which is supposed to indicate to what extent a given utterance has been understood by a listener (Hustad, 2008). A similar discussion has been going on in the field of L2 pronunciation instruction, where a distinction has been made between measures of accentedness (as opposed to nativeness) and measures of intelligibility (Levis, 2005; Munro & Derwing, 1995). Although accentedness and intelligibility appear to be related, they are distinct, partly independent dimensions. A relevant finding in this respect is that speech that is rated as heavily accented can still be intelligible (Munro & Derwing, 1995).

Having listeners orthographically transcribe whole sentences instead of isolated words, seems preferable, because sentences constitute more natural speech material. Yet, sentences have the disadvantage that the contextual information they contain may facilitate comprehension. According to Yorkston and Beukelman (1978), in this case we would be measuring comprehensibility instead of intelligibility. To circumvent this problem, Semantically Unpredictable Sentences (SUS) have been proposed, which are syntactically correct, but semantically incoherent sentences (Beijer, 2012; Benoît et al., 1996).

In general, orthographic transcriptions of regular or SUS sentences are scored at the word level: each word is scored as either correct or incorrect (Hustad, 2006; Beijer, 2012). Yet, both Hustad et al. (Hustad, 2008) and Beijer et al. (Beijer et al., 2014) argue that such word level scoring may still be quite broad, suggesting that it may be necessary to also collect judgements at even finer levels of granularity, i.e. the subword level. Intelligibility judgements on the subword level might indeed provide more detailed information about specific speech errors and may be more sensitive to changes within patients, enabling easier detection of treatment effects.

Next to human ratings of intelligibility, attempts have been made at developing objective measures of intelligibility that do not rely on human judgements. Many have employed ASR algorithms to obtain automatic measures of pathological speech quality (Berisha et al., 2013; M. J. Kim et al., 2015; Middag et al., 2009; Pellegrini et al., 2015; Schuster et al., 2006). It is, however, unclear to what extent such ASR-based metrics are valid representations of intelligibility. Firstly, they are often evaluated through comparison with benchmarks formed by human scale ratings or phonemic annotations. As explained above, it is not clear whether these benchmarks are themselves valid indicators of intelligibility. Secondly, while the automatic scoring methods that have been proposed so far are very interesting from a research point of view, they do not yet provide easy to use tools for clinical practice.

To summarize, in spite of the large body of research that has addressed intelligibility scoring of pathological speech, various issues still need clarification. The research reported in this paper aimed to contribute to this debate by investigating intelligibility measures with different degrees of granularity. We propose a procedure to automatically derive subword level intelligibility scores, i.e. scores at the phoneme and grapheme level, from orthographic transcriptions. The question we address is to what extent these subword intelligibility scores are reliable and how they relate to word level measures and utterance level ratings of intelligibility. In the following, we describe the procedures we used in collecting intelligibility evaluations of pathological speech on different levels of granularity (section 2.2), we present the results (section 2.3) and we discuss our findings (section 2.4).

2.2. Method

An online listening experiment was set up to compare evaluations of speech intelligibility of dysarthric speech on three different levels of granularity: utterance level, word level, and subword level. Utterance level evaluations were obtained using subjective rating scales (VAS and Likert scale); word and subword level evaluations were obtained using orthographic transcriptions, which were scored on both word and subword level.

2.2.1. Speakers and speech material

The speech material used in the study was selected from the recordings collected by Beijer (2012), from dysarthric speakers prior to speech therapy. To avoid speaker familiarity influencing the evaluation procedure, materials from seven different speakers were used. These were all male and suffered from hypokinetic dysarthria caused by Parkinson's disease. To investigate the different levels of granularity in intelligibility evaluation for a broad range of speech material, four different types of recordings were used: lists of single words, declarative SUS sentences, interrogative SUS sentences, and regular sentences. All samples consisted of existing Dutch words. The word lists contained three or five words, the SUS sentences all contained six words, and the length of the regular sentences varied between five and eight words. Table 2.1 provides an overview of types of speech material available per speaker.

Type of speech material	Speaker	Speech fragments
Word lists	S1	5 word lists (5 words each)
	S2	5 word lists (3 words each)
Declarative SUS sentences	S3	6 sentences
	S4	6 sentences
Interrogative SUS sentences	S5	6 sentences
	S6	6 sentences
Regular sentences	S7	8 sentences
	S1	8 sentences

Table 2.1.: *Overview of speech material used.*

Speech fragments with different levels of intelligibility, from low to high, were selected based on annotations by two listeners who did not participate in the current experiment.

2.2.2. Raters

Participants were invited by email or via Facebook. They filled in a questionnaire asking about mother tongue, gender, age, and familiarity with dysarthric speech. In total 36 listeners participated, 8 male and 28 female (age range 19-73). All listeners were native speakers of Dutch. Of the listeners, 31 had no experience with dysarthric speech and 5 had had the opportunity of listening to dysarthric speech before.

2.2.3. Procedure

The listening experiment was set up as an online experiment using the LimeSurvey application (LimeSurvey Project Team and C. Schmitz, 2015). Listeners could participate by accessing the experiment through a link. At the beginning of the experiment the listeners filled in a questionnaire (see subsection 2.2.2), and were informed about the task and the types of speech material to be rated. In addition, they were told that they would hear only real Dutch words. Then they had to rate three example speech fragments, aimed to familiarize them with dysarthric speech and the rating procedures. These examples were specially selected to contain low and high intelligible speech, in order to give raters an idea of the intelligibility range they could expect and to stimulate them to use the whole range of the rating scales.

The raters had to evaluate each of the 50 speech fragments in three different ways: two subjective sentence level ratings – a Likert scale and a Visual Analogue Scale (VAS) – and an orthographic transcription. Every screen presented to the listeners contained one speech fragment and the accompanying three evaluation methods. Orthographic transcription was done by typing in a textbox what was heard. The Likert scale ranged from 1 (“very low intelligibility”) to 7 (“very high intelligibility”).

The VAS was implemented as a slider that could be positioned on any number between 0 (“very low intelligibility”) and 100 (“very high intelligibility”). The order of the questions on the screen varied: for half of the fragments the orthographic transcription task was placed at the top, for the other half of the fragments, the two subjective rating scales were placed at the top. However, since all three scales were presented simultaneously on one screen, they could be answered in any order.

The raters could listen to the speech fragments multiple times before scoring or transcribing them. The 50 speech fragments (screens) were presented in a random order. On average it took the raters 20 minutes to rate all the material.

2.2.4. Calculating intelligibility scores

This subsection describes how we calculated the intelligibility scores from the raw judgements and transcriptions of the raters.

Intelligibility scores on utterance level

Intelligibility scores on utterance level were calculated as scores representing a percentage of intelligibility, ranging from 0 to 100. The VAS scores were already on a 0-100 scale, and were left unchanged. The scores on the Likert scale, ranging from 1 to 7, were transformed to percentage scores by first subtracting 1 and then multiplying by 16.67 (i.e. 1=0%, 2=16.67%, 3=33%, . . . , 7=100%).

Intelligibility scores on word level

The raters’ orthographic transcriptions were compared to the reference transcriptions and the number of identical word matches was counted. Subsequently, a percentage correct score was calculated.

Intelligibility scores on subword level

Intelligibility scores at the grapheme and phoneme level were automatically obtained from the orthographic transcriptions. For both the phoneme and grapheme level the Algorithm for Dynamic Alignment of Phonetic Transcriptions (ADAPT) (Elffers et al., 2005) was used. ADAPT computes the optimal alignment between two strings of phonetic symbols using a matrix that contains distances between the individual phonetic symbols. These distances are defined in terms of articulatory features and result in a distance measure expressing the phonetic similarity between the aligned transcriptions.

Listeners were instructed to not use any punctuation characters in their transcriptions. The punctuation characters we did find were removed and numerals were written out in words, which resulted in corrected orthographic transcriptions.

For the intelligibility scores on phoneme level, the orthographic transcriptions were converted to their phonemic equivalent using the canonical pronunciation variants from the lexicon of the Spoken Dutch Corpus (Oostdijk, 2000). Words that were not contained in the lexicon were manually added. As spelling errors complicated the lookup in the lexicon, those that did not affect the phonemic transcription were manually corrected. The resulting phonemic transcriptions were converted to the ADAPT symbol set (see appendix A in Elffers et al. (2005)). The ADAPT alignment algorithm and distance matrix were applied unchanged.

For the intelligibility scores on grapheme level, the phonetic symbols in the ADAPT distance matrix were replaced by the Dutch graphemes. The values of the graphemes 'articulatory feature' columns were all set to 0.0 except for the diagonals, which were set to 1.0. Using this matrix, the algorithm aligned the orthographic transcription with the reference transcription and calculated the distance between them. Each insertion and deletion was graded with distance 2.0 and substitution with distance 3.0.

2.3. Results

In total, five measures of intelligibility were collected for each speech fragment: two scale ratings on utterance level (Likert scale and VAS), a word level scoring of the orthographic transcription (OTW), and two subword level scorings of the orthographic transcriptions, at phoneme level (OTP) and at grapheme level (OTG). In this section we present the results regarding the reliability of these measures, and their relations.

2.3.1. Reliability

The reliability of each of the five intelligibility measures was calculated using Intraclass Correlation Coefficients (ICC) based on groups of raters, as we do not intend to devise an intelligibility measure relying on the judgements of a single rater. The ICC values for all 36 raters together were very high, ranging from .95 (OTP, OTG) to .97 (Likert, VAS, OTW). As such a large number of raters may not always be achievable, we also calculated ICCs based on smaller samples of raters, randomly drawn from our sample of 36 (for each sample size 10 random samples were drawn, and average ICCs were calculated). On average, for the utterance and word level scorings sufficient reliability is obtained with four raters (resulting in mean ICC values ranging from .79 to .84), while for subword scorings at least six raters are required (resulting in mean ICC values ranging from .79 to .80).

2.3.2. Intelligibility scores and correlations

Intelligibility scores for each fragment were calculated by averaging over the 36 raters. Mean scores for the five intelligibility measures, and the correlations between them are shown in Table 2.2. Correlations between all measures were significant (p < .01). The two utterance level measures were very highly correlated (r = .998), and the correlation between the two subword level measures was also very high (r = .954).

	M (SD)	Correlations (Pearson r)			
		VAS	OTW	OTP	OTG
Likert	63.1 (21.1)	.998	.733	-.763	-.773
VAS	63.2 (19.0)		.732	-.755	-.764
OTW	78.3 (16.1)			-.805	-.869
OTP	8.0 (6.5)				.954
OTG	8.9 (7.4)				

Table 2.2.: *Means (SDs) and correlations of the five intelligibility measures (n = 50 speech fragments).*
VAS: Visual Analogue Scale,
OTW: Orthographic Transcription at Word level,
OTP: Orthographic Transcription at Phoneme level,
OTG: Orthographic Transcription at Grapheme level.
For Likert, VAS and OTW, higher scores correspond to higher intelligibility (higher percentage correct); for OTP and OTG higher scores correspond to lower intelligibility (higher distance). All correlations were significant (p < .01).

To be able to directly compare subword level scores to word and utterance level percentage scores, we transformed the subword scores to percentage correct scores (phonemes or graphemes) in the utterance. Using an ANOVA with the five intelligibility measures as a within subject factor and percentage score as the dependent variable, we found significant differences between the different measures (F(4,46) = 80.57, p < .01). The percentage scores were significantly higher (p < .01) on the word level (M = 78.3) than on the utterance level (Likert: M = 63.1, VAS: M = 63.2), while the percent correct scores on the subword level were significantly higher (p < .01) than scores on the word level: 87.3 (SD = 10.2) for the phoneme level and 85.5 (SD = 11.8) the grapheme level. The differences between the Likert and VAS scores were not significant (p > .05).

2.4. Discussion

2.4.1. Feasibility and reliability of subword scoring

In this paper, we calculated subword level scores by automatically processing the orthographic transcriptions. The ADAPT algorithm (Elffers et al., 2005) used for this purpose only requires two text files as input, i.e. all orthographic transcriptions and the reference transcriptions. Both are formatted to contain a single transcription per line. The results of the alignments are stored in a comma-separated text file that allows easy viewing and import into spreadsheet software, making it feasible and relatively easy to use in clinical and research contexts.

Results showed that subword level intelligibility scorings are slightly less reliable than scorings at the utterance or word level. This can be explained by the effect of chance agreement (Tinsley & Weiss, 1975): since phoneme and grapheme scorings have a higher level of detail, they allow for more variation. Yet, when using at least six (inexperienced) raters, sufficiently reliable phoneme and grapheme scorings can be obtained. A sample of six raters is quite feasible in most research, and the fact that orthographic transcription tasks can easily be performed online makes it less problematic to involve multiple raters. With expert raters (e.g., speech-language pathologists), one would expect higher reliability, and the required number of raters might be lower. This should be verified in future research.

2.4.2. Comparisons between scores at different levels of granularity

The results show that intelligibility measures of different levels of granularity are fairly highly correlated, which is a reassuring outcome. When comparing the percentage scores obtained for the different measures, results show, first, that word level scoring produces higher intelligibility scores than utterance level scoring. This is in line with earlier research of Hustad (2006) suggesting that when using subjective rating scales, raters tend to underestimate the extent to which speakers are intelligible. A rater may, for example, understand every word, but still judge intelligibility as less than perfect when higher-than-normal listening effort is required because of articulatory irregularities. This again suggests that orthographic transcriptions may be more objective or valid measures of intelligibility while subjective rating scales indicate comprehensibility as defined by Munro and Derwing (1995): the difficulty a listener experiences in understanding an utterance.

A second finding is that percentage scores obtained with subword scorings are higher than those obtained with word level scoring. This is a natural consequence of the fact that with subword scoring, words that are not understood correctly can still be scored as partly correct: the transcription <stad> of the prompt <stap> would be incorrect at word level while at subword level only 1 of the 4 graphemes would be incorrect. In fact, subword level scores measure a different dimension of intelligibility, focusing on intelligibility of parts of words (graphemes, phonemes) rather than entire words.

2.4.3. Phoneme vs. grapheme scoring

We investigated two types of subword scorings, phoneme and grapheme scoring, in an attempt to obtain a more fine-grained measure of intelligibility. Clearly, phoneme scoring can provide more accurate indications of specific articulation problems of speakers, as phonemes directly represent speech sounds, whereas the connection between graphemes and speech sounds is blurred by spelling conventions (e.g. when two graphemes correspond to one phoneme, such as <oe> - /u/, <ng> - /ŋ/).

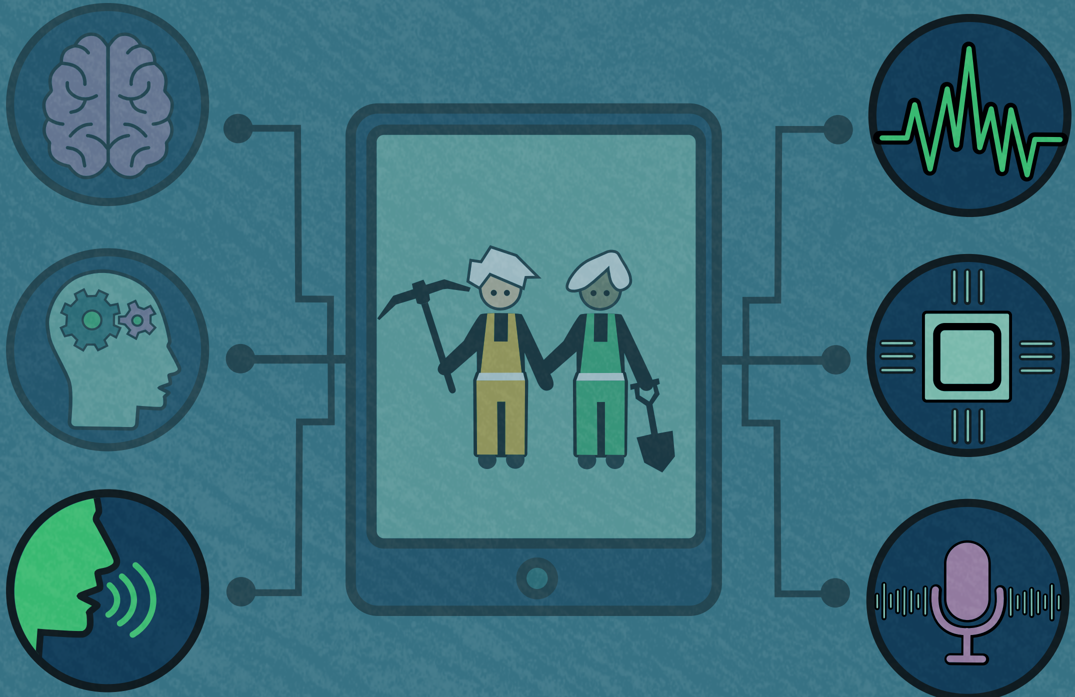
However, phoneme scoring is less easily performed, as it requires that a grapheme string is converted to a phoneme string, e.g. by means of a lexicon look-up (as we did in the current study) or a grapheme-to-phoneme conversion algorithm. The high correlation between the overall distance scores obtained from phoneme and grapheme scorings suggests that, at least when focusing on intelligibility per se, and not on specific articulation problems, grapheme scoring is a suitable alternative for phoneme scoring in cases where phoneme scoring is less feasible.

2.5. Conclusions

We conclude that automatic scoring of orthographic transcriptions on the subword level is a feasible way of obtaining a more fine-grained measure of intelligibility. While scoring at the phoneme level is more informative with respect to identifying specific articulation problems, scoring at the grapheme level is a reasonable alternative in cases where phoneme scoring is not possible, i.e. in clinical practice.

One can argue that utterance, word, and subword level scorings of intelligibility each measure a different dimension of intelligibility, with subjective utterance level ratings measuring comprehensibility as defined in Munro and Derwing (1995), word level scorings of orthographic transcriptions measuring actual intelligibility of words, and subword level scorings measuring intelligibility of parts of words. As each of these dimensions of intelligibility are relevant in both clinical practice and research contexts, we suggest the use of subword scorings as a supplement to utterance level and word level scorings, not as a replacement. While subword scorings may require a few extra raters to achieve reliable measures, they provide worthwhile information at a finer level of granularity.[1]

[1] **Acknowledgements**
The authors would like to thank Lilian Beijer of the St. Maartenskliniek rehabilitation centre for collecting the dysarthric speech fragments used in this research. Thanks also extend to Ilona van der Linden and Sandra Schoemaker for collecting the listener judgements and their initial analyses.



Chapter 3

Combining non-pathological data of different language varieties to improve DNN-HMM performance on pathological speech

This chapter has been reformatted and slightly modified from:
Yılmaz, E., Ganzeboom, M., Cucchiarini, C., Strik, H. (2016)
Combining Non-Pathological Data of Different Language Varieties to Improve DNN-HMM Performance on Pathological Speech.
Proceedings of Interspeech 2016, 218-222,
doi: 10.21437/Interspeech.2016-109.

3.1. Introduction

Motor speech disorders including dysarthria caused by neuromuscular control problems (Duffy, 2019) lead to decreased speech intelligibility and communication impairment (Kent & Kim, 2003). Consequently, the life quality of dysarthric patients is negatively affected (Walshe & Miller, 2011) and they run the risk of losing contact with friends and relatives and eventually becoming isolated from the society.

Research has shown that intensive therapy can be effective in (speech) motor rehabilitation (Bhogal et al., 2003; Kwakkel, 2006; Ramig et al., 2001; Rijntjes et al., 2009), but various factors conspire to make intensive therapy expensive and difficult to obtain. Recent developments show that therapy can be provided without resorting to frequent face-to-face sessions with therapists by employing computer-assisted speech training systems (Beijer & Rietveld, 2011). According to the outcomes of the efficacy tests presented in (Beijer et al., 2014), the user satisfaction appears to be quite high. However, most of these systems are not yet capable of automatically detecting problems at the level of individual speech sounds, which are known to have an impact on speech intelligibility (De Bodt et al., 2002; Popovici and Buică-Belciu, 2012; Van Nuffelen et al., 2009; Yunusova et al., 2005; chapter 2).

Despite long-lasting efforts to build speaker- and text-independent ASR system for people with dysarthria (cf. section 3.2), the performance of state-of-the-art systems is still much worse on this type of speech than on normal speech. One main reason is the lack of pathological speech data to train automatic speech recognition systems which can provide accurate enough recognition and speech assessment.

Training deep neural networks (DNN)-based acoustic models on large amount of pathological data to capture the within- and between-speaker variation is generally not feasible due to the limited size and structure of existing pathological speech databases. The number of recordings in dysarthric speech databases is much smaller compared to normal speech databases. Moreover, these databases contain mostly very restricted speech tasks such as reading out word and sentence lists with varying linguistic complexity.

As a remedy, combining in-domain and out-of-domain English speech data to train DNNs used for feature extraction has been proposed in (Christensen et al., 2013). In this chapter, we describe another such solution to train a better DNN-hidden Markov model (HMM) system for the Dutch language which has fewer speakers and resources compared to English. We investigate combining non-dysarthric speech data from different varieties of the Dutch language to train more reliable acoustic models for a DNN-HMM ASR system. The included varieties are Northern Dutch and Flemish (Southern Dutch) which have the same phonetic alphabet and share a large amount of vocabulary. Most prominent phonetic differences between these varieties are diphthongized long vowels of Northern Dutch and articulation of several consonants. This work has been done in the framework of the CHASING project[1],

[1] https://www.helmer-strik.nl/chasing/, last accessed on October 28, 2025.

in which a serious game employing ASR is being developed to provide additional speech therapy to dysarthric patients.

The rest of the chapter is organized as follows. In section 3.2 previous relevant work on ASR of dysarthric speech is described. In section 3.3 the rationale behind the selection of speech corpora is explained. A summary of the fundamentals of DNN-based ASR is provided in section 3.4. It also reports on the details of the DNN training scheme applied in this chapter. The experimental setup is described in section 3.5 and the recognition results are shown and discussed in section 3.6. The final section of this chapter (section 3.7) presents our conclusions based on the obtained results.

3.2. Related work

Several researchers have investigated ASR performance on pathological speech. In a very recent work, T. Lee et al. (2016) has reported the ASR performance on Cantonese aphasic speech and disordered voice. A generic DNN-HMM system provided significant improvements on disordered voice and minor improvements on aphasic speech compared to a GMM-HMM system. Takashima et al. (2015) proposed a new feature extraction scheme using convolutional bottleneck networks for dysarthric speech recognition. They tested the proposed approach on a small test set consisting of 3 repetitions of 216 words by a single male speaker with an articulation disorder and reported some gains over a system using MFCC features.

Shahamiri and Salim (2014) proposed an artificial neural network-based system trained on digit utterances from nine non-dysarthric and 13 dysarthric individuals affected by Cerebral Palsy (CP). They reported word recognition accuracies of 74.7%, 67.4% and 51.7% on mild (66-99% speech intelligibility), moderate (33-66% speech intelligibility) and high (less than 33% speech intelligibility) dysarthric speaker as an independent test set. Christensen et al. (2012) trained their models solely on 18 hours of speech of 15 dysarthric speakers due to CP leaving one speaker out as test set. The different degrees of severity were reported through classes and percent intelligibility scores from listening tests with unfamiliar listeners: ‘Very low 2-17%’, ‘Low 28-43%’, ‘Mid 58-62%’, and ‘High 86-95%’. There were 4, 3, 3, and 5 speakers in every class, respectively. Recognition results for ‘Very low’ ranged from 4.1-12.9%, ‘Low’ 7.0-22.2%, ‘Mid’ 30.3-50.2% , and ‘High’ 46.6-68.5%. This shows that there is overlap between classes and that the recognition results do not always exactly match the intelligibility scores given by listeners.

Rudzicz (2007) compared the performance of a speaker-dependent and a speaker-adaptive GMM-HMM systems on the Nemours database (Menéndez-Pidal et al., 1996). Their system was trained on the WSJ corpus. The test set consisted of speech from 11 dysarthric speakers due to CP or head trauma. Every speaker recorded 74 nonsense sentences. The recognizer provided recognition rates below

10% on the speech of 4 speakers with severe dysarthria. For moderately and mildly dysarthric speakers, recognition accuracy was between 11-30% and 31-60% respectively. Seong et al. (2012) proposed a weighted finite state transducer (WSFT)-based ASR correction technique applied to a recognition system trained also on the WSJ corpus. They reported an average accuracy of 47.1% when recognizing the speech of 10 dysarthric speakers from the same dysarthric database. Similar work had been proposed by Caballero-Morales and Cox (2009) previously.

Mengistu and Rudzicz (2011) combined dysarthric data of eight dysarthric speakers with that of seven normal speakers, leaving one out as test set and obtained an average increase by 13.0% in comparison to models trained on non-dysarthric speech only. They also noted that context-dependent HMMs showed little improvement over context-independent ones. In one of the earliest work on Dutch pathological speech by E. Sanders et al. (2002), a pilot study was presented on ASR of Dutch dysarthric speech data obtained from two speakers with a birth defect and a cerebrovascular accident. Both speakers were classified as mild dysarthric by a speech pathologist.

From the previous descriptions, it appears that it is difficult to fully compare results between these publications due to the differences in types of speech materials, types of dysarthria, reported severity, and dataset used for training and testing. Additionally, dysarthric speech is highly variable in nature, not only due to its various etiologies and degrees of severity, but also because of possible individually deviating speech characteristics. This may negatively influence the capability of speaker-independent systems to generalize over multiple dysarthric speakers.

Possible improvements may come from recent advances in DNN-based acoustic model yielding impressive results in the field of non-dysarthric speech recognition (Hinton et al., 2012). These results show promising increases in the speaker-independent recognition accuracies when compared to those obtained with traditional GMM-HMMs. Therefore, we investigate how the DNN-HMM system trained on normal speech perform on the recognition of dysarthric speech with a focus on the amount of available training data from different varieties of the Dutch language.

3.3. Speech corpora selection

Given the limited availability of dysarthric speech data, we investigate to what extent already existing databases of Dutch normal speech can be employed to train DNNs and optimize their performance on dysarthric speech. The ASR technology to be developed in the CHASING project is primarily intended for dysarthric patients who live in the Netherlands and speak the Northern Dutch variety. However, we thought it would be interesting to also investigate the usability of speech databases of the Southern variety of Dutch spoken in Flanders and also known as Flemish. First, because these two varieties are mutually intelligible and their phonetic alphabets are arguably similar, apart from some well described phonological and phonetic differences. Second, because using Flemish speech would open up the possibility of adapting the game that is now been developed for patients in the Netherlands for use by patients in Flanders.

Fortunately, there have been multiple Dutch-Flemish speech data collection efforts (Cucchiarini et al., 2008; Oostdijk, 2000) which facilitate the integration of both Dutch and Flemish data in the research reported in this chapter. For training purposes, we used the CGN corpus (Oostdijk, 2000), which contains representative collections of contemporary standard Dutch as spoken by adults in the Netherlands and Flanders. The components with read speech, spontaneous conversations, interviews and discussions are used for training the acoustic models in the present experiments.

For testing purposes, we decided to use the largest collection of pathological speech that is available for the Dutch language, the Flemish COPAS database (Middag, 2012). In the meantime, a database of Northern Dutch dysarthric speech has been compiled (Yılmaz et al., 2016). In the course of the CHASING project, this collection will be augmented with additional material and then used for further experiments to optimize ASR back-end used in the developed serious game.

The COPAS corpus contains recordings from 122 Flemish normal speakers and 197 Flemish speakers with speech disorders such as dysarthria, cleft, voice disorders, laryngectomy and glossectomy. The dysarthric speech component contains recordings from 75 Flemish patients affected by Parkinson's disease, traumatic brain injury, cerebrovascular accident and multiple sclerosis who exhibit dysarthria at different levels of severity.

The word reading tasks used in this chapter are taken from the Dutch Intelligibility Assessment (DIA) (De Bodt et al., 2006) material which contains 35 versions of 50 consonant-vowel-consonant (CVC) words organized in 3 subgroups. Moreover, all sentence reading tasks with annotations, namely 2 isolated sentence reading tasks, 11 text passages with reading level difficulty of AVI 7 and 8 and Text Marloes, are also included in the test data.

3.4. Training DNNs for Dysarthric Speech

3.4.1. Fundamentals of DNN-based ASR

A single artificial neuron, which is the basic element of the DNN structure, receives N input values $\mathbf{v} = [v_0, v_1, ..., v_{N-1}]$ with weights $\mathbf{w} = [w_0, w_1, ..., w_{N-1}]$ and an offset value b. To compute the neuron output y, a non-linear function f is applied the weighted sum z of all outputs of the previous layer and the offset, i.e., $y = f(z) = f(\mathbf{w}^T\mathbf{v} + b)$. A DNN consists of L layers of M artificial neurons and the output of the $(l-1)^{\text{th}}$ layer with M_{l-1} neurons is the input of the l^{th} layer with M_l neurons which is formulated as $\mathbf{v}_l = f(\mathbf{z}_l) = f(\mathbf{W}_l\mathbf{v}_{l-1} + \boldsymbol{b}_{\mathrm{l}})$ where the dimensions of $\mathbf{v}_l$, $\mathbf{W}_l$, $\mathbf{v}_{l-1}$ and$\boldsymbol{b}_{\mathrm{l}}$ are M_l, $(M_l \times M_{l-1})$, M_{l-1} and M_l respectively. M_0 is the number of neurons in the input layer which is equal to the dimension of the speech features. The non-linear activation function f maps an M_{l-1} vector to an M_{l-1} vector. The activation function applied at the output layer is the softmax function in order to get output values in the range $[0, 1]$ for the HMM state posterior probabilities

$$\mathbf{v}_{L+1} = P(q_i|\mathbf{o}) = \frac{e^{z_L^i}}{\sum_m^{M_{L+1}} e^{z_L^m}}, \tag{3.1}$$

where M_{L+1} is equal to the number of HMM states.

The DNN-HMM training is achieved in three main stages (Dahl et al., 2012; D. Yu & Deng, 2015). Firstly, a GMM-HMM setup is trained to obtain the structure of the DNN-HMM model, initial HMM transition probabilities and training labels of the DNNs. Then, the pretraining algorithm described in Hinton (2010) is applied to obtain a robust initialization for the DNN model. Finally, the back-propagation algorithm (Hecht-Nielsen, 1989) is applied to train the DNN that will be used as the emission distribution of the HMM states.

3.4.2. Tuning DNNs on Flemish Speech

The DNN training applied in this chapter is organized in two steps. In the first step, the DNN training is performed on the combined speech data containing Flemish and Northern Dutch normal speech. Both varieties sharing the phonetic alphabet, we learn several hidden layers and an output layer on both varieties with the aim of learning more reliable hidden layers. The amount of training data used during the initial training phase can be increased by including more speech data from different speech types such as elderly and children speech. In the scope of this work, we only consider using normal speech to analyze the impact of the data merging on the recognition performance.

In the second step, the softmax layer of this DNN is retrained only on the Flemish data. Retraining the softmax layer achieves the fine-tuning of the DNN targets on the target Flemish speech. This two-step training approach resembles the multilingual DNN training scheme for cross-lingual knowledge transfer which is commonly used for obtaining acoustic models for under-resourced languages (Swietojanski et al., 2012; Huang et al., 2013). In these studies, considerable improvements have been reported on both low- and high-resourced languages thanks to the hidden layers trained on multiple languages.

3.5. Experimental Setup

3.5.1. Databases

The CGN components with read speech, spontaneous conversations, interviews and discussions are used for acoustic model training. The duration of the normal Flemish (FL) and northern Dutch (NL) speech data used training is 186.5 and 255 hours respectively. The combined training data (FL+NL) contains 441.5 hours in total.

For testing the acoustic models, we have classified the speech material in the COPAS database based on the type of the speaker (normal vs. pathological) and speech material (word vs. sentence) resulting in 4 test sets. The speech segments in which the speaker do not utter the target word are discarded to be able to evaluate the recognizer errors only. There are 687 different words and 212 different sentences in the test data. The test set containing the word tasks uttered by normal speakers (WordNor) and speakers with disorders (WordDys) consists of 6154 and 8648 utterances with a total duration of 1.5 and 2 hours, respectively. The test set containing the sentence tasks uttered by normal speakers (SentNor) and speakers with disorders (SentDys) consists of 1918 (15,149) and 1034 (8287) sentences (words) with a total duration of 1.5 and 1 hour, respectively.

3.5.2. Implementation Details

The recognition experiments are performed using the Kaldi ASR toolkit (Povey et al., 2011). A standard feature extraction scheme is used by applying Hamming windowing with a frame length of 25 ms and frame shift of 10 ms. A conventional context dependent GMM-HMM system with 40k Gaussians and 5925 triphone states is trained on the 39-dimensional MFCC features including the deltas and delta-deltas. This system is used to obtain the state alignments required for DNN training. We also trained a GMM-HMM system on the LDA-MLLT features as a second baseline system.

The DNNs with 6 hidden layers and 2048 sigmoid hidden units at each hidden layer are trained on the 40-dimensional log-mel filterbank features with the deltas and delta-deltas. The DNN training is done by mini-batch Stochastic Gradient Descent with an initial learning rate of 0.008 and a minibatch size of 256. The time context size is 11 frames achieved by concatenating ±5 frames. A unigram (trigram) language model trained on the target transcriptions of the word (sentence) tasks is incorporated in the recognition of the word (sentence) tasks.

3.6. Results and Discussion

We have performed ASR experiments using the speech data described in Section subsection 3.5.1. The recognition results obtained on the word and sentence tasks uttered by normal and pathological speakers from the COPAS database are presented in the columns of Table 3.1. The lowest WER for each column is marked in bold. The recognition performance obtained on the sentence readings task is more relevant compared to isolated word recognition in the context of the developed CHASING serious game. We report results on both word and sentence task results for completeness.

Acoustic models	Training Data	WordDys	WordNor	SentDys	SentNor
GMM+MFCC	FL	77.2	56.1	38.2	13.0
GMM+MFCC	FL+NL	78.7	61.0	37.4	14.7
GMM+LDA-MLLT	FL	74.4	50.9	37.4	11.3
GMM+LDA-MLLT	FL+NL	74.9	55.0	37.0	12.5
DNN+FBANK	FL	65.0	37.9	28.1	4.7
DNN+FBANK (w/o retraining on FL)	FL+NL	64.9	38.4	26.8	4.7
DNN+FBANK (with retraining on FL)	FL+NL	**63.7**	**36.2**	**26.3**	**4.4**

Table 3.1.: *Word error rates in % obtained on the word and sentence COPAS test sets.*

The conventional GMM-HMM trained on FL data using the MFCC features provides a WER of 38.2% on the dysarthric sentence utterances and a WER of 77.2% on the dysarthric word utterances. This large gap in recognition accuracy obtained on the sentence and word recognition tasks is due to the very challenging nature of recognizing phonetically similar, monosyllabic words and pseudowords in isolation. The GMM-HMM system trained on FL+NL data reduces the normal speech recognition from 13.0% WER to 14.7% in sentence reading tasks and from 56.1% WER to 61.0% in word reading tasks, while increasing dysarthric sentence recognition accuracy from 38.2% to 37.4%. The performance reduction in normal speech is comprehensible, since adding Northern Dutch data increases the mismatch between the training and testing conditions. Training GMM-HMM on the combined data does not always improve dysarthric speech recognition with 0.8% decrease on sentence tasks and 1.5% increase on word tasks in the WER.

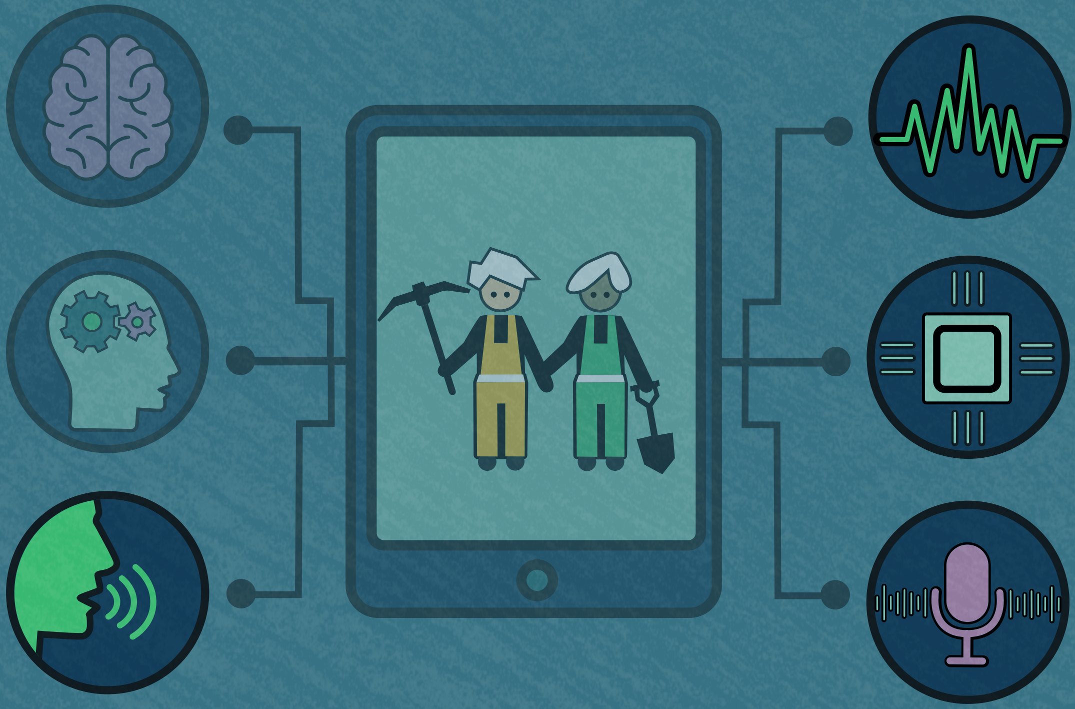
Compared to MFCC features, using LDA-MLLT-transformed features considerably reduces the WERs obtained on normal speech as expected. On the other hand, the gains obtained on pathological speech by using these features are minimal. This is due to the fact that there is a significant mismatch between the type of speech on which the transformation is learned and applied to in the case of pathological speech.

The DNNs trained on FBANK features provide considerable improvements on all speech types and reading tasks. These improvements are as large as 8.9% WER on dysarthric sentence utterances and 9.9% WER on dysarthric word utterances. Training the DNN-HMMs on FL+NL data improves the performance on dysarthric sentence reading tasks with an absolute improvement of 1.3% without retraining the softmax layer on Flemish data. The same system does not improve the recognition accuracy of dysarthric word reading tasks. After the final step of applying softmax layer retraining for tuning the DNN targets to Flemish phones, we can get an improved recognition performance in all cases compared to the baseline DNN system trained only on Flemish data. To be specific, the WER obtained on dysarthric sentence reading task decreases from 28.1% to 26.3%, while the WER obtained on dysarthric word reading task reduces from 65.0% to 63.7%.

It is relevant to point out that, although some of the above differences in WER may be small, they do show that training DNN-HMM systems on training data containing speech from different varieties of a language can improve the recognition performance at moderate levels. Despite the large gap between the performance on pathological and normal speech, the presented speaker-independent recognition results obtained on different types of pathological speech at different severity levels are encouraging. Building text- and speaker-independent ASR systems that can be used in clinical applications appears to be within reach, even for languages with more limited resources than English.

3.7. Conclusions

In this chapter, we have investigated combining speech data from different varieties of a mid-sized language for training a DNN-HMM system. The DNN-based acoustic models were trained on normal Flemish and Northern Dutch speech and speaker-independent recognition experiments were performed on pathological Flemish speech. The results have shown that the proposed data merging technique in this context can improve the recognition of pathological speech, especially after a second training phase in which the DNN targets are tuned to the phones of the specific variety involved in the testing setup, Flemish in this case. These results are promising for future work aiming to develop useful ASR-based pathological speech applications for languages that are smaller in size and less resourced than English.



Chapter 4

Multi-stage DNN training for automatic recognition of dysarthric speech

This chapter has been reformatted and slightly modified from:
Yılmaz, E., Ganzeboom, M., Cucchiarini, C., Strik, H. (2017)
Multi-Stage DNN Training for Automatic Recognition of Dysarthric Speech.
Proceedings of Interspeech 2017, 2685-2689,
doi: 10.21437/Interspeech.2017-303.

4.1. Introduction

Speech disorders caused by neuromuscular control problems (Duffy, 2019) like dysarthria can reduce speech intelligibility and cause communication impairment (Kent & Kim, 2003). This can negatively affect the life quality of dysarthric patients (Walshe & Miller, 2011) who run the risk of losing social contact and eventually becoming isolated from society. Recent research has shown that intensive therapy can be effective in (speech) motor rehabilitation (Bhogal et al., 2003; Kwakkel, 2006; Ramig et al., 2001; Rijntjes et al., 2009). Conventional speech therapy provided by a speech therapist is costly. Recent developments show that therapy can be provided by employing computer-assisted speech training systems (Beijer & Rietveld, 2011). According to the outcomes of efficacy research on computer-assisted speech training systems (Beijer et al., 2014), user satisfaction towards such a system appears to be quite high. However, most of these systems are not yet capable of automatically detecting problems at the level of individual speech sounds, which are known to have an impact on speech intelligibility (chapter 2, De Bodt et al., 2002; Popovici and Buică-Belciu, 2012; Van Nuffelen et al., 2009; Yunusova et al., 2005). Our goal is to develop more robust acoustic models for pathological speech and incorporate automatic speech recognition (ASR) technology to detect these problems.

Despite long-lasting efforts to build speaker- and text-independent ASR systems for people with dysarthria, the performance of state-of-the-art systems is still considerably lower on this type of speech than on normal speech. Past ASR experiments on dysarthric speech mostly included GMM-HMM systems (Christensen et al., 2012; Menéndez-Pidal et al., 1996; Mengistu & Rudzicz, 2011; Rudzicz, 2007; E. Sanders et al., 2002; Shahamiri & Salim, 2014). More recently T. Lee et al. (2016) reported ASR performance on Cantonese aphasic speech and disordered voice. A generic DNN-HMM system provided significant improvements on disordered voice and minor improvements on aphasic speech compared to a GMM-HMM system. Takashima et al. (2015) proposed a new feature extraction scheme using convolutional bottleneck networks for dysarthric speech recognition.

Training robust deep neural networks (DNN)-based acoustic models to capture the within- and between-speaker variation in dysarthric speech is generally not feasible due to the limited size and structure of existing pathological speech databases. The number of recordings in dysarthric speech databases is much smaller compared to that in normal speech databases. Moreover, these databases mostly contain very restricted speech tasks such as reading out word and sentence lists with varying linguistic complexity.

To remedy the data scarcity problem, in-domain and out-of-domain English speech data were combined to train DNNs for improved feature extraction (Christensen et al., 2013). In chapter 3, we described a similar solution to train a better DNN-hidden Markov model (HMM) system for the Dutch language, a language that has fewer speakers and resources compared to English. In particular, we investigated combining non-dysarthric speech data from different varieties of the Dutch language to train more reliable acoustic models for a DNN-HMM ASR system. This work was conducted in the framework of the CHASING project[1], in which a serious game employing ASR is being developed to provide additional speech therapy to dysarthric patients. In this research we employed a 6-hour Dutch dysarthric speech database that had been collected in a previous project (E-learning-based Speech Training, EST, Yılmaz et al., 2016). The serious game developed in the CHASING project also serves as a useful data collection tool for pathological speech research. The dysarthric speech material recently collected during the CHASING field studies, which we refer to as the CHASING01 speech database, is used for testing, while the EST database is employed for training purposes.

In the present work, we apply a multi-stage DNN training procedure using a large amount of out-of-domain and a small amount of in-domain data. A two-stage version of this training procedure has been applied to multilingual training of DNNs which is commonly used to obtain acoustic models for under-resourced languages (Huang et al., 2013; Swietojanski et al., 2012). In these studies, considerable improvements have been reported on both low- and high-resourced languages thanks to the hidden layers trained on multiple languages.

In the first stage of the training, we train models on normal Dutch and Flemish speech, which, in chapter 3 has been shown to provide improved recognition of dysarthric speech compared to training on only one variety. The background model obtained in the first stage is retrained on normal and adult Dutch speech only for language adaptation and the EST dysarthric speech database is used for domain adaptation in subsequent training stages. The final models are then applied to the recently collected dysarthric speech data from the CHASING01 database.

The rest of this chapter is organized as follows. In section 4.2 details of the DNN training scheme applied in this chapter are described. The selection of various speech corpora for the proposed training scheme is explained in section 4.3. The experimental setup is described in section 4.4 and the recognition results are shown and discussed in section 4.5. The final section of this chapter (section 4.6) presents our conclusions based on the obtained results.

4.2. Multi-stage DNN training

The DNN training applied in this chapter is organized in multiple steps. In the first step, a background DNN is trained on large quantities of normal speech data. The amount of training data used during the initial training phase can be increased by including speech data from different speaker groups such as normally speaking elderly people and children. In the following step, the layers of this DNN are retrained using only speech data that resembles the target speech, e.g. using Dutch

[1] http://hstrik.ruhosting.nl/chasing/

dysarthric speech and/or Dutch elderly speech. The aim of the second step is to tune the DNN on dysarthric speech as this is the type of speech to be recognized.

We have investigated multiple parameters that may influence the accuracy of the final model, such as the number of retrained layers and the learning rate. Moreover, various types of speech data have been used to explore their impact on the modeling accuracy of the final DNN model. Normal speech has been used due to its abundance compared to other deviant speech types. Since the majority of dysarthric speakers are older than 50, elderly speech data is also relevant in this scenario. Finally, normal speech data from a related variety of Dutch, namely Flemish, is included to obtain the background model. Using speech data from different language varieties led to a mild improvement in recognition accuracy in a previous study (chapter 3). Since both varieties share the phonetic alphabet, we learn several hidden layers and a softmax layer on both varieties with the aim of learning more reliable hidden layers. The following section continues this chapter by describing the speech corpora that have been used during the experiments.

4.3. Speech corpora selection

Given the limited availability of dysarthric speech data, we investigate to what extent already existing databases of Dutch normal speech can be employed to train DNNs and optimize their performance on dysarthric speech. There have been multiple Dutch-Flemish speech data collection efforts (Cucchiarini et al., 2008; Oostdijk, 2000) which facilitate the integration of both Dutch and Flemish data in the present research. For training purposes, we used the CGN corpus (Oostdijk, 2000), which contains representative collections of contemporary standard Dutch as spoken by adults in the Netherlands and Flanders. Considering that the high median age in our database of dysarthric speech is 66.5 years, we have also included elderly speech data from the JASMIN corpus (Cucchiarini et al., 2008) to the Dutch normal speech in the training phase.

The EST Dutch dysarthric speech database (Yılmaz et al., 2016) contains dysarthric speech from ten patients with Parkinson's Disease (PD), four patients who have had a Cerebral Vascular Accident (CVA), one patient who suffered Traumatic Brain Injury (TBI) and one patient having dysarthria due to a birth defect. Based on the meta-information, the age of the speakers is in the range of 34 to 75 years with a median of 66.5 years. The level of dysarthria varies from mild to moderate. The dysarthric speech collection for this database was achieved in several experimental contexts. The speech tasks presented to the patients in these contexts consist of numerous word and sentence lists with varying linguistic complexity. The database includes 12 Semantically Unpredictable Sentences (SUSs) with 6- and 13-word declarative sentences, 12 6-word interrogative sentences, 13 Plomp and Mimpen sentences, 5 short texts, 30 sentences with /t/, /p/ and /k/ in initial position and unstressed syllable, 15 sentences with /a/, /e/ and /o/ in unstressed syllables, production of 3 individual vowels /a/, /e/ and /o/, 15 bisyllabic words with /t/, /p/ and /k/ in initial position and unstressed syllable and 25 words with alternating vowel-consonant composition (CVC, CVCVCC, etc.).

As mentioned above, for testing purposes we use the CHASING01 dysarthric speech database that was recently collected in the first stage of the CHASING project. This database contains speech of 5 patients who participated in speech training experiments and were tested at 6 different times during the treatment. For each set of audio files, the following material was collected: 12 SUSs, 30 /p/, /t/, /k/ sentences in which the first syllable of the last word is unstressed and starts with /p/, /t/ or /k/, 15 vowel sentences with the vowels /a/,/e/ and /o/ in stressed syllables, appeltaarttekst (*apple cake recipe*) in 5 parts. Utterances that deviated from the reference text due to pronunciation errors (e.g. restarts, repeats, hesitations, etc.) were removed. After this subselection, the utterances from 3 male patients remained and were included in the test set. These speakers are 67, 62 and 59 years old, two of them having PD and the third having had a CVA.

4.4. Experimental Setup

4.4.1. Database details

The CGN components with read speech, spontaneous conversations, interviews and discussions were used for acoustic model training. The duration of the normal Flemish (FL) and northern Dutch (NL) speech data used for training is 186.5 and 255 hours, respectively. The combined training data (Nor. FL+NL) contains 441.5 hours in total. The total duration of the elderly speech recordings in the JASMIN database (Eld. NL) is 10 hours and 10 minutes.

The EST Dutch dysarthric speech database (Dys. NL) contains 6 hours and 16 minutes of dysarthric speech material from 16 speakers (Yılmaz et al., 2016). The speech segments with pronunciation errors (e.g. restarts, repeats, hesitations, etc.) were excluded from the training set to maintain integrity of the results on ASR performance evaluation. Additionally, the segments including a single word and pseudoword were also excluded, since the sentence reading tasks are more relevant in our project context. The total duration of the dysarthric speech data eventually selected for training is 4 hours and 47 minutes.

The CHASING01 speech database, which was used for testing, contains 721 utterances (6231 words) with corresponding manual transcriptions that match the reference text. The total duration of this speech data is 55 minutes.

4.4.2. Implementation Details

The recognition experiments were performed using the Kaldi ASR toolkit (Povey et al., 2011). A standard feature extraction scheme was used by applying Hamming windowing with a frame length of 25 ms and frame shift of 10 ms. A conventional context dependent GMM-HMM system with 40k Gaussians and 5925 triphone states was trained on the 39-dimensional MFCC features including the deltas and delta-deltas. We also trained a GMM-HMM system on the LDA-MLLT features, followed by training models with speaker adaptive training using FMLLR features. This system was used to obtain the state alignments required for DNN training.

The DNNs with 6 hidden layers and 2048 sigmoid hidden units at each hidden layer were trained on the 40-dimensional log-mel filterbank features with the deltas and delta-deltas. The DNN training was done by mini-batch Stochastic Gradient Descent with an initial learning rate of 0.008 and a minibatch size of 256. The default initial learning rate of 0.008 was used in the first training stage. The time context size was 11 frames achieved by concatenating ±5 frames. A trigram language model trained on the target transcriptions of the sentence tasks was used during recognition of the sentence tasks.

4.5. Results and Discussion

We performed several ASR experiments using the speech data described in subsection 4.4.1. Firstly, we explored the impact of the number of retrained layers and initial learning rate on the recognition accuracy in a two-stage training setting. The Word Error Rates (WER) obtained on the CHASING01 test set after having trained models on the normal speech database (Nor. NL) and retrained on EST's dysarthric speech database (Dys. NL) are presented in Table 4.1. The lowest WER is marked in bold. Two recognizers trained on the Nor. NL database and Dys. NL separately provide a baseline WER of 21.3% and 17.3%, respectively. WERs yielded by the recognizers with two-stage training are considerably lower than those of the baseline systems and vary between 11.0%-13.6%. The recognition accuracy is obtained by only retraining the softmax and the last hidden layer with an initial learning rate that is the same as the initial learning rate used in the first stage. The results for different numbers of retrained layers do not follow a pattern, hence, it is difficult to formulate a superior retraining strategy. However, we can conclude that retraining only the softmax layer with a relatively low learning rate results in a reduced recognition accuracy of 13.6% with respect to retraining more layers with the same learning rate or retraining with a higher learning rate. It is important to mention that the amount of in-domain data used in the retraining stage will have an impact on the choice of the number of retrained layers.

Training	Retraining	# of Retr. Lay.	Retr. Init. LR	WER (%)
Nor. NL	-	-	-	21.3
Dys. NL	-	-	-	17.3
Nor. NL	Dys. NL	all	0.008	12.1
Nor. NL	Dys. NL	5	0.008	12.6
Nor. NL	Dys. NL	4	0.008	12.8
Nor. NL	Dys. NL	3	0.008	12.0
Nor. NL	Dys. NL	2	0.008	12.4
Nor. NL	Dys. NL	1	0.008	**11.0**
Nor. NL	Dys. NL	softmax	0.008	11.9
Nor. NL	Dys. NL	all	0.0008	11.6
Nor. NL	Dys. NL	5	0.0008	11.8
Nor. NL	Dys. NL	4	0.0008	11.8
Nor. NL	Dys. NL	3	0.0008	11.8
Nor. NL	Dys. NL	2	0.0008	12.0
Nor. NL	Dys. NL	1	0.0008	12.2
Nor. NL	Dys. NL	softmax	0.0008	13.6

Table 4.1.: *Word error rates in % obtained on the test set for different number of retrained layers (# of Retr. Lay.) and retraining initial learning rate (Retr. Init. LR)*

In Table 4.2, we present a similar set of results by varying the content of data used in the initial training phase. Speech from a related Dutch language variety, Flemish, was used. Background models were trained on both the Northern and Flemish varieties of Dutch (Nor. NL+FL) instead of the Northern variety only, as was done in the previous paragraph. The goal of these experiments was to investigate the impact of adding speech from a related language variety to the training procedure on the modeling accuracy of the final models tuned on the Dys. NL data. The best recognition accuracy was provided by the system obtained with retraining of the softmax layer with a relatively high initial learning rate. That system has a WER of 11.3% which is comparable with, but not better than the previous best performing system presented in Table 4.1. Using in-domain speech data for training, the performance gain reported in previous experiments (chapter 3) cannot be obtained in this scenario.

Finally, the impact of retraining the background acoustic model on the EST Dysarthric data (Dys. NL) and speech data from Dutch elderly (Eld. NL) is shown in Table 4.3.

Training	Retraining	# of Retr. Lay.	Retr. Init. LR	WER (%)
Nor. NL+FL	Dys. NL	all	0.008	12.8
Nor. NL+FL	Dys. NL	5	0.008	12.9
Nor. NL+FL	Dys. NL	4	0.008	12.6
Nor. NL+FL	Dys. NL	3	0.008	12.5
Nor. NL+FL	Dys. NL	2	0.008	12.5
Nor. NL+FL	Dys. NL	1	0.008	12.2
Nor. NL+FL	Dys. NL	softmax	0.008	**11.3**
Nor. NL+FL	Dys. NL	all	0.0008	12.0
Nor. NL+FL	Dys. NL	5	0.0008	11.9
Nor. NL+FL	Dys. NL	4	0.0008	12.0
Nor. NL+FL	Dys. NL	3	0.0008	12.3
Nor. NL+FL	Dys. NL	2	0.0008	12.3
Nor. NL+FL	Dys. NL	1	0.0008	12.0
Nor. NL+FL	Dys. NL	softmax	0.0008	12.2

Table 4.2.: *Word error rates in % obtained on the test set for different number of retrained layers (# of Retr. Lay.) and retraining initial learning rate (Retr. Init. LR)*

Training	Retraining	# of Retr. Lay.	Retr. Init. LR	WER (%)
Nor. NL	Dys.+Eld. NL	all	0.008	15.1
Nor. NL	Dys.+Eld. NL	5	0.008	15.1
Nor. NL	Dys.+Eld. NL	4	0.008	15.5
Nor. NL	Dys.+Eld. NL	3	0.008	15.1
Nor. NL	Dys.+Eld. NL	2	0.008	14.7
Nor. NL	Dys.+Eld. NL	1	0.008	14.7
Nor. NL	Dys.+Eld. NL	softmax	0.008	**13.9**
Nor. NL	Dys.+Eld. NL	all	0.0008	14.8
Nor. NL	Dys.+Eld. NL	5	0.0008	15.1
Nor. NL	Dys.+Eld. NL	4	0.0008	15.3
Nor. NL	Dys.+Eld. NL	3	0.0008	14.7
Nor. NL	Dys.+Eld. NL	2	0.0008	15.1
Nor. NL	Dys.+Eld. NL	1	0.0008	15.0
Nor. NL	Dys.+Eld. NL	softmax	0.0008	15.2

Table 4.3.: *Word error rates in % obtained on the test set for different number of retrained layers (# of Retr. Lay.) and retraining initial learning rate (Retr. Init. LR)*

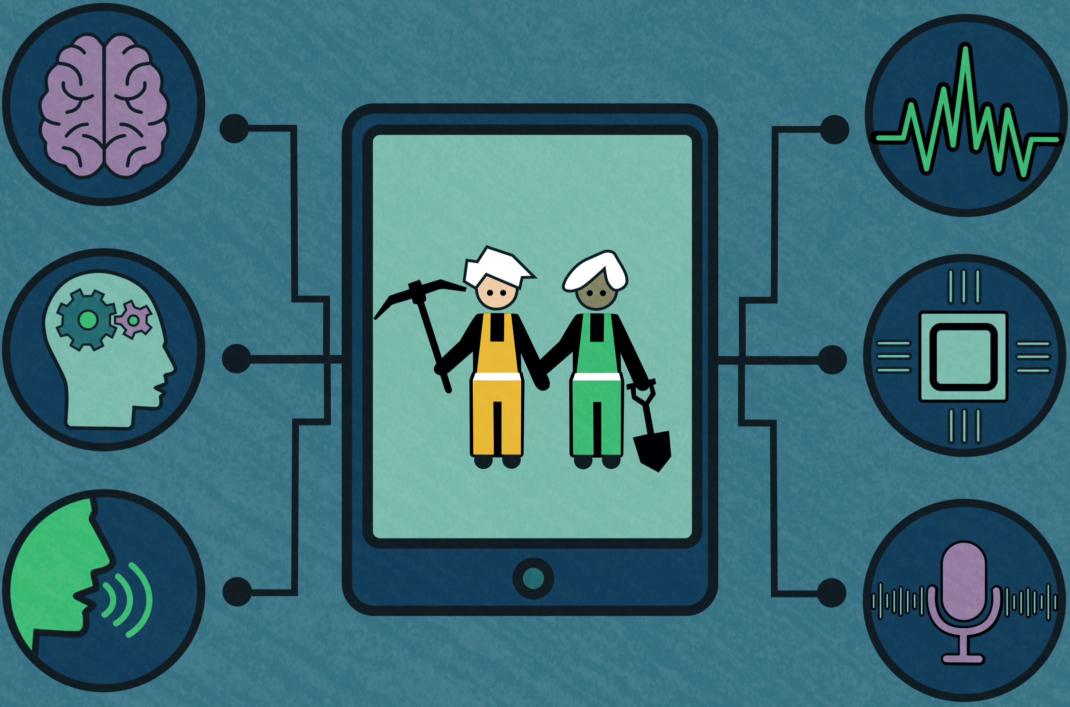
The presented WER results vary between 13.9%-15.5% for different training parameters. From these results, it can clearly be seen that the performance of the final acoustic models deteriorate when elderly speech is used for retraining in all scenarios. The impact of the mismatch between the elderly and dysarthric elderly speech, e.g. reduced speaking rate and articulation skills, appears to be more salient than the increase in the amount of retraining data on the recognition accuracy.

To summarize, we can conclude that using in-domain data in the described two-stage training scheme improves the recognition performance significantly whereas merging different speech types in a single-stage training scheme provides only minor improvements (Ganzeboom et al., 2016). Adding relevant types of data, i.e., the Flemish Dutch variety during background model training and using elderly speech data for retraining, does not improve the recognition accuracy of the final models.

4.6. Conclusions

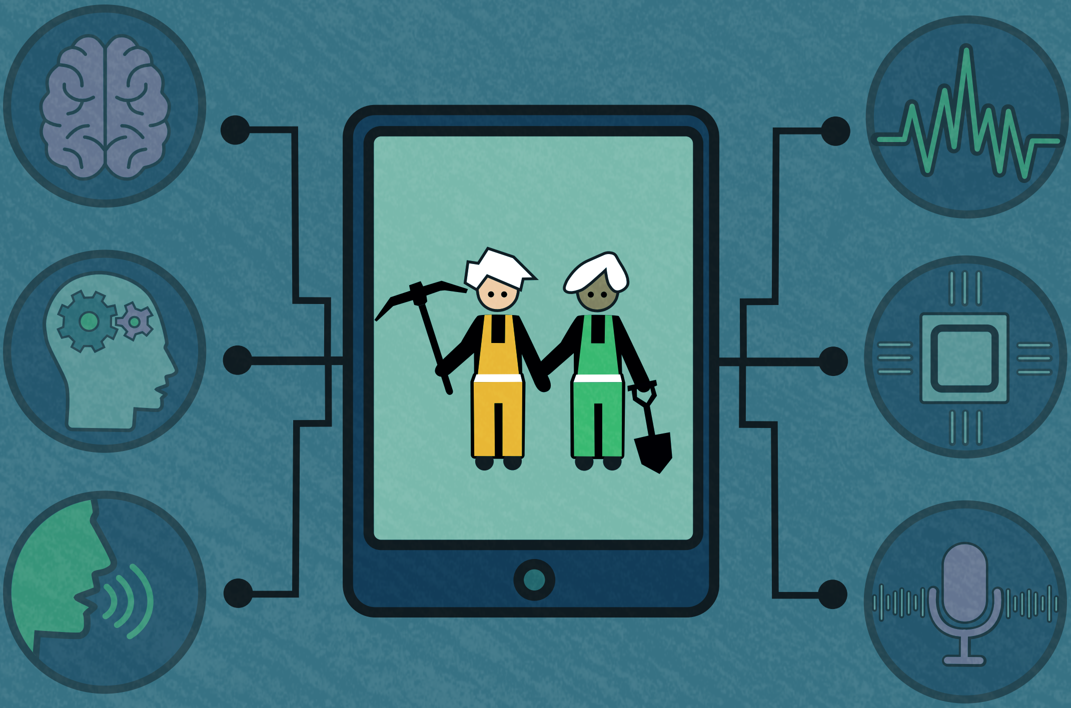
In this chapter, a multi-stage DNN training scheme was applied to obtain robust acoustic models in the framework of a serious game to be used as an individualized speech therapy tool. These models are applied to Dutch dysarthric speech, which is more challenging to recognize than normal speech due to its increased variation. The data recently collected through the game could be used for testing, while the dysarthric data already available from the EST database were used for training. The applied multi-stage training approach aims to learn a background model trained on more general data in the initial stage. That model was then retrained on in-domain data in the second stage to get a domain-specific model.

We performed several ASR experiments by varying two training parameters, namely the number of retrained layers and the initial learning rate used in the second stage. The results have shown that this kind of training provides large improvements in recognition accuracy compared to baseline systems trained either on normal speech or on dysarthric speech. Moreover, we investigated the inclusion of various speech types such as normal speech from a related language variety for background model training and elderly speech for retraining in further recognition experiments. The recognition results suggest that adding normal speech data from a language variety does not bring improvement compared to a recognizer trained on only normal speech from the target language. Adding elderly data reduced the recognition performance compared to retraining only on dysarthric speech, most likely due to the increased mismatch between the training and target speech.



Part III.

Designing a serious game for speech training and exploring its efficacy



Chapter 5

Lessons learned from the design, development, and use of a serious game for speech training in older adults

This chapter has been submitted to the *International Journal of Serious Games*.

5.1. Introduction

As countries around the world are facing an ageing population (World Health Organization, 2022), one of the potential consequences is a growing incidence in neurological disorders like Parkinson's Disease (PD) and Cerebral Vascular Accident (CVA or stroke). Both are known to cause dysarthria, a speech disorder that affects speech intelligibility and causes communication problems (Duffy, 2019). A well-known therapy developed in the USA is the Lee Silverman Voice Treatment (LSVT; Ramig et al., 2001). The LSVT aims to improve a patient's articulatory precision by focusing on increasing the intensity of speech. However, a potential side effect of an increase in intensity is that patients often raise their pitch, putting extra strain on the vocal cords and tiring them. The Pitch Limiting Voice Treatment (PLVT, Kalf et al., 2011), is an intensive speech training program that is currently standard practice in the Netherlands and focuses on increasing voice intensity while maintaining a low pitch. Its goal is to train patients to 'speak loud and low'.

In line with the PLVT protocol (Kalf et al., 2011), speech therapists provide the therapy in four face-to-face sessions per week for four consecutive weeks. Given the rise in the number of patients and the limited resources in healthcare, it becomes increasingly more difficult to provide the necessary level of care.

A potential solution may be provided by eHealth, 'the cost-effective and secure use of information and communication technologies (ICT) in support of health and health-related fields', as defined by the World Health Organization (WHO, World Health Organization, 2005). Using ICT, eHealth solutions enable patients to train intensively and independently in their home environment while still being monitored by a speech therapist. Previous research focused on courseware-like approaches in which patients followed prescribed courses of exercises online (Beijer et al., 2014; Frieg et al., 2017). One of these approaches employed a drill-and-practice training method (Beijer et al., 2014).

In addition to finding beneficial effects on speech intelligibility, the authors reported that patients described difficulties in transferring the improved speech skills to daily-life situations. The patients also described a lack of variation in speech training exercises that reduced their motivation to train. In this regard, our project 'Challenging Speech Training in Neurological Patients by Interactive Gaming' (CHASING, Strik, 2014), explored the use of serious games for providing remote, intensive speech training. It was conducted in collaboration with the Creative Care Lab at Waag Society (Creative Care Lab, Waag Society, 2014). The project adopted a widely accepted definition of serious games: 'games that do not have entertainment, enjoyment, or fun as their primary purpose' (Laamarti et al., 2014). Compared with more drill-and-practice approaches, serious games are more suitable for simulating speech training scenarios that are closer to daily-life situations. For example, a scenario in which the patient needs to ask for directions. Additionally, serious games have the potential to be more engaging, triggering patients' intrinsic motivation to train as a result.

In the CHASING project, two versions of the serious game Treasure Hunters were designed, developed, and part of clinical trials to explore their efficacy. Both versions were two-player cooperative tablet games in which players navigated a virtual map and needed to help each other to find the treasure by exchanging information via an online voice connection. The first version provided automatic feedback on loudness and pitch in accordance with PLVT prescriptions. It was part of a clinical trial to explore its effects on speech intelligibility and user satisfaction (chapter 6). In the second version, game elements were added that required the pronunciation of longer utterances enabling an intended increase in the intensity of the speech training. Additionally, pronunciation exercises were added to address the difficulties dysarthric patients have with articulation. Patients received feedback on their pronunciation using automatic speech recognition technology developed earlier in the project. The second version was also part of a clinical trial to assess its effects on the same dimensions (chapter 7). After analyzing the results, a significant difference in speech intelligibility ratings was found after the participants trained with the serious game, indicating positive effects of the serious game-based speech training.

This paper is an extension of our previous paper (Ganzeboom et al., 2016), which only describes parts of the design process and some findings on our initial designs of the first version of the game. This work adds a full description of our design and development method including the results that were obtained with both versions of the game. Retrospectively, this paper also adds the lessons that were learned during the design and development process. Guidelines are formulated at the end of this paper for the benefit of future research into serious games for speech training.

5.2. Related work

A substantial body of literature exists on gaming for children with varying speech disorders (Nasiri et al., 2017). However, for elderly with speech disorders is far less available. Early literature researched the use of technical diagrams and vocal tract schemes to show feedback on speech (Hutchins, 1992). These were often difficult to interpret without additional explanation. Later research also included user interaction (S. Ferguson et al., 2012) and focused on visual representation of speech (Pietrowicz & Karahalios, 2014).

5.2.1. Gaming for speech training

Krausse et al. (Krause et al., 2013) were the first to report on a study that incorporated game elements to provide challenging speech therapy in older adults with dysarthria. In addition to improving their reduced voice intensity, the game also aimed to make speech training more engaging by using gamification. A widely accepted definition of gamification is: "the process of incorporating game design elements into non-game contexts, such as business, education, or healthcare, to engage and motivate people to achieve their goals" (Damaševičius et al., 2023). Krause et al. (2013) added game elements to traditional drill-and-practice exercises in which players were motivated to virtually break different items by producing sufficiently loud and long speech sounds. The items broke at varying intensities and durations, which were calibrated per player beforehand. Every item also had an associated score which was awarded to the player after breaking the item. The goal of the game was to break all items and obtain all the points available. All study participants achieved that goal. They also appeared to be engaged and enthusiastic about continuing playing.

Research was also reported on a game system for speech rehabilitation in which players controlled 3D game scenarios with movement of their tongue (Shtern et al., 2012). The tongue's movement range is often reduced due to dysarthria, affecting the player's pronunciation ability. Tongue movement exercises are beneficial to improve this ability. Prototype game scenarios to support these exercises were moving a wooden plank up and down a deformable ball by moving the tongue down and up, respectively. In a subsequent version, the game was changed to a cartoon-like bee avatar that was controlled by the player's tongue movements. The goal of the game was to let the bee land on flowers and collect pollen. Points were awarded to the player with every successful collection. Clinicians could also personalize the game to the player's exercising needs. In theory, this game may have been expanded to include different courses with varying difficulty for the training of players. However, in a paper reporting on a pilot study using this gaming system (Yunusova et al., 2017), very different game scenarios were used. A fishing net that expanded as the player's tongue movement range increased while pronouncing a sentence and a dragon's range of spewing fire also increased similarly. In this paper's perspective, these are scenarios containing gamified speech exercises similar to those described by Krausse et al. (Krause et al., 2013).

More recently, research has been reported that included a requirements analysis and design of a serious game to aid speech rehabilitation (Baranyi et al., 2024). While it was still in the early design phases, a prototypical implementation included the main game elements and automatic feedback through speech recognition (Weber, 2025). Being a 2D platform game, the main game elements consisted of moving a character from one platform to the next. This was done by making it jump when the corresponding utterance was pronounced correctly. While this is a flexible game design that can support types of utterances for multiple speech impairments and can be played by children and adult players, it may not be motivating enough to play for multiple weeks in an intensive speech intervention. Additionally, being able to only control the character using utterances can become frustrating when continuously pronouncing the utterance incorrectly. It may also provide a more 'drill-and-practice' experience that potentially demotivates players. Both, the frustration and demotivation is what this paper's research is aiming to avoid.

A different form of gamified speech exercises was included in researching the use of a mobile app employing crowdsourcing (McNaney et al., 2016). Generally, users recorded a sentence using the app and sent it to an online crowdsourcing platform. Other crowd workers connected to the platform rated the loudness, pitch variation, and speaking rate of the recorded speech on a scale of 0 - 100. The median of the ratings was returned and visualized in the app. Users were also given goals by a clinician for improving their speech. During the qualitative interviews after the study, the users described that they were enthusiastically challenged to achieve these goals. Additionally, users were motivated to do another exercise whenever they received a rating that was lower than the previous one. Although this app does not employ game scenarios as described in the older literature, elements of gaming can be identified. Obtaining a median rating is perhaps similar to scoring points in a game. The player is motivated to increase both.

In the previous literature described, games were designed on the basis of traditional speech training and interaction design. However, to provide training that players are motivated to do intensively for longer periods, it is necessary to design game scenarios that evoke players intrinsic motivation. To this end, motivational theories from the field of psychology can prove valuable. One such theory that has been applied to the development process of gamified speech training is the Self-determination theory (Mühlhaus et al., 2017). The researchers designed gamification elements to enable integration of all the player's needs described by the theory.

5.2.2. Game design for older adults

Literature on game design for older adults in general may provide valuable insights for designing a serious game for speech training. An early study on heuristics for the design of mobile serious games provided insightful recommendations (Machado et al., 2018). A more recent study reaffirms those of previous studies and described additional recommendations (S. Lee et al., 2021).

In addition to design heuristics and recommendations, older adults attitude towards serious games was also investigated and showed that it can change from negative to positive after having tried a game (Pyae et al., 2017). Gaming preferences were also investigated (Tabak et al., 2020). As shown by that study, older adults prefer gaming on non-console-like platforms, playing with others or alone in non-competitive scenarios, like puzzle-based games and dislike fast and stimulating gameplay. These are all preferences that are relevant to the research reported in this paper.

5.2.3. Game design approach for serious games

In order to design a serious game for speech training effectively, it is good practice to employ a structured approach. This is taken from the fields of software development and user interface design in which a user-centered design (UCD) approach is advocated (Wallach & Scholz, 2012). In the past, research has been published on an approach to design games for therapy (Amengual Alcover et al., 2018; Beristain-Colorado et al., 2021). It clearly integrates the principles of UCD by including the target users, clinicians and researchers. Also, work has been published applying these principles to serious games for health (Baranyi et al., 2020; Ratnanather et al., 2021; Laine et al., 2020) and education (Wanick & Bitelo, 2020; Pacheco-Velazquez et al., 2023). All reported positive results by including target users and stakeholders at multiple stages of the design. Recent literature includes research into little or no-code game authoring tools which require little or no-coding skills to design, develop, and deploy serious games (Laurent et al., 2022; Colado et al., 2023; Torres et al., 2025). This can make the process of developing serious games more efficient and reduce costs involved. For now, this research is mostly in the educative domain, but other domains (e.g. rehabilitation of limbs, speech, cognitive functions, etc.) may benefit in a similar manner.

5.3. Methodology

To design and develop the game, a CoDesign approach (E. B.-N. Sanders & Stappers, 2014) was chosen. CoDesign is a joint, creative process that supports multiple stakeholders in reaching their common goals. The expertise of speech therapists, patients, caregivers, designers and researchers in the area of speech training were utilized by this approach in every part of the design process. To guide the CoDesign approach, the following principles were implemented (Van Dijk et al., 2011):

- Receive feedback and inspiration from a small number of users.
- Involve users in every part of the design process right from the start.
- Visualize application ideas at an early stage and test them using prototypes.
- Work in multidisciplinary teams and share knowledge.
- Imagine being a future user and let that user experience the new possibilities.

Our CoDesign approach consisted of six phases:

A. Getting to know the target users
B. Determining game principles
C. Designing game concepts
D. Prototyping game concepts
E. Game development
F. Level design

All phases were set up using an iterative process that included interviewing target users, initial game design or development, testing the result with target users and using their feedback to start a subsequent iteration. In most phases, two or three of these iterations were completed before obtaining a satisfactory result. The following subsections provide a description of each phase.

5.3.1. Getting to know the target users

The goal of the first phase was to get to know the target users of the game. To inform important design decisions it is important to know who the game is designed for and for whom it is not, what their characteristics are as well as their capabilities. To determine who the game is designed for and for whom it is not, all stakeholders were interviewed to obtain their requirements and constraints. From those results an initial profile of the target users was written. Target users that met that profile were contacted and asked to participate. Qualitative interviews were held to get to know the characteristics and capabilities of the target users. Using the results of these interviews, multiple portraits were composed to get an in-depth description of individual target users. Portraits are composed with an existing user in mind, albeit anonymously. They can contain detailed descriptions of user characteristics (age, marital status, hobbies, etc.), personal anecdotes with respect to their likes and dislikes, their daily routines, and their surroundings. Their current capabilities and the capabilities they lost due to their neurological disorder were also included in the descriptions. Composing personas was also considered as an alternative for portraits. However, these are usually based on the average target user and lack the level of detail for a designer to empathize with such a person. Portraits also have the advantage that the designer really gets to know the user by returning multiple times to test and improve the design.

Speech therapists' expert knowledge on patient characteristics and capabilities was also included in composing portraits. Additionally, their application of current dysarthria treatments in clinical practice including ways to motivate patients comprises a valuable body of knowledge.

To access this knowledge, a creative workshop was organized with ten speech therapists to obtain insight on different topics. Two sessions were held to brainstorm about how current treatments could be made more playful and what the ideal speech training would be when no limitations existed. The attendees were divided into groups and were all asked to come up with a concept for the serious game.

The output from the above sessions was combined with that of qualitative interviews held with the academic researchers who described the possibilities and constraints in researching dysarthria treatment efficacy. Goals for the serious game were formulated from the output and prioritized by the stakeholders in a separate design session.

5.3.2. Determining game principles

Determining game principles was an essential step in our game design process. Game principles are defined in this paper as principles that focus on the player's game experience from the start to the end. They provide a foundation for how the game will be played and include a definition of gameplay elements and overall storyline.

From the interviews with the academic researchers it became clear that the most important guidelines for determining the game principles were the following:

A1 The game principles should be chosen in such a way that they challenge players to speak, also when having difficulties with speaking.

A2 They should also strike a balance between challenging players' speaking capabilities, level of difficulty and fatigue.

A3 While not every player will like the same type of game, the principles should reflect the ones that will motivate the target group to train their speech on a regular basis.

A4 The choice of game principles should take their suitability to provide feedback on players' speech into consideration.

Using these guidelines, multiple game principles for our game were identified from the previous interviews. These were then tested with target users and their partners using existing games and apps. Afterwards, the target users were interviewed about positive and negative aspects of those games and apps and asked how they imagined to practice their speech daily.

5.3.3. Designing game concepts

From the results of the game principles tests described in the previous subsection, principles were identified that were either in line with the target users' preferences or the goal of the game. Next, three very different game concepts were designed using these principles to explore users' motivation to perform speech exercises. These were tested in multiple sessions and included users that had not participated before. Subsequently, a fourth game concept was tested that combined the positive elements of the previous three. Qualitative observations and interviews were used to gather users' experiences, feedback and mode of play. From the outcomes, game concepts that contributed to the goal of the game were identified.

5.3.4. Prototyping game concepts

A first interactive digital prototype was developed for tablets using the game concepts identified as described in the previous subsection. Prototyping is a useful way to investigate the feasibility of game concepts, using only a small amount of time and resources. Different forms of low and high-fidelity prototypes were used to test the game concepts: pen and paper, existing apps and interactive visualizations. Additional prototypes were created iteratively to test those concepts that may provide improvement. They were integrated in the interactive digital prototype when improvement was found. A process that is similar to the incremental development in evolutionary prototyping (McConnell, 1996). Thorough testing of the initial digital prototypes was carried out by the designers. Afterwards, the resulting prototype was tested with a group of speech therapists and pathologists to include their feedback before testing with the target users themselves.

5.3.5. Game development

An iterative design process was chosen to develop the game. Input to this process was the knowledge obtained in the previous subsection. It focused on developing the three parts of our game design: user interaction, content, and graphical visualization. The following guidelines were used in our aim to design accessible and intuitive user interaction:

I1 Make use of target users' prior knowledge of interfaces, which was assumed to be limited.

I2 Make use of a real-world metaphor, like visualizing a button as pressed just as a physical button would be.

I3 Use, visualize, and position interface elements consistently.

Content was defined as the overarching story of the game. For its design, the following criteria were used:

C1 It should stimulate speaking loud and low and increase articulatory effort.

C2 Integrate ways to provide feedback on users' speech.

C3 Be sufficiently challenging to continue practising at home after clinical speech therapy sessions ended.

C4 Keep motivating users to play for at least 15 minutes, four times a week, for multiple weeks.

C5 It should fit in with the actual world and interests of the target users.

Graphical visualization was defined as the graphical look and feel of the game's interface and world. The following guidelines were used for designing the graphics:

G1 Provide guidance on where to focus during play, due to potential cognitive impairment.

G2 Show any text as large as possible, because of potential visual impairment.

G3 Use high contrastive colors.

G4 Do not only contrast in colors, but also in shape where possible, because of potential color blindness.

5.3.6. Level design

Levels are commonly used to progress through a game's storyline in smaller parts. Users may complete these parts by solving puzzles and finding particular information. The design of these levels is important to continuously challenge users and motivate them to play repeatedly. To achieve that, applying a method of persuasion can be beneficial. The Hooked Model (Eyal, 2014) is such a method. It is often implemented in a commercial context to get users 'hooked' on a particular app. However, the principles of this method may also be beneficial to achieving positive behavioral change. For that reason, it was added to the following list of guidelines to facilitate level design:

L1 Introduce the user to the game basics at the start.

L2 Increase the level of difficulty gradually.

L3 Create a level design that is easily expandable.

L4 Apply the four principles of the Hooked Model (Eyal, 2014) to intrinsically motivate users to play.

Level designs were internally tested at the Creative Care Lab before testing with target users and observing their experiences to iteratively improve the designs.

5.4. Results

5.4.1. Getting to know target users

From the interviews with all stakeholders an initial profile of the target users was written. The target users were adults aged 25 years or older. However, it was expected that most users would be 55 years or older, considering the prevalence of neurological disorders in that group. They are affected by an acquired neurological disorder, like CVA (or stroke), PD, or TBI. These disorders cause decreased speech intelligibility due to dysarthria. A reduced physical condition is also expected. For example, increased fatigue, decreased hand-eye coordination, vision problems, and/or reduced mobility.

Cognitive limitations are also likely. For example, decreased attention span, decreased short-term memory, decreased focus, and/or visual neglect, in which the user exhibits a lack of response to stimuli in one half of their visual field.

Users that are affected by language production disorders, like aphasia, were excluded from this profile.

Collaborating speech therapists were asked to contact target users that met the above profile. Portraits were then composed of several participating users. An anonymized example is included in Appendix 5A.

In the creative workshop with speech therapists, multiple concepts for a game were imagined:

1) A scenario playing game. For example, a player asks their partner, who is busy in the kitchen, to bring something in a loud enough voice. The player is rewarded when the partner comes with the right item.

2) Speaking loud and low results in reward. For example, making an object visible when speaking loud enough and moving it down the screen when simultaneously speaking low.
3) Immediate feedback on voice loudness and pitch: a screen reveals more of a personal picture when correct levels are reached.
4) Digital board-like game: the roll of the dice will have a higher outcome in the next turn when speaking at the right levels. After every game, the player is rewarded by receiving a part of something greater.

The above output combined with the interviews of the stakeholders resulted in a list of game goals which can be found in Appendix 5B.

5.4.2. Determining game principles

Using the guidelines described in the previous section, the following game principles were identified from the interviews and workshops:

1) Exploring: to control an object with your voice or make certain decisions.
2) Knowledge testing: to participate in quizzes.
3) Experiencing: to immerse in a conversation or story by means of a role-playing game.
4) Performing: to recite a story or to create one with a partner or grandchildren.
5) Creating: make music with your voice or draw/build a world.

These game principles were tested with target users using existing games and apps. What follows is an overview of users' positive and negative feedback.

1) **Exploring**
   A 'boat puzzle' game was used in which players had to move waves with their voice to lead their red boat to the exit of the puzzle. The game itself, a 3-dimensional puzzle, was disliked by all target users. It was found too easy, childishly visualized, and it provided more frustration than joy.
2) **Knowledge testing**
   This principle was tested using two quizzes. First, a multiple-choice quiz in which fictitious money was earned. Notable positive comments were that it trains the brain, the rewards are fun, and questions are adaptable to personal interests. Negative comments that were shared by most target users were that the music was irritating and loud, and answering incorrectly was demotivating as you lost all your money. Secondly, a word-guessing quiz was used in which a picture with a few letters was provided. Most target users found it rather difficult to switch focus between the picture, word and letters. Also, it only triggered speech when the player thought out loud.

3) **Experiencing**
A voice changer and speech-to-text apps were used to test this game principle. The former changed the player's voice in real time using filters. Positive comments were that it focused on the perception of one's voice and it is fun to do. However, it is difficult to keep speaking when hearing one's voice distorted and many felt that it did not contribute much to speech training. The speech-to-text app was used to compose messages to friends and family. Most target users noted that it provides feedback on players' speech, and it is fun to contact people you know. A negative comment was that the provided feedback contained no detail.

4) **Performing**
To test this principle, a newsreader was simulated using an autocue app. The player is recorded while presenting the news messages and can listen to it afterwards. Many commented that it was fun, one can choose interesting topics and multiplayer is possible. Additionally, it was found instructive to hear one's own voice, albeit not always pleasant to hear that improvement is needed.

5) **Creating**
An app was used that provided visual feedback on the usage of one's voice. Most players found it too distracting, making it difficult to simultaneously focus on speaking and the visual feedback. Another game employed dice with narrative icons. Every roll of the dice added an icon to an evolving story. Players noted that it sparked their imagination and creativity, it was fun to visually interact with language, it enabled multiplayer, and motivated them to speak.

5.4.3. Designing game concepts

The principles 'Experiencing' and 'Performing' showed the most potential for our serious game. Three very different game concepts that included them were designed and tested with players. The following is an overview of their comments.

1) *Newsreader*: The newsreader concept described in the previous subsection was expanded to let players improve their news reading and get promoted to regional or even national newsreaders. They could organize a schedule to read news bulletins, weather forecasts, and/or traffic news. Players described that they liked doing that and it was well suited for single play. However, we observed that some had difficulties and felt they were stumbling along. Players also commented that reading to an audience could increase motivation.

2) *Voice coach*: This concept is based on the idea of an app providing feedback on speech continuously, at convenient and appropriate times. It could also be used for doing voice warming up exercises. Using a Wizard of Oz setup, the player was provided feedback through headphones while presenting a text.

The tests showed that the feedback was either ignored or found disturbing. In some cases players started reciting the provided feedback. This improved when the feedback was given during breaks in between speaking.

3) *Audio adventure*: In the audio adventure concept, two players needed to find each other. Both players were put on a map with icons of which only four were visible at any time. The goal of the game was to find each other by moving over the map and describing to each other what they see. They spoke to each other via an online audio connection. During the tests, many players had difficulties orienting on the map. For that reason, the icons were made bigger and the maps less random by using themed icons (e.g. recognizable landmarks). An example of the paper prototype is shown in Figure 5.1. Players also described that they could easily play this game daily and liked that they could play it with others close to them.

Positive elements of these three were then combined into a fourth concept that was titled ‘Imaginative stories’. However, it was not well received in user tests. Players described that reading out loud sentences to move left or right around a stylized, two-dimensional globe made no sense and felt like mandatory speech exercises. Additionally, this concept did not provide the freedom of movement experienced in the audio adventure concept. In our observations, both reduced the overall game experience. Figure 5.2 shows the stylized globe.

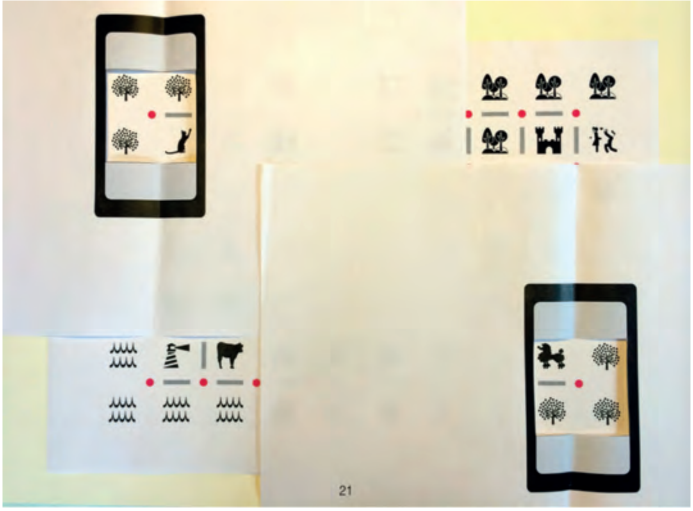


Figure 5.1.: *Paper prototype of the audio adventure, displayed here with the two player tablets for illustrative purposes.*

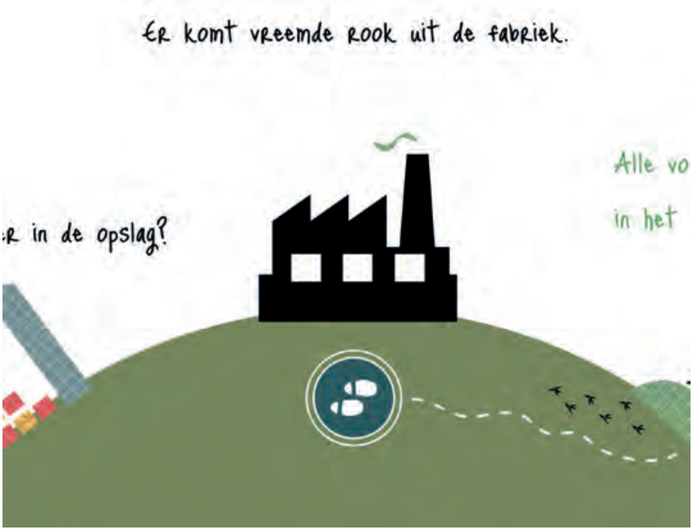


Figure 5.2.: *The stylized globe in the 'Imaginative stories' concept. Players can only go left or right along the globe by pronouncing the corresponding Dutch sentences and attempt to find the other player. (Dutch text on top: 'Strange smoke is coming from the factory'; to the left: '(..) in the storage?'*

5.4.4. Prototyping game concepts

A first digital prototype was made from the audio adventure followed the concept of player one having an interactive map of the entire city who guides player two who is on the street and can only see the immediate surroundings. Concretely, player one had to catch crooks who walked around the city using an interactive map and player two's identification and descriptions of the crooks' surroundings. Figure 5.3 shows player one's interactive map. The prototype was developed for two tablets and tested while the players were communicating with each other in the same room. They were prevented from seeing each other's screen. Positively, this prototype motivated players to speak a lot, but the overall tempo of the game was too high for the target users. There were too many variables to process, and the crooks moved along too quickly before their surroundings could be sufficiently described.

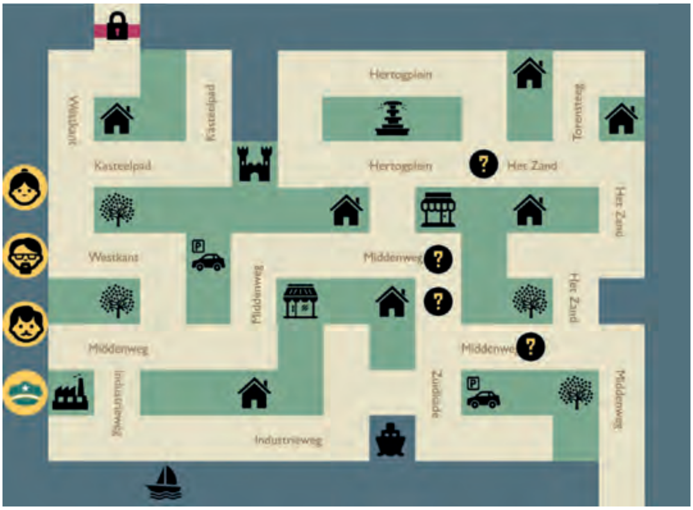


Figure 5.3.: Player one's interactive map in the 'catching crooks' concept. The question marks are one of the crooks on the left or player two and they keep moving around.

In subsequent prototypes, the pace of the game was reduced by changing the goal of the game, but sticking to the same concept. For example, finding lost dogs, finding and excavating artifacts, and solving detective-like mysteries. Positively, this concept made it easier for players to navigate the game world. However, it resulted in an imbalance in spontaneous speech between both players. In an attempt to resolve this, players were given equal roles by putting them both on the street, invisible to each other, but with access to the map. Subsequent user tests showed that the invisibility made the game more cognitively challenging. Also, access to the map caused a reduction in the amount of speech produced by both players as it had taken away the need to communicate for navigation. Removing the invisibility reduced the cognitive load, but decreased the need for players to communicate even further. To increase that need, sharing information received from other game characters was introduced. The players needed to find their own pieces of information and verbally share them with each other to accomplish the goal of the game. Final user tests showed that this increased the amount of speech produced to a sufficient level.

5.4.5. Game development

In accordance with guidelines I1-I3, the interface layout was kept consistent as well as the purpose of the interface elements. In the end, all buttons were labeled with a string representing their actions, because tested metaphors were often not understood.

During its development, the game content or story underwent several changes to optimize its adherence to guidelines C1-C5. Other story concepts in addition to the ‘catching crooks’ one described in the previous subsection were considered and tested. However, they either made no sense to players, did not motivate players to play daily, or were found too abstract. Players preferred a more everyday concept.

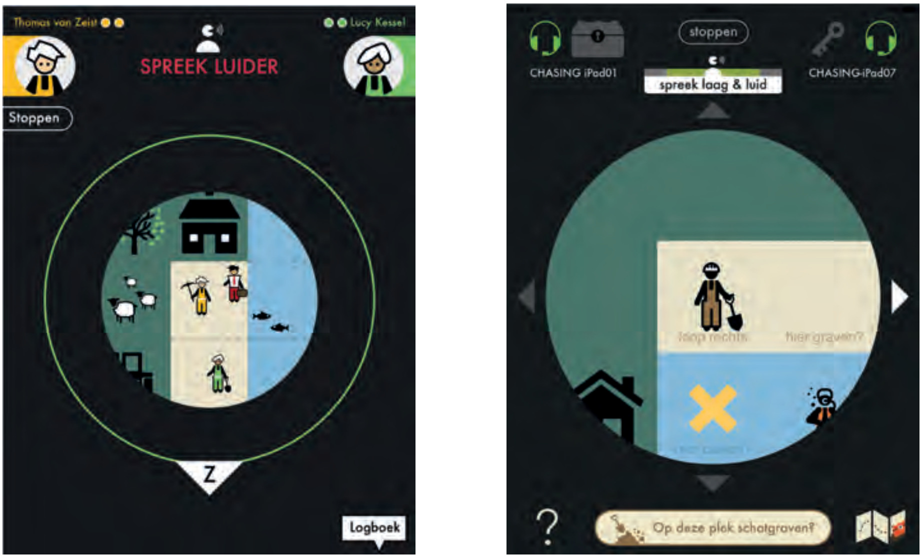
The final story concept ‘Treasure Hunters’ managed to motivate players and adhere to our guidelines. Players played two archaeologists that needed to work together to find treasures and artifacts. They navigated the same map, but started at different locations. Gathering clues about the location of the artifacts, helping each other to find the way, and getting to an artifact location together were also part of the story.

From a graphical perspective, the map was two-dimensional and divided in squares. To focus the player’s attention and reduce cognitive load, only a circular part of the map surrounding the player was made visible. Figure 5.4a shows how this was visualized to players. For navigation, they could use the buttons at the edge of the circle that were visualized similar to a compass.

This metaphor was chosen as we observed less confusion among players in orienting and providing navigational instructions. Real-time feedback on players’ speech was given automatically on loudness, pitch, and pronunciation. Initial game prototypes were tested that integrated indirect feedback with the gameplay. Those tests showed that players found this complex and difficult to process. In the end, a more direct approach using a loudness meter and textual notification was found easier to understand and process. Figure 5.4b visualizes this approach in the first version of our game (Ganzeboom et al., 2016).

The clinical trials (chapter 6) showed that players found it difficult to switch their focus between the playing field and the loudness meter. For that reason, we returned to user testing indirect methods. After multiple designs, the one shown in Figure 5.4a showed promising results and was included in the second version of the game. This way, players were intrinsically motivated to speak loud, as they would then see more of the playing field. To reinforce this method of feedback, a direct method was also included using textual notifications as shown at the top of Figure 5.4a.

As the results of the user satisfaction questionnaire showed chapter 7, players noticed the feedback and were able to use it.



(a) *Screenshot showing what player 'yellow' can see when speaking too softly. The inner circle widens as large as the green circle when speaking loud enough. Textual feedback is given at the top: 'Spreek luider' (English: Speak louder).*

(b) *Screenshot of the first game version showing the direct method of feedback at the top.*

Figure 5.4.: *Screenshots of Treasure Hunters version 1.*

5.4.6. Level design

Many aspects of the level design were taken into account to find a usable combination. For example, the starting positions of the two players, the clues provided by non-player characters (NPCs) that must be collected, the position of all the items describing the surroundings of the playing field, and the storyline within and across levels. Figure 5.5 shows the level editor that was created to design the levels for the game. All of them were individually tested on their difficulty to play, the amount of communication necessary between the players, as well as the build-up and variation in the level of puzzles presented. During these tests the observation was made that players spoke to each other the most when there was confusion about something in the level. Balancing the amount of confusion in each level was then used to motivate players to speak.

Players almost immediately started to play the levels in the initial prototypes with only a brief visual instruction provided by the game beforehand. Tests with players showed that such an instruction was too little for them and resulted in frustration

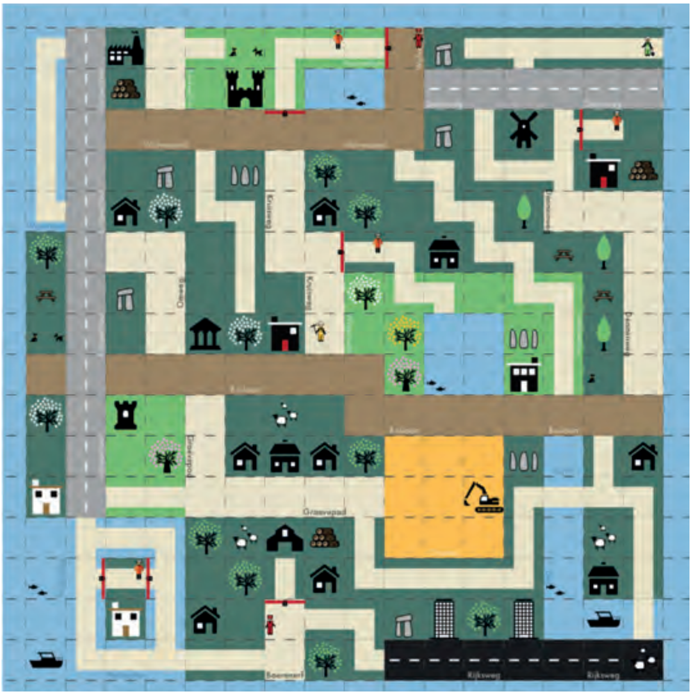


Figure 5.5.: *Level editor showing all the elements used in the level design: player's starting positions, descriptive icons, street names, and Non-Player Characters.*

with the game. For that reason, introductory levels were developed that provided a step-by-step explanation of how to play the game and how to use its interface. In the first version of the game (chapter 6), shown in Figure 5.4, players had to find the other player's colored X, denoting the location of a part of the treasure to be found. They then needed to guide the other player there before being able to retrieve the treasure. Over the course of the research it became clear that players found this repetitive and were less motivated to continue. For that reason, the second version of the game (chapter 7) introduced the collection of clues to solve puzzles or mysteries. These were given by NPCs to only one of the players in order to facilitate the need for sharing and speaking. The collection of clues allowed for more variety in the design of the levels making them less repetitive. It was also used to gradually increase the level of difficulty. For example, by making clues more cryptic using puzzles. As such, clues could be devised in a variety of ways and within many themes. This enables an easily expandable level design that can also be personalized to players' interests and required speech exercises.

In our aim to intrinsically motivate players, the four principles of the Hooked Model were applied. The NPCs who had unread clues for the players were 'triggers' in this respect. The corresponding ‘action’ was to move to this NPC’s location to obtain the clue. What the clue would tell, was always a surprise and in that respect a ‘variable reward’. The player ‘invests’ in the game by spending additional time to find and share the other clues. The Hooked Model was also implemented in the view on the map. As shown in Figure 5.4a, players only saw the part of the map surrounding their character, which triggered exploration by tapping the navigation buttons. They got rewarded each time by seeing a different part of the map. Players invested in the game by spending additional time in exploring the map and gathering knowledge on its topography.

5.5. Discussion

As the result section showed, our research uncovered a number of important design decisions. The next subsections summarize them as lessons learned per area.

5.5.1. Game concept

During the concept design phase, the following was found about the users this research’ targeted. They prefer:

- A cooperative game as opposed to a single player or competitive game.
- Freedom of movement in a game world as opposed to a preset path to follow.
- Everyday concepts to abstract ones.
- A fast-paced game was found unsuitable.

5.5.2. Speech production

Consider the following to motivate players’ speech:

- Players should have equal roles or tasks in the game to balance the amount of speech between the two.
- Sharing information between players that contributes to achieving game goals motivates speech.
- Confusion between players can be used to increase the production of players’ speech. However, too much confusion could cause unwanted frustration.

5.5.3. Feedback on speech

To provide feedback on speech, the following is recommended:

- Providing an indirect method of feedback to this target user group in a real-time interactive game can be effective. Especially when integrated with the game view that is in the player's main focus. To not hinder that focus, it is recommended to not manipulate the complete view and its contents.
- Reinforcing an indirect method of feedback with a direct one can be beneficial.

5.5.4. Game introduction and navigation

In this area, the following was learned:

- Introductory levels that demonstrate to the player step-by-step how the game works proved beneficial for this target user group in motivating them to play.
- It is recommended to not use brief textual instructions with accompanying images for this target user group. Introductory levels were easier to follow.
- The compass metaphor was beneficial to reduce confusion in communicating navigation instructions.

5.5.5. Design methods

The following was learned in the area of designing and developing a serious game:

- As described in previous literature, following a design process that involves users at every stage is highly recommended.
- Prototyping with target users early on using existing apps or pen and paper prototypes is very useful to identify game concepts users are motivated to play.
- It is useful to apply a method of persuasion (e.g. the Hooked Model) to the design of the game to motivate users to keep playing.

The above lessons learned can be considered as guidelines to benefit future research into serious games for speech training. This research acknowledges that these design guidelines are limited to the target group described in the method section. For example, the game concepts preferred by these users may be different from those preferred by younger adults with neurological disorders. Also, younger adults may not need the elaborate introductory levels and can suffice with explanatory texts and/or visuals. Furthermore, the goal of this research was to explore the design of a serious game for speech training in older adults and how to design for this target group. For that reason, testing the consequences of design choices was limited to qualitative user observations and interviews.

Future research may study these design choices in a quantitative manner and obtain more insight into the effects on users' motivation to play, to speak, and the efficacy of different modes of feedback.

The patients turned out to prefer a cooperative game to a single player game. Our choice to design a game of their preference was reinforced by the idea that they may have to use it intensively for a long time. However, a cooperative game is more complex in design, development, maintenance and use, because every session requires the availability of a co-player. That could be an acquaintance of the patient or another dysarthric speaker but not everyone is probably equally suitable, which makes organizing game sessions even more complex.

To maintain users' level of speech throughout their lives, longer and more intensive training periods then used in this research would be needed. How this design performs under such circumstances is an interesting topic for future research.

5.6. Conclusions

Designing a serious game for speech training in older adults affected by a neurological disorder presented its challenges. These challenges were met by following the CoDesign process that involved users at every stage of the design. As a result, two versions of a serious game were developed and deployed in clinical trials that, in the end, proved to be beneficial to users' speech intelligibility. Looking back on the design process, valuable lessons were learned in five areas. The following highlights the most distinctive ones per area:

1. Game concept design: older adults prefer a cooperative game that uses everyday concepts.
2. Motivating players' speech production: both players should have equal roles in the game, and a moderate amount of confusion can stimulate players' speech.
3. Visualizing feedback on pronunciation: reinforcing an indirect method of feedback with a direct one is recommended.
4. Game introduction and navigation: do not use brief textual instructions with accompanying images to introduce the game to a new player. Introductory levels were easier to follow by older adults.
5. Design methodologies: prototyping with target users early and throughout the design stage is useful to identify game concepts users are motivated to play.

Guidelines were formulated from the lessons in the previous areas. They are intended to guide future research into designing serious games for speech training.

Appendix 5A Example of anonymized user portrait

What follows is an automatic translation (Google Translate, September 5, 2023) of a user portrait written in Dutch that was used in this paper's research. The automatic translation has only been reviewed globally by the first author. The second subsection has the original text in Dutch[1].

English translation

Portrait - Marcel

Male, age 58

Marcel is 58 years old and has his own consultancy firm, he works from home and visits his clients by car all over the country. He lives in Amsterdam with his wife. His three daughters have all left home now. Marcel gets excited about imagination, originality, people who have different perspectives on the same thing. For example, why don't we investigate placebos instead of drugs? A red box with placebos for pain, a green box for rest, etc.

Twelve years ago, the onset of Parkinson's Disease was discovered in Marcel. He received pills as a first treatment, but stopped taking them after 3 years. Then he underwent surgery, a deep brain stimulation operation and that worked: He now hardly suffers from tremors anymore and has integrated the disease into his life. But writing is becoming increasingly difficult, especially if you have to write on the whiteboard for work.

Marcel gets up at 07:00 and does his physio exercises before breakfast. In the morning between 10:00 and 12:00 he visits a customer and after lunch in the afternoon between 13:00 and 15:00 another customer. Occasionally, he also meets another customer, between 4 and 6 pm, but then the day is really full. He prefers to be home by 17:00 or 17:30. Then drink a whiskey to relax, knowing that one does not have to do anything anymore. Marcel enjoys that. Other fun moments in a day are: Getting up, watching a nice movie or laughing in front of the TV at the TV program 'De Wereld Draait Door' (English: As the world keeps on spinning'). He likes to read the newspaper on Saturdays. Marcel also often enjoys his work, especially the inspiring conversations and the variety between the customers appeal to him.

Marcel uses his voice a lot in his work. He gives presentations for groups and conducts consultations. Due to Parkinson's, his voice quickly becomes rather soft and hoarse. His voice also skips sometimes. The natural reaction is then to clear one's throat and stop talking hoarsely, but that has no effect of course. During

[1] From report https://waag.org/sites/waag/files/2018-06/Chasing-ontwerpproces-rapportage.pdf, last accessed on September 21, 2025.

speech therapy he's done home exercises for 2 to 3 months: voice exercises once a day, but that stopped afterwards. He must remember to give his voicing more power. He is not often aware when talking too softly. Feedback in a positive way from his environment works for him to get his attention back to his voicing. He is often absent, busy with thinking about work. "You don't quickly worry about how you walk or talk." "But now you really have to think about everything." The less pleasant moments in a day are often when he gets tired in the late afternoon and comes back from work, just before the whiskey. He also doesn't like heavy traffic on the road, having to rush or being late. "Going to bed extremely tired is no fun." After dinner, Marcel and his wife often watch TV or read together. They go to sleep between 22:00 and 23:00.

People Marcel sees every day are his wife and customers, but always different customers. Once a week he goes to the physiotherapist. Less frequently but every month he sees his daughters (22, 26, 29), friends, acquaintances and relatives. He has four brothers and two sisters. He sees the speech therapist once every three months.

They will be moving soon to a smaller town outside the city, where his wife has more social connections and can build her own life next to him, because of his impairment from Parkinson's.

Marcel plays very few games, the occasional crossword puzzle, also with his wife. He used to play golf or tennis on the Wii game computer, but not anymore. He sometimes did archery with others in a cottage in France, but it soon turned into work there, more or less: maintenance around the house or in the garden, mowing the lawn, digging, etc. When playing the board game 'Triviant' he cannot stay silent and just has to say the answer to a question that is posed to someone else.

He doesn't like rugby or football. He is more of an athletics or running guy. Having to hit something with precision is something that Marcel does not like.

Actually, he has always hated sports. He reluctantly does physical therapy, but it's fun once he gets started. When Marcel does not do his exercises in the morning, he is also immediately punished during the day: Then his back will hurt. The manual therapist no longer allows him to bend over or lift anything. "That gives you a stiff back, so don't do it." Actually, a game for speech exercises should also have something like this in it, thinks Marcel. "You have to do physio exercises, that has now become part of my daily system."

He wouldn't want to play a game with another patient. Preferably independent, not having to depend on others. It is your own responsibility to save yourself. What works for Marcel is to hear himself speak, then he hears how his voice sometimes sounds rather "dull, flat and slow". He'd like to have a more vibrant voice. A voice with power and variety. To be aware of this is often enough to do something about it. When Marcel has to give a presentation for a group, it works to consciously practice the first part at home. So that he primes himself to use the technique for a good clear sound, speaking loud and low.

A good reminder for him is if his wife just says "Hey, your voice is dropping again." Then you can do something about that.

Dutch original

Portret - Marcel

man, 58 jaar

Marcel is 58 jaar en heeft zijn eigen adviesbureau, hij werkt vanuit huis en bezoekt zijn klanten met de auto door het hele land. Hij woont in Amsterdam samen met zijn vrouw. Zijn drie dochters zijn inmiddels allemaal het huis uit. Marcel wordt enthousiast van verbeeldingskracht, originaliteit, mensen die een andere visie op hetzelfde hebben. Waarom onderzoeken we bijvoorbeeld geen Placebo's in plaats van medicijnen? Een rood doosje tegen pijn, een groen doosje voor rust, etc.

Twaalf jaar geleden is bij Marcel beginnende Parkinson ontdekt. Daarvoor kreeg hij eerst pillen, maar is daar na 3 jaar mee gestopt. Toen is hij geopereerd, een deep brain operatie en dat werkte: Hij heeft nu zo goed als geen last meer van trillen en heeft de ziekte geïntegreerd in zijn leven. Maar schrijven wordt steeds lastiger, helemaal als je op het whiteboard moet schrijven voor je werk.

Om 7.00 uur staat Marcel op en doet zijn fysio oefeningen voor het ontbijt. In de ochtend tussen 10.00 en 12.00 bezoekt hij een klant en na de lunch in de middag tussen 13.00 en 15.00 uur nog een klant. Heel soms daarna, tussen 16.00 en 18.00 ook nog een klant, maar dan zit de dag echt wel vol. Het liefst is hij om 17.00/17.30 uur weer thuis. Drinkt dan een whisky om te ontspannen en niks meer te hoeven. Daar geniet Marcel van. Andere leuke momenten op een dag zijn: Opstaan, een leuke film kijken of lachen voor de TV bij De Wereld Draait Door. Op zaterdag leest hij graag de krant. Ook van zijn werk geniet Marcel vaak, vooral de inspirerende gesprekken en de afwisseling tussen de klanten spreken hem aan.

In zijn werk gebruikt Marcel veel zijn stem. Hij geeft presentaties voor groepen en voert adviesgesprekken. Door de Parkinson praat hij snel nogal zacht en hees. Ook slaat zijn stem nog weleens over. De natuurlijke reactie is dan om te rochelen om niet meer hees te praten, maar dat heeft geen enkel effect natuurlijk.

Tijdens de logopedie deed hij wel gedurende 2 à 3 maanden zijn huiswerk: 1x per dag stemoefeningen. Nu niet meer. Hij moet zich herinneren om meer kracht achter zijn stem te zetten. Zelf is hij het zich niet zo snel bewust als hij weer te zacht praat. Feedback op een positieve manier uit zijn omgeving werkt bij hem, om hem weer even met zijn gedachten erbij te halen. Vaak is hij afwezig, bezig in zijn hoofd met werk. "Je bent niet snel bezig met hoe je loopt of praat. Maar nu moet je echt overal bij nadenken." De minder leuke momenten op een dag zijn vaak als hij moe wordt eind van de middag en terugkomt uit zijn werk, net voor de whisky. Ook houdt hij niet van veel verkeer op de weg, haasten of te laat komen. "Doodmoe in bed rollen is geen pretje.” Na het eten kijken Marcel en zijn vrouw vaak samen TV of lezen nog wat. Tussen 22.00 en 23.00 uur gaan ze slapen.

Mensen die Marcel dagelijks ziet zijn z’n vrouw en klanten, wel telkens andere klanten. Eén keer per week gaat hij naar de fysiotherapeut. Minder frequent maar toch wel iedere maand ziet hij zijn dochters (22, 26, 29), vrienden, kennissen en familieleden. Hij heeft vier broers en twee zussen. Eén keer in de drie maanden ziet hij de logopedist.

Binnenkort verhuizen ze naar een kleinere plaats buiten de stad. Daar heeft zijn vrouw meer aanspraak en kan ze haar eigen leven naast hem opbouwen, in verband met zijn aftakeling door Parkinson.

Marcel speelt heel weinig spellen, af en toe een kruiswoordraadsel, ook samen met zijn vrouw. Eerder speelde hij nog weleens op de Wii spelcomputer golf of tennis, nu niet meer. Met anderen deed hij in een huisje in Frankrijk wel eens aan boogschieten, maar ging daar dan al gauw meer aan het werk: Klussen aan het huis of in de tuin, grasmaaien, graven, etc. Bij Triviant kan hij zijn mond niet houden om met de oplossing te komen, wanneer de vraag aan iemand anders gesteld wordt.

Hij houdt niet van rugby of voetbal, meer een atletiek of hardloop type. Met precisie ergens op moeten slaan daar houdt Marcel ook niet van, eigenlijk heeft hij altijd een hekel aan sport gehad. Fysiotherapie doet hij met tegenzin, maar als hij er eenmaal mee bezig is, is het wel leuk. Wanneer Marcel zijn oefeningen niet doet ’s ochtends wordt hij gedurende de dag ook direct afgestraft: Dan krijgt hij last van zijn rug. Van de manueel therapeut mag hij niet meer bukken of iets optillen. “Daar krijg je een stijve rug van, dus niet doen.” Eigenlijk zou een spel voor spraakoefeningen ook zoiets in zich moeten hebben, vindt Marcel. “Bij fysio-oefeningen moet je wel, dat is nu in mijn dagelijkse systeem gesleten.”

Een spel zou hij niet met een andere patiënt willen doen. Liever zelfstandig, niet afhankelijk van anderen. Het is je eigen verantwoordelijkheid om jezelf te kunnen redden. Wat voor Marcel werkt is zichzelf terug horen, dan hoort hij hoe zijn stem soms nogal “saai, gelijkmatig en traag” klinkt. Hij wil meer leven in zijn stem terug, een stem met kracht en afwisseling. Het bewustzijn daarvan is vaak al voldoende om er iets aan te doen. Wanneer Marcel een presentatie voor een groep moet geven, werkt het om het eerste stuk thuis bewust te oefenen. Zodat hij daarin de techniek toepast voor een goede duidelijke klank, laag en luid.

Een goed geheugensteuntje voor hem is als zijn vrouw enkel zegt “Hé, je stem zakt weer weg.” Dan kun je er wat mee doen.

Appendix 5B List Of Game Goals

From the results of the creative workshop with speech therapists and the interviews with the stakeholders, the following list of game goals was derived, which guided the design process. Originally in Dutch[2] (second subsection), the first section contains a manual translation to English by the first author.

English translation

We would like a game for patients with dysarthric speech due to an acquired neurological disorder that:

- Can be played regardless of location, which also includes players' home environments.
- Can preferably be played platform independently (e.g. on tablets, smartphones, and desktop PCs).
- Provides speech exercises that fit in with daily life. For example, by integrating them in an already existing daily routine.
- Assists in improving players' speech by encouraging them to speak loud and low and provides appropriate feedback.
- Provides feedback on player's speech (live or recorded) and makes 1-on-1 feedback from a speech therapist in a remote location possible (potentially via a connection with the electronic patient record).
- Provides feedback by enabling the player to hear his/her own speech.
- Motivates and assists the player in setting goals and achieve small successes step by step.
- Provides clear exercises daily that are easy to do and defined in terms of playing time.
- Can be played over a longer period of time. Preferably, a few months or perhaps years.
- Does not exhaust the player.
- Can be both used during and after speech therapy treatment.
- Can be played individually or together with partner/family.

The game:

- Is light-hearted, flexible and positive.
- Motivates to keep trying.

[2] From report https://waag.org/sites/waag/files/2018-06/Chasing-ontwerpproces-rapportage.pdf, last accessed on September 21, 2025.

- Is optimistic and exudes confidence in the patient.
- Builds increasingly in contact with the player.
- Avoids constant confrontation and prevents frustration.
- Approaches the player in a mature way.

The game takes into account:

- Players' areas of interest.
- The difference in the playing pace between patient and partner/family.
- The physical and cognitive limitations of neurologically disordered players, like tremors, quickly fatigued, and fluctuating physical condition.
- Players' daily schedule.

Dutch original

Wij willen een game voor patiënten met dysartrische spraak naar aanleiding van een verworven neurologische stoornis, die:

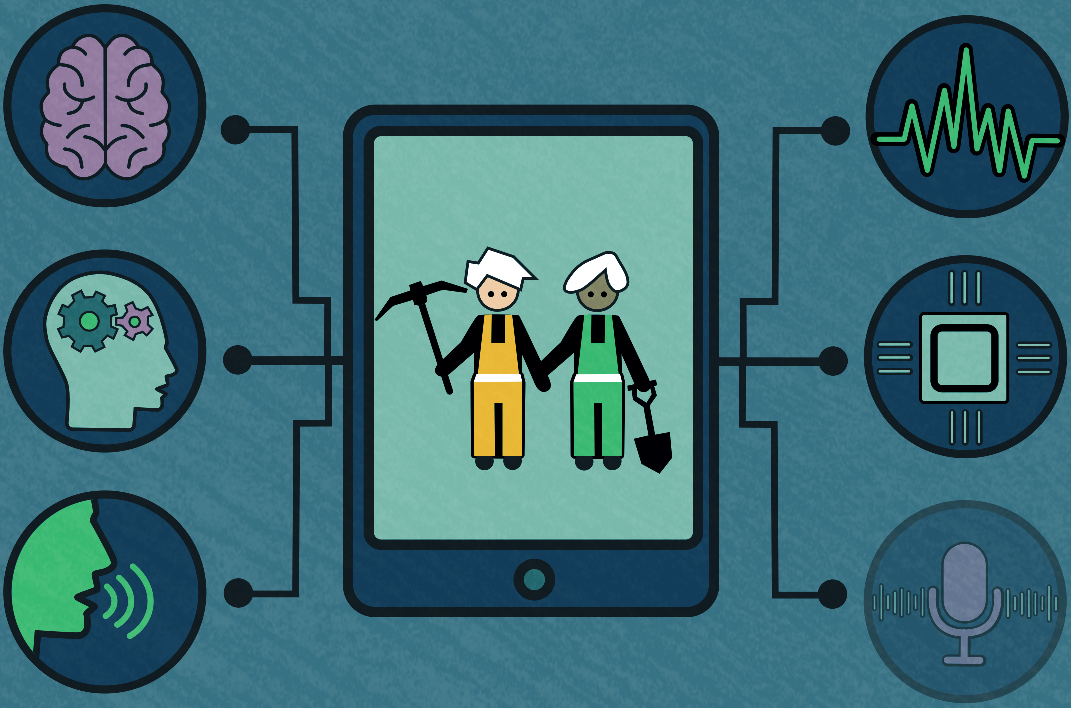
- Locatie onafhankelijk te spelen is, waaronder thuis;
- Bij voorkeur platform onafhankelijk te spelen (bijv. zowel op tablet, smartphone als desktop pc);
- Spraakoefeningen aanbiedt die aansluiten bij het dagelijks leven, bijvoorbeeld door de oefeningen te integreren in iets wat toch al 'moet';
- Stimuleert om laag en luid te spreken, feedback geeft en zo de spreektechniek van de speler helpt verbeteren;
- Feedback op de spraak (live/opgenomen) en 1-op-1 contact ("Hoe gaat het?") van een behandelend logopedist op afstand mogelijk maakt (evt. via mogelijke aansluiting op het Elektronisch Cliënten Dossier (ECD);
- Feedback kan geven waarbij de speler eigen spraak terughoort;
- Motiveert, de speler helpt doelen te stellen en stapsgewijs kleine successen laat behalen;
- Duidelijke en gemakkelijk uit te voeren oefeningen biedt voor iedere dag, afgebakend qua speelduur;
- Over langere tijd te spelen is, bij voorkeur enkele maanden of misschien wel jaren;
- De speler niet uitput;
- Zowel tijdens als na het logopedie behandeltraject inzetbaar is;
- Zowel individueel als samen met partner/familie te spelen is.

De game:

- Is luchtig, flexibel en positief;
- Motiveert om door te zetten;
- Is optimistisch en straalt vertrouwen uit in de patiënt;
- Bouwt een groeiend contact met de speler;
- Vermijdt constante confrontatie en voorkomt frustratie;
- Hanteert een volwassen benadering.

De game houdt rekening met:

- Interessegebieden van de speler;
- Verschil in speeltempo tussen patiënt en partner/familie;
- Fysieke en cognitieve beperkingen van Parkinsonpatiënten, zoals trillen, snel moe worden, schommelingen in de fysieke toestand;
- De dagindeling van de patiënt.



Chapter 6

Speech training for neurological patients using a serious game

This chapter has been reformatted and slightly modified from:
Ganzeboom, M., Bakker, M., Beijer, L., Rietveld, T. and Strik, H. (2018),
Speech training for neurological patients using a serious game.
British Journal of Educational Technology, 49: 761-774.
doi: 10.1111/bjet.12640.

6.1. Introduction

Patients with acquired neurological conditions such as stroke and Parkinson's disease (PD) often experience communication problems due to distorted speech, caused by dysarthria. Dysarthria is a motor speech impairment which negatively affects speech dimensions such as articulation and loudness (De Bodt et al., 2002). This results in diminished speech intelligibility in these patients, which often hinders communication in daily life. Traditionally, dysarthric patients regularly visit a speech therapist to exercise their speech. Given the increasing number of patients due to our ageing population and limited resources in healthcare, eHealth applications for speech training are gaining interest. Such computer-based applications enable patients to train their speech and receive feedback in their own home environment without the need to travel. In this way, speech training can be prolonged and intensified, which is known to have beneficial effects on motor speech rehabilitation (Maas et al., 2008; Palmer et al., 2007). Computer-based speech training systems have been investigated quite often in the last two decades (Y.-P. P. Chen et al., 2016; Palmer et al., 2007; Ritterfeld et al., 2016). The results indicate that computers could provide a method of delivering effective dysarthria training without placing high demands on therapy resources.

One of the computer-based speech training applications of interest is E-learning based Speech Therapy (EST), which focuses on drill-and-practice training for dysarthric patients with stroke or PD (Beijer et al., 2014; Schaefer et al., 2016). Exploratory research indicated the potentials of EST for beneficial effects and patients appreciated the possibility to train in their home environment at a time of their own choice (Beijer, 2012). Nevertheless, patients did not fully appreciate the usability of the web application and considered this an obstacle for frequent practice. They also indicated to prefer an ecologically more valid way to practice speech. That is, patients perceived the gap between drill-and-practice exercises on the one hand and functional daily communication too large. Also, the traditional exercises were not considered motivating for therapy compliance. These comments, questioning the potentials for therapy compliance and the appropriateness of exercises for daily communication, are typical for computer-based speech training (Y.-P. P. Chen et al., 2016; Palmer et al., 2007).

Rather than traditional drill-and-practice speech therapy, serious games have potential to trigger patients' intrinsic motivation for therapeutic practice through fun and enjoyment (e.g., Kari, 2017; Z. H. Lewis et al., 2016), thus enhancing therapy compliance. For both stroke and PD, the literature on rehabilitation using serious games primarily concerns physical fitness (these games are also known as exergaming). We refer to Appendix 6A for an overview of that literature. In the field of speech rehabilitation in dysarthric speakers, to the best of our knowledge only one study investigating the effectiveness of serious gaming has been reported (Krause et al., 2013). The game in this feasibility study focused on increasing loudness by producing isolated 'a' sounds in order to break a glass. Their work demonstrated the potential of integrating loudness exercises and game design to provide a useful tool for vocal training. The present study contributes to how serious gaming can be utilized to increase loudness in a more functional communicative context.

In the project 'CHallenging Speech training In Neurological patients by interactive Gaming' (CHASING) the goal was to develop a serious game (Treasure Hunters) for patients with dysarthria due to PD or stroke, aiming to improve their speech intelligibility in functional communication (Ganzeboom et al., 2016), and subsequently evaluate it. In the present pilot study we aimed to explore the added value of game-based speech training (using the game Treasure Hunters) compared to non-game computer-based speech training (EST). Two main research questions were addressed:

1. How does game-based speech training compare to non-game computer-based speech training with respect to dysarthric patients' speech intelligibility outcomes?
2. How does game-based speech training compare to non-game computer-based speech training with respect to patient satisfaction?

In the sections below we will describe the method and results of this experimental research project. Finally we will interpret and discuss the outcomes.

6.2. Methods and materials

6.2.1. Design

As shown in Figure 6.1, a crossover repeated measures design was employed. This design was chosen because of the exploratory nature of our study and the advantage of participants using both EST and the game. Each participant received both a four-week game-based speech training intervention (using the game Treasure Hunters) and a four-week non-game computer-based speech training intervention (EST). Intervention order was counterbalanced by having one group of participants start with the game intervention (group 1) and another group with the EST intervention (group 2). Repeated measures (i.e. speech recordings) were conducted before and after each intervention, at T1, T2, T3 and T4. Thus, the effects on speech intelligibility of both the game intervention and the EST intervention could be established as well as the effects of the within-subject variable 'game versus non-game (EST)' for each participant. At T4 participants also completed a user satisfaction questionnaire and a paired comparisons preference task (hereafter denoted as 'preference task').

6.2.2. Participants

Patients with dysarthria due to PD or stroke were recruited via speech pathologists. The patients had completed face-to-face speech training at least six months before.

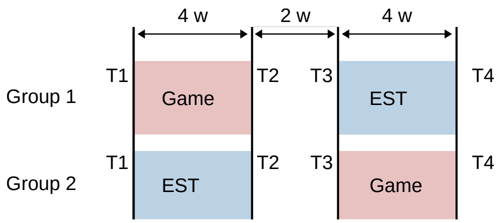


Figure 6.1.: *The crossover repeated measures design used to study the effects of the game and EST interventions (w = weeks). T1-T4 are the speech pretests and posttests.*

Excluded were patients with aphasia, reported severe cognitive problems or other disabilities that would hamper 15 minute training sessions with the game or EST.

Seven participants (five PD, two stroke), all male, were found willing to participate. Five of them (three PD, two stroke) completed both interventions and could be included in our study. Two participants had to withdraw due to health related reasons. Participant characteristics are provided in Table 6.1.

Participant	Gender	Diagnosis	Age (yrs)	Time since diagnosis (yrs)	Mobility limitations	Perceived impact on daily communication	Experience with computers
				Group 1			
p1	male	PD	62	15	none	large	considerable
p2	male	PD	59	7.5	none	large	little
p3	male	PD	69	4	severe	moderate	little
				Group 2			
p4	male	stroke	68	0.75	none	moderate	considerable
p5	male	stroke	67	3	none	none	hardly

Table 6.1.: *Participant characteristics. Participant 3 was additionally diagnosed with Hereditary Motor Sensory Neuropathy.*

6.2.3. Speech training interventions

Both the game intervention and the EST intervention were based on the Pitch Limiting Voice Treatment (PLVT; De Swart et al., 2003), which is currently standard practice for dysarthria therapy in the Netherlands (Kalf et al., 2008). PLVT is an adapted version of the Lee Silverman Voice Treatment (Ramig et al., 2001) and focuses on speaking 'loud and low'. This therapy is known to positively affect voice intensity (perceived loudness) in patients with PD (De Swart et al., 2003). The aim of this therapy is to improve a patient's intelligibility: the extra effort required to increase loudness indirectly has a positive influence on articulatory function, which has a beneficial effect on intelligibility (Sapir et al., 2007). For this reason, stroke patients also potentially benefit from this increased effort. However, due to this increased effort patients often also raise their pitch. This could have a negative impact on voice quality. Therefore, the PLVT prescribes to speak ’loud and low’. In order to encourage patients to speak ‘loud and low’ in both our interventions, feedback is provided on loudness and pitch. In accordance to the prescriptions of PLVT both speech training interventions were scheduled 15 minutes four times a week during four weeks.

Game-based speech training: Treasure Hunters

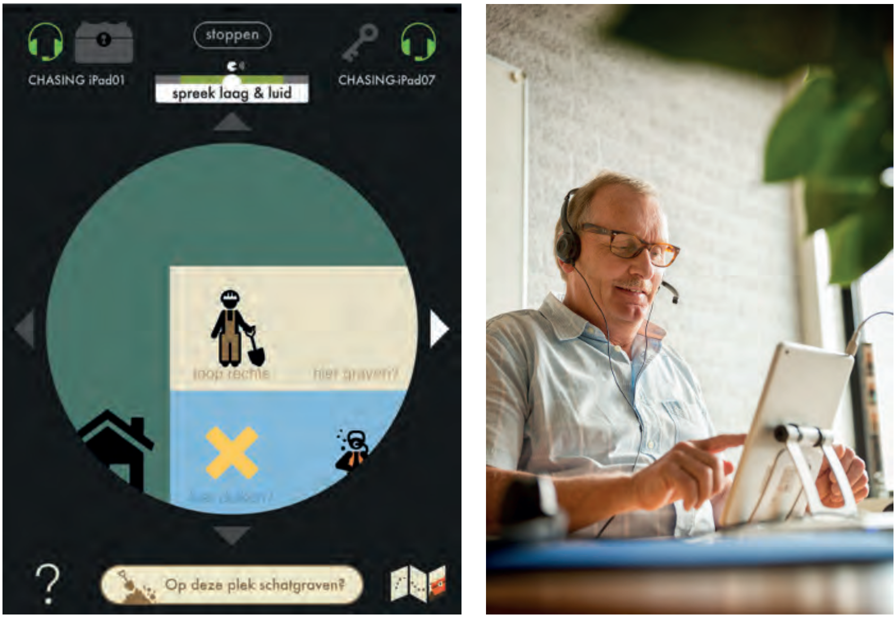
During the game intervention, the participants practiced their speech with the serious game Treasure Hunters (Dutch: ‘Schatzoekers’; Ganzeboom et al., 2016) which was developed in collaboration with Creative Care Lab at Waag Society[1], following a user-centred design approach. The game targeted elderly patients, which is largely representative of the population of patients with dysarthria due to PD or stroke. Both patients and speech therapists were involved in different stages of the design and development process.

Interviews and tests with different game concepts showed that patients preferred concepts with the emphasis on gaming, without a specific focus on the therapeutic aspects. In addition, multiplayer concepts were considered more appealing due to their social aspects. The interviews and tests also revealed that patients preferred a cooperative playing style. Additional details on the game design process are provided in Appendix 6B.

Treasure Hunters (see Figure 6.2a) is a two-player cooperative game in which players navigate a virtual map and need to help each other to find the treasure. One player plays the character ‘digger‘ and needs to find the treasure chest that is buried in the ground and the other player is a diver who searches the water for the key to open the treasure chest. The location of the chest is only visible to the diver and the location of the key is only visible to the digger. This way, both players are encouraged to speak, as they have to guide each other to the right location. Players are connected through the internet and communicate using a voice chat connection. While playing, the players receive automatic feedback on voice loudness and pitch, based on PLVT practices. This feedback is based on automatic analysis of the audio from the voice chat connection employing speech analysis algorithms. Initial game prototypes integrated indirect feedback on loudness and pitch with the gameplay. User tests showed that patients found this complex and difficult to process. A more direct approach in a later prototype was found easier to understand and process and resulted in providing feedback in the form of a loudness meter, which turns red if the voice is too soft. Feedback on pitch is given in the form of ‘speak low and loud’ (Dutch: ‘spreek laag & luid’) notifications when the pitch is too high (see the top of Figure 6.2a; Ganzeboom et al., 2016). For all participants, adequate levels of

[1] Creative Care Lab’s project CHASING page: http://waag.org/project/chasing, last accessed on September 30th, 2025.

loudness were set at 60 decibel or above (based on Rietveld and Van Heuven, 2016) and for pitch below 170 Hz (based on Kalf et al., 2008).



(a) *Screenshot of the game Treasure hunters.* **(b)** *Example of a typical game setting. Photography: Radboud University / Dick van Aalst.*

Figure 6.2.: *Images of the game-based speech training.*

All participants played the game using the same model tablet (Apple iPad Air) on a desk stand and headset (see Figure 6.2b). They played the game together with a coplayer through the internet. The coplayers in this study were university student assistants, who were instructed to act as a cooperative coplayer and encourage the participant to speak. The reason for using student assistants was to minimize variability between participants in the way the intervention proceeded. Importantly, these coplayers were instructed not to give feedback on participants' pronunciation. They were only allowed to ask for clarification like in regular conversation (e.g. 'Could you repeat that?').

During the game intervention, game sessions of around 15 minutes were scheduled four times a week for four consecutive weeks. Due to health-related reasons, complex agendas, and technical issues not all participants completed all 16 game sessions: participants 1 to 5 completed 11, 16, 16, 13, and 15 sessions, respectively.

Non-game computer-based speech training (EST)

During the EST intervention, the participants practiced their speech individually by drill-and-practice speech exercises in the EST application. In EST (see Beijer et al., 2014), the patient reads utterances, listens to target audio samples of these utterances, imitates these target samples by reading aloud the utterances, and aurally compares their own speech with the target sample. The latter is supported by automatic visual feedback on loudness and pitch employing the same type of speech analysis algorithms and the same loudness and pitch thresholds as in the game intervention.

The speech utterances used in the EST intervention were adapted from Beijer et al. (2014). All participants practiced with EST individually in their home environment, using the same model laptop and headset. Similar to the game intervention, the EST intervention comprised four sessions each week during four consecutive weeks. Sessions could be planned by the participants themselves within a two day time period; each session included a specific set of exercises in EST which took about 15 minutes to complete. Participants were contacted once a week by telephone to review their progress. All five participants completed all 16 sessions.

6.2.4. Measurement instruments

Speech recordings

As Figure 6.1 shows, participants' speech was recorded in their home at four points in time (T1-T4). An additional recording session was held some weeks before T1, which served as a feasibility test (i.e. test internet connection, find a low noise recording location, and fill out information and consent forms). The recordings at T1 and T2 served as pretest and posttest for the first intervention, while the recordings at T3 and T4 served the same purpose for the second intervention. Participants' speech was recorded using the same laptops and headsets as used for the EST intervention.

In the recording sessions, participants were asked to read aloud sentences and texts. The same sentences and texts were used in all recording sessions. The materials to be read aloud comprised 30 sentences containing a word with /p/, /t/, or /k/ as the initial sounds (Beijer et al., 2014), the short story 'Papa en Marloes' (Van de Weijer & Slis, 1991), and an Apple pie recipe text (constructed by the authors) to stimulate the realisation of more functional speech (participants were asked to imagine reading the recipe to a friend who was busy baking the pie).

During the speech assessments the recording of the speech materials was restricted to a single attempt, except for cases where reading errors, stutters, or restarts occurred. To limit their occurrence, participants read each sentence or text silently before reading aloud.

Listener judgements of intelligibility

To evaluate the effects of the interventions on participants' speech intelligibility, speech samples recorded in the speech assessment sessions were judged on intelligibility by inexperienced listeners using a paired comparisons judgement task. For each participant, eight inexperienced listeners (students from a university of applied sciences) compared pairs of speech samples recorded at times before and after the interventions (i.e. T1-T2 and T3-T4, respectively). They were asked which of the two samples was most intelligible to them and to what extent. In accordance with Scheffé's procedure for paired comparisons (Scheffé, 1952), ratings were assigned using a scale from -3 (the first realization was much better) to +3 (the second realization was much better), excluding 0.

Speech samples for the judgement task were selected from the recorded speech materials. Generally, the problems dysarthric speakers experience (i.e. decrease of loudness, increase of pitch and articulatory imprecision) often occur towards the end of sentences due to effects of fatiguing speech organs. For that reason, it is likely that the first parts of the speech samples do not show a potential reduction in intelligibility. That is why we chose to cut the /p/, /t/, /k/-sentences into two parts of comparable length and to include the latter parts as much as possible. From the text fragments, we used parts containing 2 or 3 successive sentences, always excluding the first sentence from a fragment.

Recordings containing unrecoverable reading errors and (background) noise were excluded from the judgement task. Furthermore, the selected recordings were balanced in length, occurrence of non-frequent words, and number of occurrences of the /p/, /t/, and /k/ sounds and consonant clusters. The final set totaled 40 speech samples. All speech samples were normalized to an average loudness of 68 dB (calibrated to an artificial ear in dB(A)).

In total, every listener was asked to judge 80 pairs of speech samples of a randomly assigned participant (40 pairs T1-T2 and 40 pairs T3-T4). The samples were digitally provided to the listeners in an OpenSesame experiment (Mathôt et al., 2012). The order in which the two samples within a pair were presented was counterbalanced, and the order in which the different pairs were presented was randomized.

For pairs that were presented in counter-chronological order, the listener ratings were reversed, so that positive ratings always represent an improvement in intelligibility from before to after the intervention. The listener ratings for each pair were then averaged to obtain an intelligibility gain score per pair of speech samples.

User satisfaction and preference

To evaluate participants' appreciation of game-based speech training (Treasure Hunters) in comparison with a non-game computer-based speech training (EST), they were asked to fill in two user satisfaction questionnaires (one about each system) and complete a preference task.

The user satisfaction questionnaires, based on Beijer (2012, ch. 8), contained four items about either the game or EST, related to satisfaction with the interface, ease of use, attractiveness and the overall system. Each question was rated on a 10-point scale, ranging from '1' (extremely unsatisfied) to '10' (extremely satisfied). Internal consistency of the 4 items was high (game: Cronbach's alpha = .82; EST: Cronbach's alpha = .84). A fifth question was added enabling dysarthric speakers to add written comments. The questionnaires about the game or EST were administered during the recording session directly after the respective interventions.

In the preference task, based on Beijer (2012, ch. 8), the participants were asked to choose between game-based speech training (using a two-player tablet game) and non-game computer-based speech training in four hypothetical scenarios. Hypothesized levels of speech improvement were attached to the two interventions and were either a 'slight improvement' (+-) or a 'strong improvement' (++). An example of a scenario is shown in Figure 6.3, where the participant had to choose between a game-based intervention with strong hypothetical improvement and a non-game computer-based intervention with slight hypothetical improvement. Each scenario was rated on a scale from -3 (strong preference for the left option) to 3 (strong preference for the right option). The ratings showed sufficient internal consistency (Cronbach's alpha = .77). The hypothetical scenarios were presented visually to participants through E-prime 2.0 and were orally explained by the experimenter.

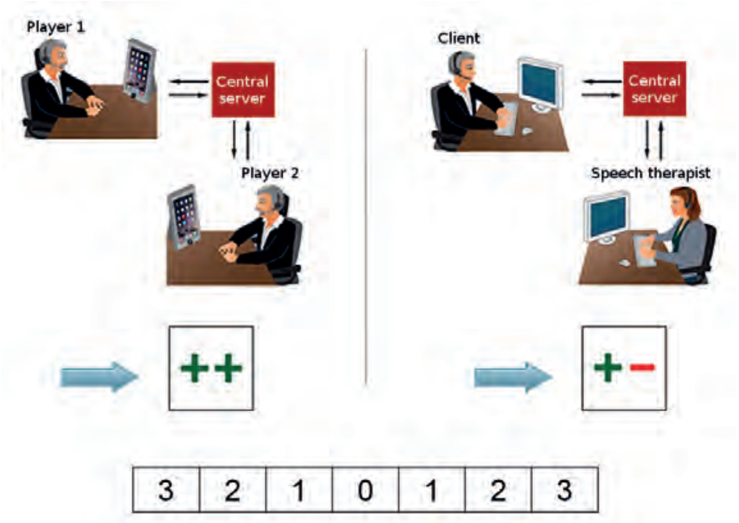


Figure 6.3.: Example of a scenario in the preference task.

6.2.5. Data analysis

For each participant and each intervention a one-sample t-tests was carried out to test whether the mean intelligibility gain score obtained from the listener judgement task was significantly different from 0. The unit of analysis was the speech sample pair (n = 40 speech sample pairs per participant per intervention, or less in cases where some speech samples could not be used). Kolmogorov-Smirnov's test did not reveal significant deviations from normality[2] and also skewness and kurtosis values were not significant. As we hypothesized that the interventions would positively affect patients' intelligibility, we employed one-tailed tests.

The intelligibility gain scores of the game intervention and the EST intervention were compared using a paired samples t-test for each participant. Also here there were no significant deviations from normality. As we did not have a hypothesis about which of the two interventions would result in higher improvements in intelligibility, we used two-tailed tests here.

All analyses were carried out with SPSS, version 23. A significance level of 0.05 was adopted. Cohen's d effect sizes were calculated for all analyses.

For the user satisfaction data and the user preference data (research question 2), descriptive statistics were calculated.

6.3. Results

6.3.1. Intelligibility

Table 6.2 displays for each participant the means and standard deviations of the intelligibility gain scores obtained from the listener ratings, both for the EST intervention and for the game intervention. Positive scores indicate an increase in intelligibility from pretest to posttest; negative scores indicate a decrease in intelligibility.

The one-sample t-test results for the game and EST (Table 6.2) indicate a large variation between participants. For participant 1 there was a marginally significant increase in intelligibility during the game intervention (d = 0.22) and no significant change in intelligibility during the EST intervention. Participant 2 had a marginally significant increase in intelligibility both during the game intervention (d = 0.26) and the EST intervention (d = 0.25). Participant 3 showed a significant decrease in intelligibility during the game intervention (d = -0.53) and a significant increase during the EST intervention (d = 0.36). Participant 4 had no significant changes in intelligibility in any of the interventions. Finally, participant 5 showed no change

[2] Except for participant 5 for the game intervention. For this case, however, a bootstrapped t-test led to similar results as the regular one-sample t-test.

Participant	Intelligibility gain score game intervention				Intelligibility gain score EST intervention				game vs. EST	
	n	M (SD)	t^a	d	n	M (SD)	t^a	d	t^b	d
p1	40	0.16 (0.70)	1.41†	0.22	40	0.16 (0.95)	1.07	0.17	-0.02	0.00
p2	40	0.18 (0.67)	1.65†	0.26	39	0.16 (0.64)	1.59†	0.25	0.26	0.04
p3	39	-0.44 (0.84)	-3.28**	-0.53	29	0.37 (1.02)	1.96*	0.36	-3.65**	-0.68
p4	37	-0.10 (1.05)	-0.57	-0.09	40	0.01 (0.75)	0.08	0.01	-0.45	-0.07
p5	40	-0.05 (0.78)	-0.43	-0.07	40	-0.34 (0.67)	-3.24**	-0.51	1.70†	0.27

Table 6.2.: *Descriptives and t-test results of intelligibility gain scores per participant.*
n = number of speech sample pairs. [a]Positive t-values were tested one-tailed, as they were in line with our hypothesis. Negative t-values were tested two-tailed. df = n - 1.
[b]Tested two-tailed. df = smallest n - 1.
† p < .10.
** p < .05.*
*** p < .01.*

in intelligibility during the game intervention and a significant decrease during the EST intervention (d = -0.51).

The paired-samples t-tests comparing the mean gain scores for the game and EST (rightmost columns in Table 6.2) revealed that for participant 3 the increase in intelligibility was significantly higher for EST than for the game, while for participant 5 the impact on intelligibility was marginally significantly higher for the game than for EST. For the other three participants there was no significant difference between the game and EST in their impact on intelligibility.

6.3.2. User satisfaction and preference

The participants' average user satisfaction ratings for the game and EST are displayed in Table 6.3. We see that participant 1 was more satisfied with the game than with EST, while participants 2 and 3 were more satisfied with EST than with the game. Participants 4 and 5 were highly satisfied with both the game and EST.

Written comments by the participants explained low user satisfaction ratings for the game by the occurrence of internet problems during the game sessions (p1), and by not being used to working with a tablet computer (p3). The contact with the coplayer was mentioned by some participants as a positive aspect of the game intervention (p2, p4).

In the preference task, the participants were asked to choose between game-based speech training and non-game computer-based speech training in four hypothetical situations. Table 6.4 presents the participants' ratings from the preference task (positive scores indicate a preference for game-based speech training). We see that in scenarios where slight improvement (+-) was compared to strong improvement (++), most participants opted for the strong improvement intervention, regardless

Participant	Game	EST
p1	6.50	5.75
p2	6.25	6.75
p3	5.50	7.00
p4	8.75	8.50
p5	8.00	8.00

Table 6.3.: *Average user satisfaction ratings per participant. Average of four items on a 10-point-scale.*

of whether this involved game-based or non-game computer-based speech training. In scenarios where the levels of hypothetical improvement were the same for both options, some participants chose game-based speech training, while others chose non-game computer-based speech training. When looking at the average over the four ratings (rightmost column in Table 6.3), we see that, across scenarios, participants 2, 4 and 5 showed a preference for game-based speech training, while participants 1 and 3 showed a (slight) preference for non-game computer-based speech training.

Participant	Preference per scenario				Average preference game vs. non-game
	game++ vs. non-game++	game+- vs. non-game+-	game++ vs. non-game+-	game+- vs. non-game++	
p1	-2	-2	3	-3	-1.00
p2	3	3	3	-2	1.75
p3	-2	0	2	-2	-0.50
p4	2	2	2	-3	0.75
p5	3	3	3	3	3.00

Table 6.4.: *Preference ratings per participant.*
Ratings range from -3 to 3, with positive values indicating a preference for game-based speech training.
game = game-based speech training using a two-player tablet game. non-game = non-game computer-based individual speech training. ++ = strong hypothetical improvement. +- = slight hypothetical improvement.

6.3.3. Summary of results

The results from Table 6.2, Table 6.3, and Table 6.4 are summarized per participant in Table 6.5.

6.4. Discussion

In the current research we compared game-based and non-game computer-based speech training: our game (Treasure Hunters) and EST. We focused on three vari-

Participant	Intelligibility	User satisfaction	User preference
p1	Game: marg. sign. improvement (d = 0.22) EST: no change in intelligibility	Game: 6.50 EST: 5.75	preference for non-game (-1.00)
p2	Game: marg. sign. improvement (d = 0.26) EST: marg. sign. improvement (d = 0.25)	Game: 6.25 EST: 6.75	preference for game (1.75)
p3	Game: sign. decline (d = -0.53) EST: sign. improvement (d = 0.36)	Game: 5.50 EST: 7.00	slight preference for non-game (-0.50)
p4	Game: no change in intelligibility EST: no change in intelligibility	Game: 8.75 EST: 8.50	slight preference for game (0.75)
p5	Game: no change in intelligibility EST: sign. decline (d = -0.51)	Game: 8.00 EST: 8.00	strong preference for game (3.00)

Table 6.5.: *Summary of results per participant.*
Ratings range from -3 to 3, with positive values indicating a preference for game-based speech training.
game = game-based speech training using a two-player tablet game. non-game = non-game computer-based individual speech training. ++ = strong hypothetical improvement. +- = slight hypothetical improvement.

ables: intelligibility of dysarthric speakers, user satisfaction, and user preference. Overall, Table 6.5 shows a substantial variability in the outcomes for each variable. In consequence, no consistent relations between the three variables were established. The lack of consistent outcomes is not surprising due to the heterogeneous participant group. Different background variables such as underlying neurological condition and experience with computers may have interfered with the interventions, and might (at least partially) explain why not in all cases a significant improvement was observed. As described in the Method section, not all participants completed the intended 16 practice sessions and some experienced technical issues during practice. Furthermore, part of the experimental design was that participants carried out the tasks in different orders. We checked for effects of these factors, but could not find any clear relations.

Another reason why the improvements obtained were maybe lower than what would ideally have been possible is related to the threshold values used for loudness and pitch, which were recommended in literature, indicating the average perceived loudness of speech in normal conversation (i.e. 60 dB), and a threshold slightly above an average male's maximum pitch (i.e. 170 Hz), respectively. The feedback provided in the two interventions was based on these threshold values, which were applied for all participants. It might be the case that these thresholds are 'on the safe side' and do not particularly challenge dysarthric speakers in increasing their loudness and simultaneously lower their pitch. Furthermore, it probably would be better to use personalized thresholds, that might even have to be adapted within one patient over time. After all, the goal is that patients' speech improves, and if a patient's speech improves it might be better to make the thresholds for that patient somewhat more challenging; and vice versa if a patient's speech deteriorates. The literature on the personalisation of exergames also point in this direction (Goršič et al., 2017; Laver et al., 2018).

Obviously, there are clear differences between our game and EST. The advantage of a non-game based application like EST is that it is built on a standardized training procedure (as described in the method section) supporting neurological patients who commonly experience cognitive impairments. The required auditory feedback for comparing spoken speech with the target sample is supported by automatic visual feedback and provided for every spoken utterance. Therefore, EST provides feedback on each spoken utterance, which is not the case in our game. Communication in our game is more varied and complex, more like in daily-life communication. Where EST's 'drill-and-practice exercises' may not be supporting for therapy compliance, they may be more suitable for patients with cognitive demands. On the other hand, where dynamic daily-life communication is expected to enhance intrinsic motivation and thus therapy compliance, in some cases it may be too demanding for dysarthric speakers with cognitive constraints.

The main limitations of our study are the limited number of participants and the relatively short duration of the interventions (about 4 hours over a four-weeks period). After this limited amount of training, the participants' appreciation was on average similar to that of EST. Our study included only elderly patients as they are largely representative of our target population. Unfortunately, no female participants could be included. This may partially be related to the higher prevalence of PD among men. Consequently, the results of our study are limited to elderly male patients.

A relevant question is what would happen after many years of intensive practice. We believe that the concept of our game can remain engaging over a long period, because new visual content and types of virtual maps can easily be added to the game. Our game could be a fun and motivating alternative for general practice in addition to EST. The latter could be used for remedial exercises to train certain aspects of speech that benefit from exercises of a more 'drill-and-practice' nature. In the game patients talk to another person, the speech is more conversational, the ecological validity of training is expected to be higher, and it has the potential to provide a better transfer to real life situations.

6.5. Conclusions

In this exploratory study, the comparison of game-based speech training (Treasure Hunters) and non-game computer-based speech training (EST) yielded no clear evidence for overall differences between the two with regard to speech intelligibility and user satisfaction. The results among participants were mixed, sometimes having a positive tendency towards our game, sometimes towards EST. In general, participants appreciated playing our game, hinting at the potential of our game for speech training.

With regard to the three variables in our study: speech intelligibility, user satisfaction, and user preference, substantial variation between participants was observed in outcomes of these variables and their relations. This indicates that, a 'one size fits all' approach does not apply. Instead, a personalized approach to speech rehabilitation is needed, in line with the suggestions in studies on exergames, as was reported in the discussion. Regarding the PLVT aspect, thresholds for pitch and loudness could be adapted to match individual patients' needs (which may vary for different phases in the treatment). The other aspects of the intervention (e.g. type of speech exercises, difficulty, etc.) should be personalized as much as possible, taking into account the abilities and the preferences of the patients. To achieve this goal and to assess its effects more research is needed. In particular, research that includes larger numbers of participants will increase statistical power and will enable us to draw more robust conclusions. In addition to studying the immediate effects on intelligibility, the long-term effects (e.g. maintenance or after longer periods of training) should be assessed as well as the effects on therapy compliance. As only male participants in the age range from 59 to 69 years participated in this study, it would be interesting to also test the game with female patients and patients of different ages, and to investigate the impact of gender and age on the effects and appreciation of the game-based intervention. Such a study may also include investigating possible differences between patients with dysarthria due to different causes. A clinically relevant question for future study is which types of people the game-based intervention is most suitable for. As our study evaluated the feasibility of providing game-based feedback on loudness and pitch, a next step would be to investigate the effects of additional and different kinds of feedback.

Appendix 6A Literature overview on serious games for non-speech rehabilitation in patients with Parkinson's Disease or stroke

Serious games for rehabilitation in patients with Parkinson's Disease

In the literature on serious games for patients with Parkinson's Disease (PD), various types of games can be identified, focusing on the movement of different parts of the body: upper and lower extremities and postural control. Multiple examples of recent studies that focus on the rehabilitation of the upper extremities (i.e. shoulder, arm, and hand) in patients with PD exist. Pachoulakis and Papadopoulos (2016) demonstrated the feasibility of game design using the Kinect infrared sensor from Microsoft's Xbox commercial game console[3]. In the game, designed for patients with mild to moderate PD, patients pop balloons which drop along vertical posts using controlled arm gestures. Patients increase their score with every balloon they pop and receive continuous feedback on their performance. Various parameters to change the difficulty level were included. The telerehabilitation system developed in Cikajlo et al. (2017) also utilises the Kinect sensor. Patients with PD played a game in which they picked moving targets by moving their arm and hand (i.e. collecting apples from a tree as they were growing slow or fast). In a feasibility study that included 28 patients, Cikaljo et al. reported significant improvements in movement accuracy.

Examples of research utilising games for rehabilitation of the lower extremities and postural control also exist. Leblong et al. (2017) reported results on a four week self-rehabilitation program that includes the use of a tablet, motion sensor, and games. Participants trained using conventional exercises (e.g. hip and knee flexion, sit-to-stand, stretching, etc.). The motion sensor was used to register the movements and the games provided feedback on the movement. A pilot study showed an improvement on walking tests in all seven participants with mild to moderate PD. For improving postural control (i.e. to align the body and stabilize its centre of mass), Albiol-Pérez et al. (2017) reported on a study involving games in which weight-transference exercises in sitting and standing positions are practiced using the Wii Balance Board from the Nintendo Wii game console[4]. In these games, the balance board is used to avoid obstacles (e.g. in a racing game) or to exit a maze-like game and not fall into its traps. The pilot study included 15 patients with PD and showed a trend to an improvement in all sitting and standing positions.

In addition to physical rehabilitation, the potential contributions of serious games to cognitive rehabilitation of PD patients have also been subject of research. van de Weijer et al. (2016) described a protocol for a study in which participants play a

[3] Microsoft Xbox game console: http://www.xbox.com, last accessed on April 20, 2018.

[4] Nintendo Wii game console: https://www.nintendo.co.uk/Corporate/Nintendo-History/Wii/Wii-636022.html, last accessed on April 20, 2018.

game aiming to improve attention, working memory, episodic memory, psychomotor speed and executive function. Participants play a game in which they explore the ocean in a rover and complete specific tasks by taking pictures of fish. Garcia-Agundez et al. (2017) reported on the implementation of a game concept that uses a dance mat or a Wii Balance Board to navigate a city map, having to remember the target location they need to reach. This game aims to train mentally drawing, visuospatial and working memory functions. As with the previously described study, no effects of this cognitive training game have yet been reported.

For additional literature on serious gaming for rehabilitation in patients with PD, comprehensive literature reviews exist (Barry et al., 2014; Cancela et al., 2014).

Serious games for post-stroke rehabilitation

The same subcategories as used in research on PD can be identified in literature on post-stroke physical rehabilitation. Focusing on the rehabilitation of the upper extremities, Standen et al. (2017) designed and evaluated three different games in a randomized controlled feasibility trial. The games included moving a virtual spacecraft through a course of obstacles, releasing a ball to hit a target, and popping a balloon by grasping and moving it to a pin. Patients trained their arm, hand, and finger movements, respectively, whenever they wanted during a period of eight weeks. The positions of the arm, hand, and fingers were tracked by a virtual glove. In total, 18 patients completed the final outcome measures and positive effects on motor functions were found. Türkbey et al. (2016) reported a single-blind randomized controlled pilot study to evaluate the feasibility and safety of training the upper extremities with a commercial game console (i.e. Microsoft's Xbox game console with Kinect infrared sensor). Twenty participants were included in the study and those who were allocated to the experimental group received additional training using the game console, on top of the conventional training. During 20 sessions, they played two commercially available exergames that require the use of the upper extremities. The experimental group showed significantly greater improvement than the control group on post-therapy measures (i.e. Box and Blocks Test, Wolf Motor Function Test, and the Brunnstrom Motor Recovery Stage for the upper extremity).

For lower extremity rehabilitation, Ramírez et al. (2017) designed a digital version of a game of dominos that includes exercises for hip abduction and hamstrings. Participants control the game with a custom-made game controller: a shoe that contains an inertial motion sensor and measures the movements of the foot while performing the exercises. To put the domino stones in the desired position on the board, participants perform different hip abduction and hamstring exercises. The authors aimed to contribute to the facilitation of unsupervized lower limb rehabilitation and reported on feedback from user tests that included three stroke patients. Using this feedback, they improved the design of the shoe and the game. Ghanouni et al. (2017) reported on a pilot study exploring and comparing the effects of balance games on commercial gaming platforms (i.e. Nintendo Wii versus Microsoft Xbox with Kinect sensor). The study included 14 stroke patients that were divided into two groups which used either the Wii or Xbox platform. The games were played in a group setting once a week for a 16-week period. In both groups participants showed improvements in balance, mobility and daily function. No significant differences were found between using the two gaming platforms.

Like PD, stroke can also affect patients' cognitive functions. Gamito et al. (2017) studied the effectiveness of a virtual reality serious gaming environment for post-stroke cognitive rehabilitation. Twenty stroke patients were included in a randomized controlled study in which participants trained with exercises to improve their attention and memory functions. Overall results indicated positive effects in these areas.

Additional literature on serious games for stroke rehabilitation can be found in comprehensive literature reviews (Laver et al., 2018; Tamayo-Serrano et al., 2018).

Appendix 6B Treasure Hunters: a summary of the game development process

Authors original report: D.-S. Boschman, L. Loos, J. Ongering, Creative Care Lab, Waag Society, The Netherlands.
Author of this summary: M. Ganzeboom, CLS / CLST, Radboud University Nijmegen, The Netherlands

The game Treasure Hunters was designed and developed in collaboration with the Creative Care Lab at Waag society[5]. This summary of the game development process is based on a report that is soon to be published via their project page[6].

Creative design approach

The CoDesign approach (E. B.-N. Sanders & Stappers, 2014) to creative design was chosen for this project to integrate the different inputs of all stakeholders (e.g. patients, therapists, researchers, designers). Within this project, the different expertises and experiences from patients, their caring partners, speech therapists, designers, and researchers in the area of speech training, have been brought together by utilising various design tools (Van Dijk et al., 2011), ultimately resulting in a suitable and potentially effective game for speech training in patients affected by dysarthria. By using these design tools (e.g. interviews, focus groups, codesign sessions, etc.), users are part of the design process. The CoDesign approach is guided by the following principles (translated literally from the original report):

- "receive feedback and inspiration from a small number of users;"
- "involve users in every part of the design process right from the beginning;"
- "visualize application ideas at an early stage and test them using prototypes;"
- "work in multidisciplinary teams and share knowledge;"
- "imagine to be a future user and let that user experience the new possibilities."

Design process

Using the CoDesign approach with all partners, a game concept was developed that suited the target users' preferences and capabilities. The following subsections describe the different phases of the design process and their findings.

Phase 1. Getting to know the target users In the first phase of the design process, it was important to get to know the target users. In interviews and discussions with the researchers and speech therapists the characteristics of the target

[5] Waag's Creative Care Lab: http://waag.org/en/labs/creative-care-lab, last accessed on April 20, 2018.

[6] CHASING project page at the Creative Care Lab: http://waag.org/nl/project/chasing, in Dutch, last accessed on April 20, 2018.

user group were determined. Subsequently, a number of Parkinson's Disease patients were recruited via speech therapists and were asked to contribute to the development of the game. Portraits of these patients were derived that give a description of the user and what they do and experience on a typical day in their life. This helped our designers in designing an application for real persons in a real-world context.

In a codesign session with speech therapists, our designers obtained additional knowledge on the target users and the methods that are used in speech therapy for dysarthric patients.

Possible goals of the game were identified from the qualitative interviews, portraits, and codesign sessions. These goals were prioritized in a design session with the project partners.

Phase 2. Exploring game play mechanisms Game play mechanisms are used to motivate the user to play the game, take certain actions in the game, and keep playing the game. The interviews and codesign sessions from the first phase resulted in different game play mechanisms which were found suitable for speech training and appealed to the target users:

- to discover: by controlling an object with your voice or indicating a particular choice verbally;
- to test knowledge: answering knowledge questions or guessing games, like a quiz;
- to experience: immersion in a conversation or story by playing a role in a game;
- to perform: giving a speech, reading a story aloud, or creating a story with partner and (grand)children;
- to create: creating music by using your voice, e.g. drawing or building a fictitious world.

As part of this phase, existing games and apps that include these game play mechanisms were tested with target users. A discussion with the target users after the tests provided insight in the advantages and disadvantages of the mechanisms.

Phase 3. Developing different game concepts In the third phase, four different game concepts that include the previous game play mechanisms were developed and tested with target users:

News reader - Reading the current news aloud using a preprogrammed autocue. The user can advance from local, regional to national broadcasting companies and create a personal news reading schedule.

Voice coach - A coach that provides cues to improve a user's speech continuously during the day, like an angel on your shoulder whispering in your ear.

Audio adventure - Two users need to find each other in a virtual world and can do so by describing their surroundings and position to each other via a voice chat connection.

Strong stories (i.e. 'Sterke verhalen' in Dutch) - The positive elements from the previous three were integrated in this concept. The goal in this concept is to find your coplayer on a virtual, disc-shaped world. Users can only navigate by reading aloud a sentence that is provided with the objects shown on the left and right of the user's current position. Fading footsteps provide an indication of the other user's position.

Target users indicated during the test sessions that they liked the concept of an audio adventure most, because of the social aspects it included and the freedom of navigation. However, users indicated that they found it difficult to get a sense of direction regarding the virtual map. In subsequent game concepts different options for improvement were tested (i.e. larger icons on the map, landmark icons, more thematic icons). This game concept was further developed in subsequent phases.

Phase 4. Prototyping Prototyping is an important phase to determine the feasibility of the game concept in terms of how it can be realized and how the target users interact with it. The initial prototyping stages included paper prototypes, which have the advantage that they can be made and tested quickly without the need of programming. After a successful paper prototype, a digital version can be created to further develop and test the interaction with the game. The audio adventure game concept was appreciated by the users as it includes spontaneous conversational speech (as opposed to reading aloud sentences/texts), social contact through multiplay, both users need each other's information to collaboratively complete game goals, whereas the content enables personalization (e.g. maps with personalized icons and speech material for training).

In the initial digital prototype, the goal was to identify and catch moving thieves on a city map. User tests showed this to be too dynamic and high in pace for the target users. A calm, clear game environment that provides overview was required.

This was implemented in the second prototype by changing the game concept to finding a static object (e.g. a lost key) on a city map. One user had an overview of the city map and was tasked to guide the other in finding the lost object (e.g. guide the other user to a location or person with information). Using these changes, the navigation and sense of direction of the users improved. However, it lead to an imbalance between users in the amount of speech: the user having the overview did most of the speaking by providing directions and descriptions, while the other user only asked for clarification every now and then and confirmed otherwise.

To solve this issue, the overview map was removed in the final game concept and both users got the same kind of game goals (i.e. search for an object on a city map). The users were required to find two objects that work together, a treasure chest and the key to open it. The users depended on each other because each user could only see the location of the other user's object, not the location of their own object. This game play mechanism improved the balance in the amount of speech.

Phase 5. Game development In the phase of game development, the most successful prototype was developed into a full game. Multiple iterations of designing, developing, and testing were necessary to obtain a good balance between game interaction, content, and audiovisual form.

For example, an accessible user interface is important in developing the interaction with the game. A user always has a slight learning curve to understand an interface and how to operate it. In order to ease this learning, a designer can build on target users' already existing knowledge of recognisable interfaces. However, the target users of this game were elderly people who have a relatively limited knowledge of tablet interfaces. Additionally, younger persons are generally less afraid to make mistakes as opposed to elderly persons. They are used to discovering how new digital media works using trial and error. To assist elderly persons in this direction, a metaphor from the analogue life can be helpful. Furthermore, the labels on the buttons in the final version of the game describe the button action from the user perspective.

The design of the game world has generally remained the same during game development: a flat, two-dimensional world in the form of a map with rectangular squares, similar to an analog board game. Controls for navigating the world has been visualized as a compass. The directions were limited to the north, east, west, and south, to support users' sense of direction.

An example of balancing the content of the game was the development of a motivating storyline. The storyline had to meet the following five criteria:

1. It stimulates speaking loud and low according to speech therapy principles.
2. It provides feedback in such a way that assists in improving a user's speech,
3. It is sufficiently challenging for practice at home for a longer period,
4. It remains engaging for four fifteen minute sessions a week for multiple weeks,
5. It is consistent with target users' general perception of the world.

Different storylines were conceptualized and tested, but often could not meet one of the criteria. However, the final storyline of 'Treasure Hunters' could.

The target users have been taken into account in the audiovisual design of the game. Texts were provided in font sizes as large as possible without blocking the view on the rest of the game world. The interface elements and objects in the game world have been designed with high contrast and colour blindness in mind (e.g. also different in shape besides colour).

Phase 6. Level design Most important in the whole design of the game may be the design of the levels, as many variables come together at this point: positions of the icons, start positions of both users, which icons are visible to one user and not yet to the other, the meaning and syllable composition of the street names on the map, background colours of levels, etc. Every level has been tested on difficulty, required communication between users, and the increase and variety in the level of problem solving required. Various tests showed that the amount of communication between users is highest when there is confusion about a part of the level. The challenge is therefore to design the right amount of confusion.

To support the target users in getting acquainted with the game interface and interaction, introductory levels explaining the navigation and other game mechanisms were added.

Levels also slowly increased in difficulty. For example, in the initial levels users could see each others' game character on their own tablet. In later levels this was removed to increase the amount of speech required to describe one's location to the other. In the beginning, users could also consult a map that provided a detailed overview of the game world, including the location where to guide the other. The overview map was gradually reduced in detail and finally removed in later levels, to increase communication between users and exploration of the game world.

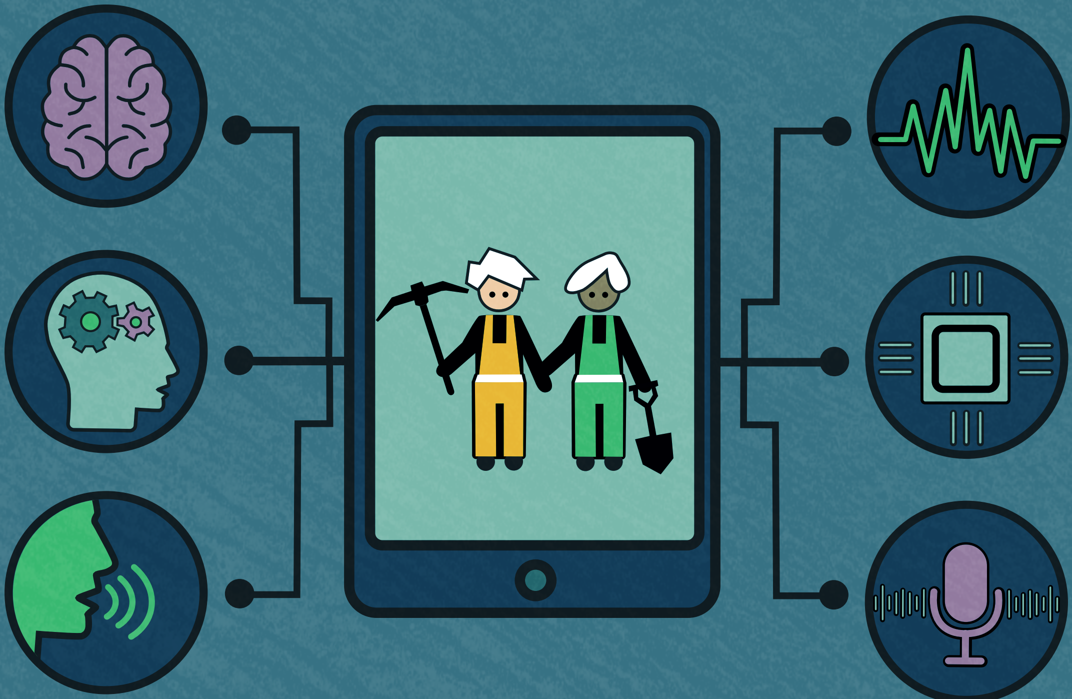
Phase 7. Motivating users to play The previous phases in the design process were responsible for producing the audiovisual appearance, the interaction, and the content. Subsequently, it is important that users would want to play the game over a longer period. In the ideal case, training speech with the game becomes a pleasant routine or habit if you will. In designing the game, principles of the Hook Model by behavioural specialist Nir Eyal[7] were used to motivate users to keep playing the game. The Hook Model describes the following cycle:

1. External trigger (e.g. a notification of a new message in a messaging app);
2. User action (e.g. a simple swipe to view the message);
3. Variable reward (e.g. a reward that is different every time and stimulates our curiosity. What was the message/photo we received?);
4. Investment (e.g. we make an investment in the messaging app by sending a response after which we will be more inclined to return to the app). Next, we wait for an external trigger again, which will restart the cycle.
5. After some time, the external trigger may become an internal trigger. We will then be intrinsically motivated to keep returning to the app (e.g. check the messaging/mailing app without having received an external trigger).

This model was used to design the interaction with the game. For example, part of the map is hidden from the user's view to stimulate exploration. At every step it

[7] Nir Eyal's website: http://www.nirandfar.com, last accessed on April 20, 2018.

will be a surprise to the user what will appear (i.e. variable reward). Perhaps your coplayer, or a hint to where to go next, or the location of the object that you need to guide the other to? If the map would have been fully visible to the user from the start, there would be no reward in exploring and users would have little need for communication.



Chapter 7

A serious game for speech training in dysarthric speakers with Parkinson's disease: Exploring therapeutic efficacy and patient satisfaction

This chapter has been reformatted and slightly modified from:
Ganzeboom, M., Bakker, M., Beijer, L., Strik, H., & Rietveld, T. (2022).
A serious game for speech training in dysarthric speakers with Parkinson's disease: Exploring therapeutic efficacy and patient satisfaction.
International journal of language & communication disorders, 57(4), 808–821.
doi: 10.1111/1460-6984.12722.

7.1. Introduction

The application of eHealth for patients with acquired neurological diseases has been gaining interest in the field of rehabilitation for the last two decades. eHealth is "the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research.", (World Health Organization, 2005, p. 121). It provides potential for the increased recovery of motor skills required for mobility and speech (Beijer et al., 2014), which are frequently affected due to neurological conditions such as Parkinson's disease (PD). In close to 90 percent of patients with PD, motor skills involved with speech are affected, resulting in dysarthria (Moya-Galé & Levy, 2019). Dysarthric speech is often characterized by impaired articulation and decreased voice intensity (De Bodt et al., 2002), negatively impacting speech intelligibility and hindering daily communication.

Dysarthric speech in patients with PD is known to benefit from intensified and long-term training (Ramig et al., 2001). The benefits of short-term intensive speech training have also recently been studied (Mendoza Ramos et al., 2021) with some encouraging results. The current standard practice for dysarthria therapy in The Netherlands is the Pitch Limiting Voice Treatment (PLVT; Kalf et al., 2011), an intensified long-term training program. In line with the Lee Silverman Voice Treatment (LSVT; Ramig et al., 2001), the PLVT focuses on increasing voice intensity. The resulting increase of articulatory effort and precision beneficially affects speech intelligibility (Sapir et al., 2007; Wight & Miller, 2015). However, a potential side effect is that patients often raise their pitch and laryngeal muscle tension as well. Therefore, the PLVT, unlike the LSVT, focuses on increasing voice intensity while keeping the pitch at a low and comfortable level. PLVT's goal 'speak loud and low' is known to positively affect voice intensity in patients with PD (De Swart et al., 2003).

As described by the Dutch government, the burden put on healthcare increases every year (Van Vilsteren et al., 2019). From this, a disbalance between speech therapeutical resources and the need for speech therapy may emerge that does not allow for highly frequent and long-term therapy, leading to a potential deprivation of optimal speech rehabilitation. As a consequence, dysarthric patients experience detrimental effects of low speech intelligibility and are highly motivated for practising their speech to support them in regaining societal participation.

eHealth provides the opportunity to both intensify and prolong speech motor training in patients' home environment. This allows self-management of speech training, and hence, the long-term improvement, and maintenance of enhanced speech intelligibility. Nevertheless, the absence of a therapist, providing instant feedback and a therapeutic relationship, and the need for patients to operate an eHealth device, may result in barriers for effective speech training. It is therefore vital to investigate the efficacy of eHealth-based speech training before investigating how such barriers can be removed. The importance of providing evidence for eHealth-based speech training in patients with acquired neurological diseases is even more obvious given the current pandemic (2020-2022) of the coronavirus, which demands physical ("social") distance between therapists and their patients for an - as yet - unknown period of time.

There are various ways of practising speech through eHealth. A drill-and-practice method was employed and investigated in a web-based speech application (Beijer et al., 2014). Data in that study indicated beneficial effects on speech intelligibility and user experience. However, a key point of feedback provided by the patients was the lack of variation in speech training exercises and the poor ecological validity of the application. That is, they reported being able to follow the program, but did not experience beneficial effects in daily-life communication. A plausible explanation is that the drill-and-practice nature of the exercises provided by the application may lack the transfer of the improved speech skills to daily-life situations. Also, a drill-and-practice approach challenges patients' therapy compliance, which is a key requirement for maintenance of regained speech intelligibility. In this respect, it is worthwhile to explore alternative approaches for remote and independent speech training. One such approach is the use of serious games. A widely accepted definition of serious games is: "games that do not have entertainment, enjoyment, or fun as their primary purpose.", (Laamarti et al., 2014). In many cases, entertainment, enjoyment or fun is used to serve the primary purpose of a serious game. For example, previous research showed that serious games can increase enjoyment during training (Kari, 2017; Z. H. Lewis et al., 2016), and have the potential to trigger patients' intrinsic motivation for therapeutic practice, enhancing therapy compliance as a result. Additionally, compared to more drill-and-practice eHealth applications, serious games allow speech training in ecologically valid scenarios that are more suitable to transfer the improved speech skills to daily-life situations. For example, a scenario in which a patient needs to converse with another person. Given these potentials, the current study continues our exploration of how serious games can be utilized to improve patients' speech intelligibility.

In our project 'CHallenging Speech training In Neurological patients by interactive Gaming' (CHASING) a serious game (Treasure Hunters) for patients with dysarthria due to PD or stroke was developed and evaluated. The goal of this game was to improve speech intelligibility in functional communication of these patients (Ganzeboom et al., 2016). In our previous study (chapter 6), we explored the added value of game-based speech training (using the game 'Treasure Hunters') compared to non-game computer-based speech training. Treasure Hunters was a two-player co-operative game in which players navigated a virtual map and needed to help each other to find the treasure by exchanging information via speech. The game provided automatic feedback on loudness and pitch in accordance with PLVT prescriptions. The results of our previous study (chapter 6) were mixed. No clear evidence of the efficacy of the game-based speech training was found and there was substantial variation in user satisfaction. In our continued aim to find a modality in serious gaming with clear evidence of efficacy for the patients in question, an adapted version of 'Treasure Hunters' was developed ('Treasure Hunters 2'), which was evaluated in a different experimental design. The effects of the improved game-based speech training was compared with a period of no training instead of a different type of speech training. For the current study, we hoped to encourage more participants to join as they only had to follow one speech training program. 'Treasure Hunters 2' was redesigned to increase the intensity of the speech training by including game elements that required the pronunciation of longer utterances. Pronunciation exercises were also introduced in the game to address the difficulties dysarthric speakers have with articulation. In addition to automatic feedback on voice intensity and pitch, feedback on players' pronunciation was automatically provided by the Automatic Speech Recognition technology developed for this study.

We assessed the effects of our improved speech training on different aspects, firstly how the game affected speech intelligibility. Secondly, patient satisfaction with the improved game design was a critical component. In addition, we sought to gauge patient experience and interpretation of the new feedback visualization for loudness, pitch, and pronunciation. A final critical area of focus was patient preference for game-based speech training as compared to traditional face-to-face training. These interests led us to address the following four research questions:

1. Can a game-based speech training positively affect speech intelligibility in patients with dysarthria due to PD?
2. How satisfied are patients with dysarthria due to PD with game-based speech training?
3. How do patients with dysartria due to PD experience the feedback on pitch, loudness and pronunciation in a game-based speech training?
4. Do patients with dysarthria due to PD prefer game-based speech training to traditional face-to-face training?

7.2. Treasure Hunters 2

'Treasure Hunters 2' is based on PLVT and aims to provide intensive speech training focussing on increasing loudness at a low pitch, and improving articulation. It has been developed in collaboration with Creative Care Lab at Waag Society[1]. The game targets elderly patients, because this group is largely representative of the population of patients with dysarthria due to PD. Both patients and speech therapists were involved at multiple stages of the game design and development process.

[1] Project page at Waag's Creative Care Lab: http://waag.org/project/chasing, Main project page: https://www.helmer-strik.nl/chasing/.

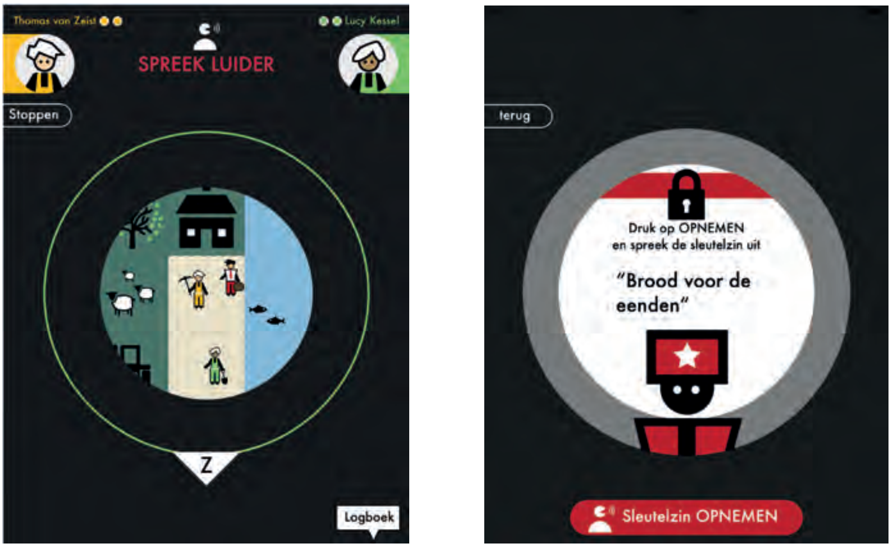
7.2.1. Game design

To integrate the practices of the PLVT, a serious game was designed that encouraged players to speak continuously and engage in highly frequent training. Additionally, it provided feedback that motivated players to speak louder at a low pitch and improve their articulation. In our previous research (Ganzeboom et al., 2016), we described that a cooperative game design facilitates intensive speech training. Players need to help each other to reach a common goal and are, consequently, encouraged to speak to each other continuously. Like the previous version of our game, Treasure Hunters 2 is also a two-player cooperative game. Players participated in different stories, such as one in which they are archaeologists searching for specific objects or another in which they are detectives tasked with solving a crime. Players saw themselves walking on a map and often did not see the other player immediately. They needed to describe their location and their surroundings to be able to help each other, encouraging them to speak. To increase the intensity of the training further, the game elicited longer utterances by requiring players to find their own specific clues that they needed to share verbally. Some clues hinted at how one player could help the other find the next clue, thus facilitating the need for sharing. Additionally, the element of opening gates or doors by correctly reading a sentence aloud was added as an exercise to train players' articulation. That was also beneficial to increasing the amount of longer utterances elicited by the game. Players played the game at a distance over the internet and spoke to each other via a voice chat connection.

7.2.2. Feedback on voice loudness and pitch

While playing, the players continuously received automatic feedback on voice loudness and pitch, by means of speech analysis algorithms. We merged the feedback on voice loudness with the circular shaped view of the playing field to have it in the focus of players' attention. The view grew larger and smaller depending on whether the player spoke loud enough or too soft. This way, we aimed to intrinsically motivate players to speak loud enough by rewarding them with a larger view of the playing field making more of the immediate environment around their character visible. Figure 7.1a shows the view on the playing field and the green circular shaped line to which it can grow. In order to add a more direct way of feedback, a 'speak louder' notification was added above the view. The 'speak louder' (Dutch: 'Spreek luider') notification is shown in red in Figure 7.1a. Feedback on pitch is provided exclusively via a 'speak lower' (Dutch: 'Spreek lager') notification at the same location. It is shown in blue to ease its identification.

Thresholds were used to determine when feedback on loudness and pitch should be given: for male and female participants, initial levels of intensity were set at 60 decibel or above (Rietveld & Van Heuven, 2016). When participants' intensity stayed above this threshold their view of the playing field grew. Oppositely, it started shrinking to its minimum size when the intensity moved below the threshold. When



Druk op OPNEMEN en spreek de sleutelzin uit

"Brood voor de eenden"

(a) *Screenshot of the game showing the 'speak louder' (Dutch: 'Spreek luider') notification in red including the green circle showing to which extent the view on the playing field can grow when speaking louder.*

(b) *The initial screen of the pronunciation exercise with at the centre, the sentence to speak aloud: "Bread for the ducks" (Dutch: "Brood voor de eenden") and the red button at the bottom to start and stop recording the exercise.*

Figure 7.1.: *Screenshots of the Treasure Hunters 2.*

this occurred, the 'speak louder' notification also showed. Initial levels for pitch were set below 170 Hz for male participants and below 260 Hz for female participants (Kalf et al., 2011). When participants raised their pitch above the threshold the 'speak lower' notification was shown. These values were additionally calibrated within the game with respect to environmental noise in a preparatory session in participants' home environment.

7.2.3. Feedback on pronunciation

In addition to feedback on loudness and pitch, Treasure Hunters 2 also provided feedback on pronunciation. Players had to read aloud a phrase to open a gate or door in the game and received feedback at word level on potential mispronunciations. We chose to provide feedback at word level assuming that feedback at a more detailed level (e.g. syllable or phoneme) may be too difficult for players to understand and interpret in a gaming environment. Players had three attempts in total to correct any detected mispronunciations. After the third attempt, the gate opened automatically, in order to not hinder the continuity of play. The screen starting the pronunciation exercise is shown in Figure 7.1b. It shows a short instruction at the top of the circle on how the player should proceed and the sentence to speak aloud in the middle, "Bread for the ducks" (Dutch: "Brood voor de eenden"). After having read the sentence in silence (to prevent possible hesitations), the player started the recording by pressing the red button at the bottom, and stopped the recording by pressing the same button again. Afterwards, the feedback screen displayed words in which a mispronunciation was detected in red, and words in which none were detected in green. Specifically for this game we developed Automatic Speech Recognition (ASR) technology that calculated confidence measures at both sentence and word level (Van Doremalen et al., 2010), which were used in the following two stages respectively. [1] Stage 1 verified whether the user actually attempted to speak the target utterance. This is to prevent the system from providing erroneous feedback on the displayed sentence. [2] Stage 2 detected mispronunciations at word level. If a word confidence measure was lower than a threshold, it was coloured red. Feedback was only provided on nouns and verbs (they provide the largest part of information). Prepositions, articles and pronouns were always coloured green. Tests with this procedure using artificially generated data showed that feedback was sufficiently reliable for our application.

7.3. Materials and Methods

7.3.1. Design

A single group repeated measures design was used to study the effects of our game-based speech training intervention (Figure 7.2). Since this study was of an exploratory nature and it was quite difficult to include matched participants in a control group due to their heterogeneous characteristics (Beijer et al., 2014; chapter 6), we refrained from a control group. Each participant received a four-weeks intervention using the game 'Treasure Hunters 2.' Repeated measures (i.e. speech recordings) were carried out before and after the intervention, at T2 and T3. Speech recordings were also made at T1, four weeks before T2. In this way, the effects of the intervention were compared to an equally lengthened period without intervention. It was also used to check the (recording and playing) conditions for the intervention at the participants' home environment, minimising and removing any noise sources (e.g. traffic, humming lights, ticking clocks, etc.). At T3 participants also completed a user satisfaction questionnaire and a paired comparisons preference task (hereafter denoted as 'preference task').

7.3.2. Participants

Patients with dysarthria due to PD were recruited via speech therapists, patient internet fora and Facebook groups. The patients had completed their latest face-to-

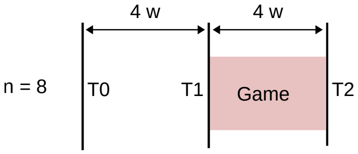


Figure 7.2.: *The repeated measures design used to study the effects of the game intervention (w = weeks). T2 and T3 were the speech pre-tests and post-tests.*

face speech training at least two months before T1. Exclusion criteria were aphasia, reported severe cognitive problems or other disabilities that would hamper 15 minute training sessions with the game. From the 15 participants that were found willing to participate, seven had to be excluded because of the aforementioned reasons.

Eight participants, five male and three female, all with dysarthria due to PD, completed the intervention and were included in our study. Table 7.1 provides the demographic data. All participants used Levodopa medication to improve motor functioning, except for participants 5 and 6 who used no medication. Participant 3 had a hearing aid. Participant 5 has received a Deep Brain Stimulation (DBS) system implant in the past and started experiencing difficulties with speaking afterwards. All participants using Levodopa were recorded in the ON condition at all measurement times. Participant 5 was equally recorded in the ON condition with the DBS implant turned on.

Participant	Gender	Age (yrs)	Time since diagnosis (yrs)	Mobility limitations	Perceived impact on daily communication	Experience with tablets
p1	m	73	4.5	little	large	a lot
p2	m	56	8.0	none	large	none
p3	m	60	4.5	none	large	considerable
p4	m	63	5.0	none	large	a lot
p5	f	53	9.0	none	large	a lot
p6	m	75	2.0	severe	large	considerable
p7	f	67	3.0	none	large	considerable
p8	f	62	3.0	little	large	a lot

Table 7.1.: *Participant self-reported characteristics. All participants experience dysarthria due to PD. Levodopa medication is also used by all, except for participants 5 and 6. Options for the participants' own perceived limitations on mobility around the home environment were 'none', 'little', or 'severe.' Options for the perceived impact on their daily communication were 'none', 'little', and 'large.' Options for the assessment of computer skills were 'a lot', 'considerable', 'little', 'hardly', and 'none.'*

7.3.3. Speech training intervention

In line with the PLVT protocol (De Swart et al., 2003), the speech intervention consisted of 15-minute practice sessions, four times per week for four consecutive weeks.

All participants played the game on the same model tablet (Apple iPad Air) on a desk stand, using a headset (Sennheiser PC 3 Chat) for voice communication. They played the game together with a coplayer through the internet. Coplayers were recruited university student assistants who studied speech and language therapy/pathology. They were instructed to act as a cooperative coplayer and to encourage the participant to speak and not to give any feedback on participants' pronunciation. They were only allowed to ask for clarification as in regular conversation (e.g. 'Could you repeat that?'). The student assistants were also instructed on how to explain the game to the participants and how to set it up in patients' home environments.

As shown in Figure 7.2, participants' speech was recorded in their home at three points in time (T1, T2, T3). Before the recordings at T1 the internet connection was tested, and a low noise location was selected. The recordings at T2 and T3 served as pre-test and post-test for the intervention. At T2 the coplayer explained and practiced the game with the participant.

7.4. Measurement instruments

7.4.1. Speech materials

In order to measure the development of participants' speech intelligibility, speech recordings were made of the same sentences and texts. During the three recording sessions, participants' speech was recorded while they read sentences and texts aloud. The same sentences and texts were read at all points in time. The materials to be read aloud were selected to facilitate the measurement of effects the intervention had on the intelligibility of daily speech. Those speech materials should also challenge participants in the affected speech dimensions (i.e. loudness and articulatory precision). Our selection comprised 30 sentences containing a word with /p/, /t/, or /k/ as the initial sounds (Beijer et al., 2014); we call the associated utterances /p/, /t/, and /k/ sentences. These sentences sufficiently challenged patients' maintenance of loudness levels and pitch as well as their articulatory precision. Additionally, the short story 'Papa en Marloes' (Van de Weijer & Slis, 1991) was included in our selection because of its focus on oral and nasal speech sounds. Lastly, the 'Apple pie recipe text' (chapter 6) was included to have a more engaging text that stimulates the realisation of functional speech. To that end, participants were asked to imagine reading a recipe to a friend who was 'busy baking the pie at the front of the kitchen while the participants were at the back.' The distance they had to cover would naturally encourage participants to speak loudly.

During the speech assessments, the recording of the speech stimuli was restricted to a single attempt. However, additional recording attempts were made if any reading errors, stutters, or restarts occurred as these would negatively bias later judgements of intelligibility. To limit their occurrence, participants read the text silently before starting the recording. Participants' speech was recorded at a sampling rate of 44.1 KHz and 16 bit PCM resolution.

7.4.2. Measuring the effect on speech intelligibility

The speech samples recorded in the speech assessment sessions were judged on intelligibility by inexperienced listeners in an adapted paired comparisons judgement task. We chose inexperienced listeners, as they represented individuals from daily life who are unfamiliar with participants and dysarthric speech in general. For each participant, ten inexperienced listeners – university students - were asked to judge the intelligibility of the speech samples recorded at all three time points (i.e. T1, T2, T3) as compared to the intelligibility of a neurologically healthy reference speaker. We chose to use neurologically intact speech as 'anchor' in the paired sample in order to facilitate judging the extent to which participants' speech deviates from typical speech. We used a female reference speaker for the female participants and a male reference speaker for the male participants, both with ages similar to those of the participants (67 and 69 years). The reference speakers recorded the same speech materials as the participants, using the same model laptop and headset as the participants.

In the paired comparisons judgement task, the speech sample of the participant was played first and that of the reference speaker second. Listeners were then asked to what extent the first sample was less intelligible than the second. They assigned their ratings using a 7-point scale from -5 (the first sample was considerably less intelligible than the second) to 1. On this scale, 0 represented both samples being equally intelligible and 1 to provide the possibility to indicate that the first sample was more intelligible than the second. A negative scale was used as we believed this would be more intuitive to indicate poorer intelligibility. Similarly, a positive integer could be used to indicate better intelligibility. Various pilot rating sessions showed no problems using this scale.

Speech samples for the judgement task were selected from the recorded speech materials. The recordings from the /p/, /t/, and /k/ sentences were used without changes. To increase the number of judgements for our analysis of efficacy, the recordings from the text fragments were cut at each sentence boundary, creating individual recordings that were rated separately.

Recordings containing unrecoverable reading errors and (background) noise were excluded from the judgement. To further prevent bias in judgements, a selection of recordings was made that was balanced in length (expressed in number of phonemes), occurrence of non-frequent words (lemma frequency < 10 per million), the number of occurrences of the /p/, /t/, and /k/ sounds and the number of consonant clusters. The final set contained 42 speech samples. All speech samples were normalized to an average intensity of 68 dB(A) (calibrated to an artificial ear, Brüel Kjaer 2610).

In total, each listener was asked to judge 126 speech samples of one, randomly assigned, participant (42 pairs per time point). The samples were digitally provided to the listeners in an OpenSesame experiment (Mathôt et al., 2012). The order in which the different pairs were presented was randomized.

7.4.3. User satisfaction

Participants filled in a user satisfaction questionnaire at time point T3, after finishing the intervention (see Appendix 7A). Google Forms was used to make this questionnaire available online. Five items of the questionnaire were based on previous research (Beijer, 2012, Ch. 8; Chapter 8). Four items asked the users about their satisfaction with the interface, ease of use, attractiveness and the overall system. Each of these were rated 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 (NUFFIC, 2022). The fifth question was open-ended and asked how users generally experienced playing the game.

At time points T1 and T2 it became apparent that most participants would find it difficult to fill in the questionnaire online due to physical limitations or fatigue. We therefore decided that the student assistants that made the speech recordings should read the questions aloud and make an audio recording of users' responses. Student assistants were instructed to limit the text they read aloud to the question on the questionnaire and, if necessary, the accompanying explanatory sentence.

7.4.4. User appreciation of feedback

Five open-ended questions were added to the user satisfaction questionnaire in order to measure the appreciation of the feedback on loudness, pitch, and pronunciation (see Appendix 7A). The questions regarding loudness and pitch assessed whether participants noticed the feedback, how they experienced its precision, and if they could act on it (i.e. change the use of their voice in accordance with the feedback). Questions regarding pronunciation probed users' general experience of the feedback and to what extent they disagreed with the provided feedback. Screenshots of the game accompanied these questions and were referred to in the question's text. The responses to these questions were audio recorded in the same session as the questions on user satisfaction.

7.4.5. User preference

In the preference task, based on (Beijer, 2012, ch. 8), we explored users' preference for game-based speech training relative to face-to-face training in scenarios where the two training methods lead to different hypothesized levels of speech improvement. The participants were asked to choose between game-based speech training (using a two-player tablet game) and traditional face-to-face speech training in four of such scenarios. Hypothesized levels of speech improvement were attached to the two interventions and were either a 'slight improvement' (+-) or a 'strong improvement' (++). An example of a scenario is shown in Figure 7.3, where the participant had to choose between a game-based intervention with strong hypothetical improvement and a face-to-face-based intervention with slight hypothetical improvement. Each scenario was rated on a scale from -3 (strong preference for the left option) to 3 (strong preference for the right option). The hypothetical scenarios were presented visually to participants through the E-Prime 2 software (Psychology Software Tools, 2012) and were orally explained by the experimenter.

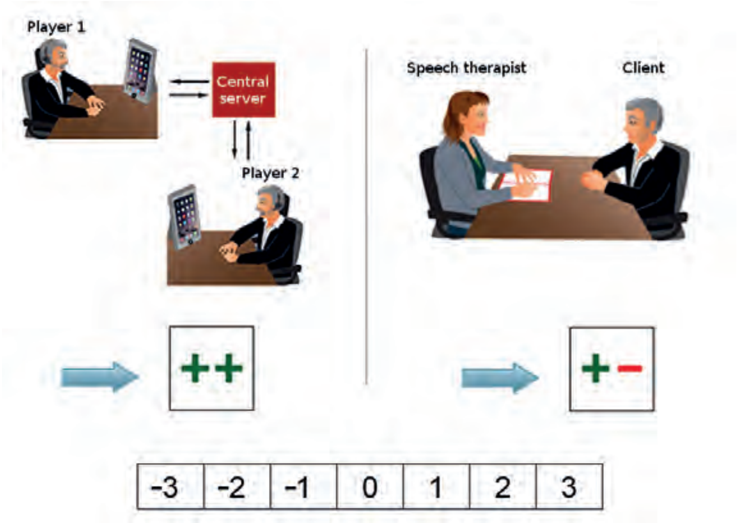


Figure 7.3.: *Example of a scenario in the preference task. The participant had to choose between a game-based intervention with strong hypothetical improvement (++) and a face-to-face-based intervention with slight hypothetical improvement (+-). Scenarios were rated from -3 (strong preference for the left option) to 3 (strong preference for the right option).*

7.5. Data analysis

To facilitate analyses, listeners' judgments were converted to a positive scale, 0 (the participant's sample was considerably less intelligible than that of the reference speaker) to 5 (both samples being equally intelligible). A higher score thus represents higher intelligibility. The original ‘1’ score, indicating that the participant’s sample was more intelligible than that of the reference speaker, was not often used (3.5% of total judgements). After inspection, we were confident that these were due to small regional variations in pronouncing individual speech sounds or recording conditions (i.e. breathing in the microphone and background noises). For that reason, the speech samples corresponding to these judgements were considered as equally intelligible and merged with the category ‘equally intelligible’ (5 in our converted scale).

From the 42 sentences recorded per speaker for every time point, 24 had usable recordings available for all three measurement times. The other sentence recordings could not be used because recordings of one or multiple measurement times contained an unrecoverable event (e.g. speech error, breathing in the microphone, environmental noise, etc.) that could bias listeners’ judgements. Consequently, judgements of 24 speech samples were used for every measurement time by 10 listeners, totalling 720 judgements per speaker.

For each of the eight speakers and for each time point (T1, T2, and T3) reliability of the listeners’ judgements was assessed by the Intraclass Correlation Coefficient, ICC(2,k). This version of the ICC family assumes that both raters and objects are random factors (Rietveld, 2021). Twenty-one of the 24 estimates of reliability (8 speakers 3 time points) were significant ($p < .05$), while three were not. The data of these three were further investigated on irregularities (i.e. possible errors in judgments, scale conversion errors, and abnormalities in judged recordings), but none were found. The mean value of the 24 estimates of reliability was 0.678.

The results of the listening experiment were analysed using the lmer function of the R package lme4 (Bates et al., 2015) for linear mixed effects modelling. The fixed factors in the model were Speaker and Time, and their interaction; the random factor was Utterance (random intercept). The effect sizes were measured with the index R^2Partial (the Kenward-Roger approach, provided by the R-procedure r2glmm[2]). To further inspect differences between time points, we carried out paired comparisons, using the R-procedure emmeans[3].

For the user satisfaction data, the recordings of the participants were listened to by the first author and the ratings they gave were noted in a table. Descriptive statistics were calculated, consecutively. The responses to the open question about how participants generally experienced Treasure Hunters 2 were summarized.

Equal to the user satisfaction data, the responses to the open questions about the appreciation of feedback on pitch, loudness, and pronunciation were also listened to and summarized.

For the user preference data, descriptive statistics were calculated.

[2] Refer to https://cran.r-project.org/package=r2glmm for more details on the r2glmm package.

[3] Refer to https://cran.r-project.org/package=emmeans for more details on the emmeans package.

7.6. Results

The intervention included 16 speech training sessions with the game. For health-related reasons and technical issues, not all participants completed all sessions: three participants completed all 16 sessions, three participants completed 15, one completed 14 sessions, and one completed 13 sessions.

7.6.1. The effect of the intervention on speech intelligibility

Speaker, Time, and their interaction were found to be significant factors ($p < .01$). The results of the fit by the R-procedure Restricted Maximum Likelihood (REML) are given in Table 7.2.

Factor	F	df1, df2	P-value	$R^2_{Partial}$
Speaker	49.77	7, 529	< 0.01	0.397
Time	9.655	2, 529	< 0.01	0.070
Speaker x Time	8.528	14, 529	< 0.01	0.062

Table 7.2.: Results of the REML procedure on the intelligibility scores.

The marginal means of each speaker at each time point are displayed in Figure 7.4.

Table 7.3 shows the results of our paired comparisons of intelligibility ratings carried out with the emmeans R-procedure.

Contrast	Estimate	SE	df	t-ratio	p
T1 - T2	0.104	0.0504	529	2.065	0.0982
T1 - T3	0.311	0.0504	529	6.184	<.0001
T2 - T3	0.207	0.0504	529	4.119	0.0001

Table 7.3.: Results of the paired comparisons of intelligibility ratings carried out with the emmeans R-procedure.

Whereas the difference of the scores obtained at T1 and T2 is not significant, that of the scores obtained at T2 (start of the treatment) and T3 was.

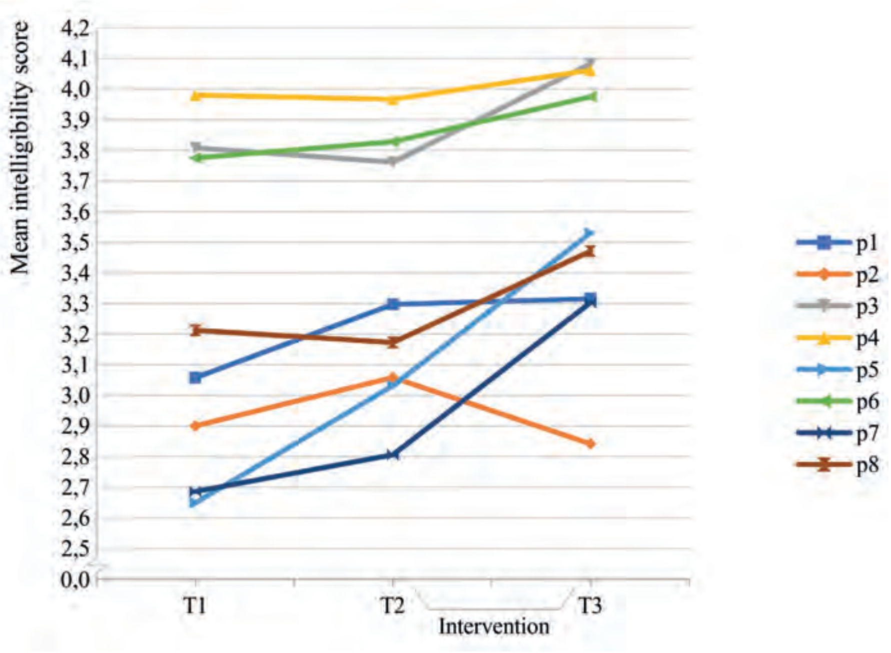


Figure 7.4.: *Estimated marginal means of intelligibility scores per speaker per time point. Note that ratings are on a scale from 0 to 5 and the figure is zoomed in on the Y-axis.*

Participant	Interface	Ease	Attractiveness	Overall
p1	8	8	8	9
p2	6	5	8	7
p3	7	7	7	7
p4	7	7	8	7
p5	9	9	8	9
p6	7	7	6	6
p7	9	9	9	7
p8	7	6	8	6
Avg. per dimension	7.50	7.25	7.75	7.25

Table 7.4.: *User satisfaction ratings per participant. Ratings on four dimensions: satisfaction with the Interface, Ease of use, game's Attractiveness, and Overall rating of the gaming experience. A 10-point-scale was used, in which 6 is considered 'sufficient'.*

7.6.2. User satisfaction

The participants' user satisfaction ratings for the game are displayed in Table 7.4. None of the participants rated the game as insufficient.

All participants but for one provided comments on their overall experience playing the game. They stated that they liked the game and assumed that it helped train their speech.

7.7. User appreciation of feedback

All participants except one stated that they frequently noticed the feedback throughout the game. For the question on the quality of the feedback, the answers were mixed. All but one participant answered positively to the question whether they could use the feedback and attempted to speak louder and lower.

Feedback on pronunciation was described as a positive experience by most participants as well. Five participants always agreed with the given feedback. Two sometimes disagreed and thought that the system should have detected a mispronunciation when it did not. One participant sometimes disagreed and was puzzled by why an initial pronunciation was not correct and a second was.

7.8. User preference

Table 7.5 presents the participants' ratings obtained in the preference task in which they were asked to choose between game-based and face-to-face speech training in four hypothetical situations (positive scores indicate a preference for game-based speech training). The available four comparisons per participant provided patterns in the criteria used in the preferences: preference for either game or face-to-face, possibly 'modulated' by the hypothetical outcome of the therapy. When the preferences are converted into binary scores, we see that four participants always prefer gaming, unless the outcome of the face-to-face therapy is strongly positive and that of the gaming only slightly positive; for two participants it is the other way round: face-to-face, unless the outcome of the face-to-face therapy is slightly positive and that of the gaming strongly positive, whereas two other participants show no consistent patterns in their preferences. These patterns are only based on observations with eight participants. Nevertheless, it makes clear which patterns can be expected, on the basis of which criteria: gaming vs. face-to-face, and the modulation of these preferences by the quality of the outcomes.

7.9. Discussion

Since there is an urgent need for independent, prolonged speech training in neurological patients' home environment, this study explored the potentials of game-based speech training for dysarthric speakers with Parkinson's disease.

Participant	Preference per scenario				Average preference game vs. face-to-face
	G++ vs. F++	G+- vs. F+-	G++ vs. F+-	G+- vs. F++	
p1	2	2	3	-2	1.25
p2	-3	-3	3	-3	-1.50
p3	1	1	2	-2	0.50
p4	3	3	3	-2	1.75
p5	-3	-2	2	-3	-1.50
p6	3	-2	3	-3	0.25
p7	3	3	3	-3	1.50
p8	2	-2	2	-3	-0.25

Table 7.5.: *Preference ratings per participant, comparing game-based speech training (G) with face-to-face speech training with a therapist (F) The hypothetical outcomes of the therapy are marked as strong (++) or slight (+-) improvement. Ratings range from -3 to 3, with positive values indicating a preference for game-based speech training over face-to-face speech training with a therapist.*

Pre-post measurements indicated a statistically significant improved speech intelligibility after the intervention (T2-T3), whereas no significant changes in speech intelligibility were observed over a 4-weeks period of no training prior to the intervention (T1-T2).

Moreover, the dysarthric speakers graded their satisfaction regarding the interface, the ease of use, the attractiveness of the game and their satisfaction overall with 'Treasure Hunters 2' as a game for speech training on average with 7.37 on a 10-point scale (no outliers). Although these preliminary results are promising, we should note that despite the statistically significant improvement, the clinical differences are relatively small. A contributing factor may have been the limited duration of our treatment. Our four weeks of treatment is based on intensive treatments in previous research as described in the Materials and Methods section. One could reason that additional weeks of treatment may improve speech intelligibility even more.

Our preliminary results do not show a consistent preference for game-based speech or traditional face-to-face speech training. Regardless of preferred type of training, users' decisions on the preferred scenario seem to be affected by the hypothesized effects on speech intelligibility. That is, our results indicate that the decisive factor seems to be the beneficial effects on speech, rather than the type of speech training. Similar results were observed in previous research (Beijer, 2012, Ch. 8; chapter 8). This also indicates no direct preference for traditional face-to-face training. Consequently, dysarthric speakers of the age group included in our study appear to positively receive game-based types of speech training as an alternative. The trade-off between type and effect of training calls for research into determinants for effective game-based training. From this perspective, our exploration of dysarthric speakers' appreciation of the type of feedback on loudness, pitch and pronunciation is relevant. That is, we included several types of feedback and explored how dysarthric speakers used and interpreted them. Overall, the dysarthric speakers in our study noticed the feedback on loudness and pitch when it was given, agreed with it most of the time and were able to adjust their speech as a result.

Generally, the types of feedback that were used seem to be well received in our study, especially the combination of implicit and explicit feedback for loudness. That is in line with Bakker et al. (2018) considerations for effective feedback. Reflecting on their work, we believe that a combination of implicit and explicit feedback can reinforce each other such that the explicit feedback makes the implicit feedback easier to understand.

In our exploration of feedback on pronunciation we employed ASR technology. As also pointed out by Bakker et al. (2018), it is important to consider the feasibility of providing reliable speech feedback. Our tests with artificially generated data made us confident that it would. However, we also know that artificially generated data does not reflect real data in every way. In our study, most participants had a positive experience and did not disagree with the provided feedback. Two participants described that the game sometimes fails to recognize mispronunciations. In ASR terms, the game sometimes falsely accepts mispronunciations. False rejects were not reported by the participants. In language learning literature false rejects are usually regarded as more detrimental to the process of language learning than false accepts (Van Doremalen et al., 2013). We assume that the same applies to speech training.

The duration of our speech training (i.e. four weeks) might be considered limited and perhaps contributed to an incomplete view of the effects of game-based training. Furthermore, no statements can be made about long-term effects of our treatment.

From a technological point of view we note that the calibration of the thresholds used for sentence and word verification was based on artificially generated data.

Future research should entail the calibration of these thresholds on real data. The data that was collected during this study can be used for that purpose. Furthermore, future studies should include a larger number of participants to enable more robust conclusions about the effects of game-based speech training. Also, the long-term effects of our treatment should be the subject of future study. Previous research showed that patients with dysarthria due to stroke also benefit from intensified and prolonged training (Wenke et al., 2008). For that reason, stroke patients should be included in future efficacy studies into serious games for speech training.

7.10. Conclusions

In this study the effects of an improved game-based speech intervention were compared to a period without intervention. We observed no significant differences between the pre and post measurements of the period without intervention, but did find a significant effect of our intervention. This finding supports game-based speech training as a positive modality to improve the overall speech intelligibility of individuals with dysarthria due to Parkinson's disease (PD).

In general, the dysarthric patients are satisfied with game-based speech training. Except for one outlier, all ratings were 6 or higher, and the total average is 7.25 on a 1-10 scale.

All but one participant had noticed the feedback on loudness, pitch, and pronunciation, and they were positive about it. They also felt they could use the feedback provided by the system to improve the quality of their speech.

Regarding their preference for game-based vs. face-to-face training, users do not prefer game-based speech training above face-to-face training. Most users have a preference for the type of speech training which provides the most hypothetical improvement.

To summarize, patients are positive about game-based speech therapy with instantaneous feedback, under the condition that a significant improvement is assured. However, these preliminary results should be interpreted with care. As in previous studies, differences between patients were observed. We can thus conclude that the game should be carefully tailored to each patient. Furthermore, for an optimal effect and user satisfaction, it should preferably not be used in isolation but in combination with face-to-face training. This way, the strengths of both types of training can be used to improve speech intelligibility in patients with PD.

Appendix 7A User satisfaction questionnaire

Note: the version below is a translation from the Dutch version used in our research.

User satisfaction Treasure Hunters

The following questions are about your satisfaction with Treasure Hunters, the game on the iPad you trained with in the last couple of weeks.

Answer the questions by clicking on one of the round buttons underneath. The scale goes from 1 to 10, 1 being very unsatisfied and 10 very satisfied.

Your participant number

Your test leader can enter this for you.

____

How satisfied are you with Treasure Hunters' user interface?

With this we mean the layout of the screen, use of the buttons, and possible help texts.

	1	2	3	4	5	6	7	8	9	10	
Extremely unsatisfied											Extremely satisfied

How satisfied are you with Treasure Hunters' ease of use?

Think about how easy it was for you to use the game for training your speech.

	1	2	3	4	5	6	7	8	9	10	
Extremely unsatisfied											Extremely satisfied

How satisfied are you about Treasure Hunters' attractiveness?

Think about how the game's visuals appealed to you.

	1	2	3	4	5	6	7	8	9	10	
Extremely unsatisfied											Extremely satisfied

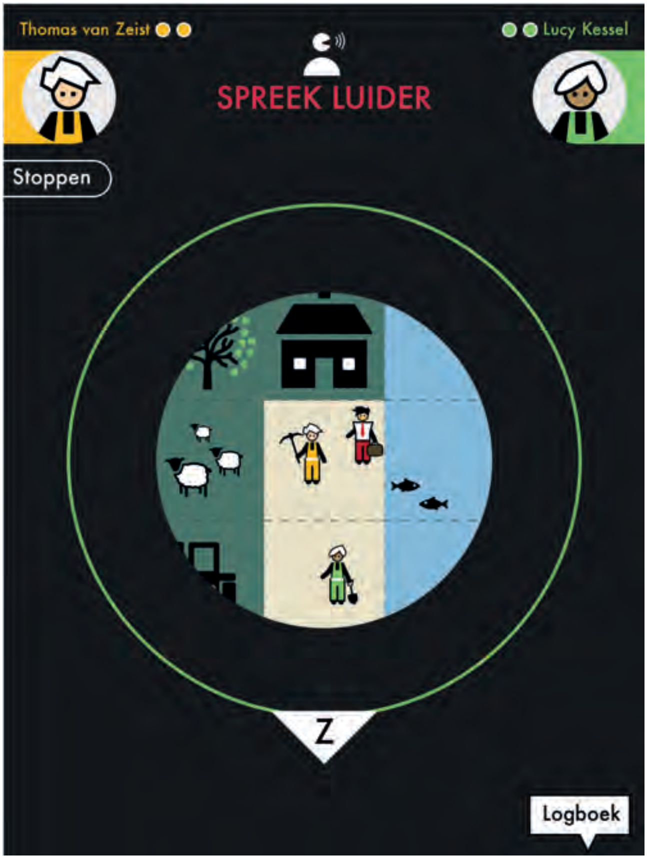
How satisfied are you about Treasure Hunters' attractiveness?

Think about how the game's visuals appealed to you.

	1	2	3	4	5	6	7	8	9	10	
Extremely unsatisfied											Extremely satisfied

How satisfied are you in general about Treasure Hunters as a game for speech training?

	1	2	3	4	5	6	7	8	9	10	
Extremely unsatisfied											Extremely satisfied



Did you notice the feedback on the pitch and loudness often?

With this we mean the resizing circle and the notifications ‘Speak louder’ and ‘Speak lower’ from the above screenshot.

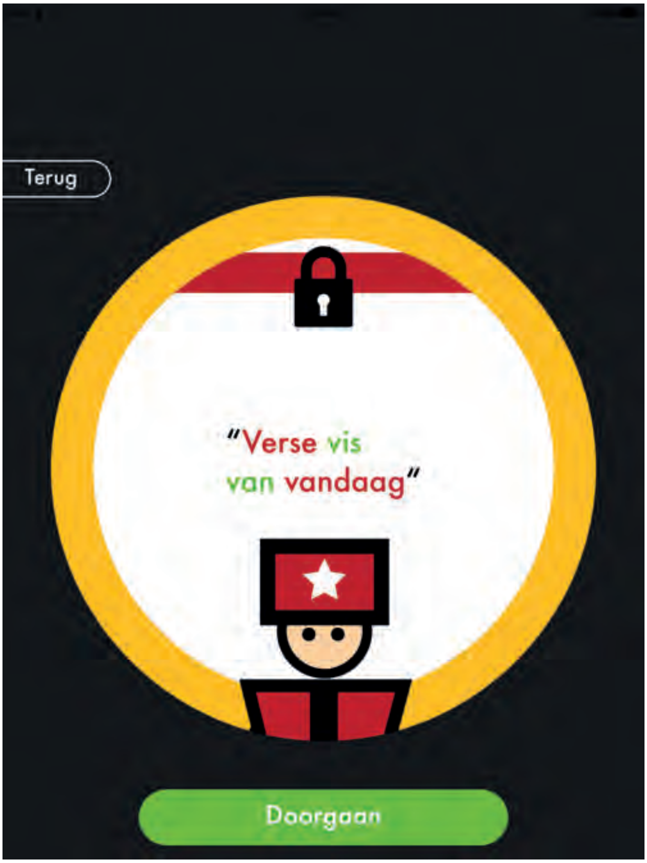
____

How did you experience the precision of the feedback?

____

What were you able to do with the feedback?

____



How did you experience the feedback on pronunciation?

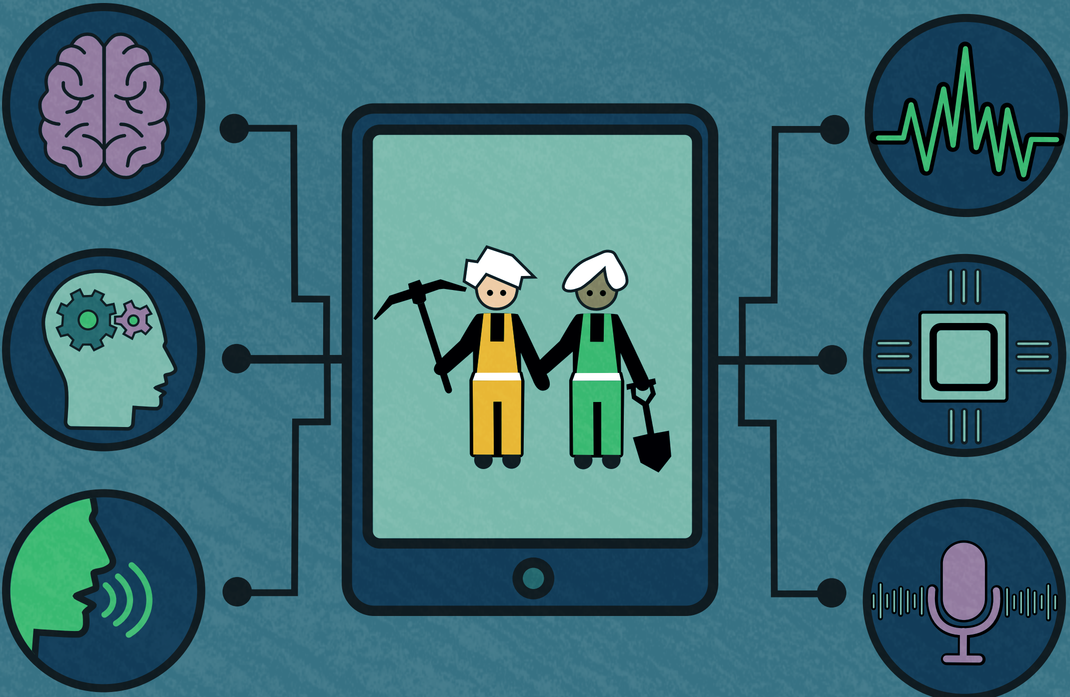
With this we mean the red/green colouring in the pronounced key phrases.

To what extent did you disagree with this feedback?

For example, words that were marked red but pronounced correctly according to you.

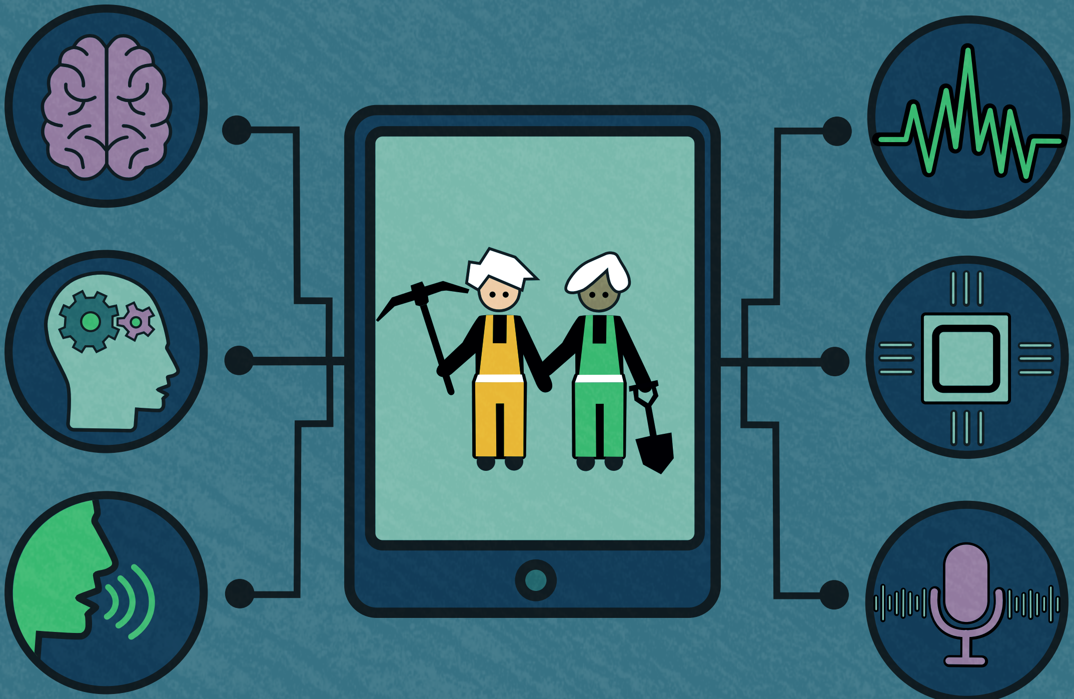
How did you generally experience playing Treasure Hunters?

Think of positive and negative aspects.



Part IV.

Discussion



Chapter 8

General Discussion

As populations around the world are ageing, healthcare in general faces a challenge in being able to provide the right amount of care to patients. Specifically, speech therapists face a growing group of patients with dysarthria due to acquired neurological disorders and are challenged in providing the necessary intensive therapy. Telerehabilitation has been one of the trends in speech rehabilitation research that has attempted to find solutions to this challenge. One of the first generation solutions enables one-on-one speech training sessions with a speech therapist remotely via a video link. This already removes travel to a rehabilitation centre or a patient's home resulting in more availability and flexibility in planning these sessions. A second generation solution is that of web-based courseware systems that provide automatic feedback on exercises that can be performed individually at home and monitored by a speech therapist remotely. Such systems can assist the speech therapist in providing a program that often contain 'drill-and-practice' exercises incorporating the needed training intensity. Additionally, these systems allow for remote changes or the introduction of new exercises to the program as patients' therapy progresses.

The current thesis is the result of the research project CHAllenging Speech training In Neurological patients using interactive Gaming (CHASING). As a partner in the CHASING project, the St. Maartenskliniek rehabilitation centre[1] provided their expertise on the topics of designing speech training programs, efficacy measurement, and clinical expertise on speech interventions. Waag Society's Creative Care Lab [2], the second partner in the project provided their expertise on user-centered design and game concept design. With these partners, the current thesis researched a third generation solution that explored the use of serious gaming (also called applied gaming) and Automatic Speech Recognition (ASR) based pronunciation evaluation for speech training. The current thesis includes research on the design of speech exercises and game concepts to enable training that is less drill-and-practice in nature. Also, this solution enables exercises that are more ecologically valid reducing the potential gap between drill-and-practice exercises and functional daily communication as reported in previous research (Beijer, 2012). Additionally, research on improving dysarthric ASR and its utilisation in providing automatic feedback on patients' speech is also addressed. The experiments that were conducted to explore the potential of the designed speech interventions are typically phase II clinical trials. In the general discussion below, the results obtained in the current thesis' research are reflected upon and the research questions from section 1.7, briefly described below, are evaluated:

1. Can speech intelligibility measures be obtained that provide evaluations at multiple levels of detail for the outcomes of different types of therapy?
2. What are ways to effectively improve the automatic recognition of dysarthric

[1] The Sint Maartenskliniek rehabilitation centre, Ubbergen, The Netherlands, http://www.maartenskliniek.nl, last accessed on October 28, 2024.

[2] Waag Society's Creative Care Lab, Amsterdam, The Netherlands, https://www.waag.org/nl/project/chasing, last accessed on October 28, 2024.

speech due to Parkinson's disease, stroke, and traumatic brain injury?

3. How can a serious game that provides automatic feedback on speech be designed for a speech intervention at dysarthric speakers' home environment?
4. How does game-based speech training compare to non-game computer-based speech training with respect to speech intelligibility outcomes and patient satisfaction?
5. Can a game-based speech training positively affect speech intelligibility in patients with dysarthria?
6. Do patients with dysarthria prefer game-based speech training to traditional face-to-face training?

In addition to discussing the research questions, limitations and directions for future research are also discussed in each section of this chapter, followed by concluding remarks. Each section of this chapter discusses one of the main research topics in the current thesis. That discussion is started in section 8.1 with the current thesis' research on speech intelligibility measures. The second section, section 8.2, evaluates Automatic Speech Recognition (ASR) technology in speech rehabilitation, how ASR can be improved for recognizing dysarthric speech and pronunciation evaluation. Subsequently, section 8.3 discusses the current thesis findings on designing a serious game for speech training. The fourth section, section 8.4, reflects upon the findings of the current thesis' research on the efficacy of the two serious game-based speech interventions. Next, section 8.5, discusses the results obtained on users' satisfaction and preferences for game-based speech training in comparison to face-to-face and non-game computer-based training. This chapter concludes with a brief summary of the current thesis' findings and implications in section 8.6.

8.1. Speech intelligibility measures

Measures of speech intelligibility are necessary to detect changes in intelligibility during speech rehabilitation. The research reported in chapters 2, 6, and 7 have contributed on this research topic and have shown that different contexts of measuring intelligibility call for different measures. Objective measures, like the loudness of the produced speech (measured as the intensity in decibels) and pitch (in Hertz) are suitable for automatic evaluation of intelligibility during gaming. Subjective measures, like human listener judgements on comprehensibility and listening effort are suited for before and after a period of speech training. That period of speech training can represent a speech intervention, in which the speech intelligibility measures can be used to measure the intervention's efficacy. In chapter 6, a subjective measure based on listeners' paired comparison judgements of recorded speech samples from before and after the intervention was used to explore the efficacy of a game-based versus non-game computer-based intervention. From these judgements, an intelligibility gain score was calculated for every pair. In chapter 7, a similar procedure was used to explore the efficacy of a game-based speech intervention versus a period of no intervention. However, judges were now asked to compare the dysarthric speech sample to that of a neurologically healthy speaker and rate on a 7-point scale to what extent the dysarthric speech samples were less intelligible, where equally intelligible and more intelligible were also options. The subjective measure in chapter 6 gave inconsistent results and the one in chapter 7 showed a statistically significant improvement in speech intelligibility. Reviewing these results, it is likely that this difference in results is partially due to that it is complex for inexperienced listeners to observe differences in intelligibility between two speech samples recorded from before and after an intervention. Especially when speakers are still fairly intelligible due to mild or moderate dysarthria. As a result, any differences between speech samples may be small and perhaps undetectable for the inexperienced listener. Differences between dysarthric and neurologically healthy speech are larger to begin with and for that reason perhaps easier to detect. However, how every listener reaches the given judgement remains a black box. For example, any differences in regional dialects or accents between two compared speech samples, albeit controlled for, may still play a role in listeners' judgements as well as differences in background noise. On the other hand, having two speech samples of which one is the anchor, will assist in transferring the knowledge to the listener on what is considered intelligible and less intelligible speech. Both chapter 6 and chapter 7 show that such scale-based measures can detect intelligibility changes in short utterances caused by a speech intervention. From this follows a nuance to the claim that scale-based measures are more suitable for longer utterances (section 1.4): scale-based measures based on paired comparisons of short utterances are suitable for measuring changes in speech intelligibility.

In chapter 2 of the current thesis, research into transcription-based measures of speech intelligibility is reported. As it is often more common to calculate these measures on the word level, this chapter specifically investigated whether more fine-grained measures could be obtained. Measures at word, and subword levels were studied and correlated to subjective utterance-level ratings. Arguably, subjective utterance-level ratings can inform on speakers' comprehensibility, word-level transcription-based measures on the intelligibility of words, and similar subword-level measures on the intelligibility of parts of words. From a clinical perspective, subword-level measures are relevant to provide more insight into speakers' specific articulation problems. They form a valuable addition to utterance-level subjective ratings and word-level transcription-based measures to inform on different dimensions of intelligibility.

The research reported in chapter 2 answers the first research question of the current thesis (see section 1.7). Speech intelligibility measures can be obtained that provide evaluations at multiple levels of detail for the outcomes of different types of therapy by using word- and subword-level transcription-based measures. A limitation of that research is that the word- and subword-level transcription-based measures have been investigated in the context of a single intelligibility assessment and not as part of a speech intervention. However, given that the main difference is that a speech intervention often has multiple intelligibility assessments over time, the measures are believed to be able to inform on changes in speech intelligibility. Future research should confirm that by statistically analysing the outcomes of a speech intervention using these measures. Recently, researchers have shown a growing interest in measuring Perceived Listening Effort (PLE) and its relation to speech intelligibility (Van der Bruggen et al., 2023). Such a measure can potentially be useful in measuring the effects of speech interventions. A downward trend of PLE measured over time may be beneficial to speech intelligibility. Future research should investigate PLE in speech interventions. A paired comparisons measure, as used in chapter 6 and chapter 7, may provide better anchored judgements instead of the often chosen visual analogue scale.

While the research reported in chapter 2 has been conducted in 2015-2016, the results and insights gained remain relevant in choosing measures for speech intervention efficacy research. The insights also remain relevant in advancing the research into intelligibility measures (Xue, Van Hout, Boogmans, et al., 2021; Xue, Van Hout, Cucchiarini, & Strik, 2021; Xue et al., 2023).

8.2. ASR technology in speech rehabilitation

ASR technology can be used in speech rehabilitation in different contexts. Two of such contexts are automatic measures for speech intelligibility assessment and automatic feedback on pronunciation. The current thesis focused on the latter as one of the goals was to explore the possibilities of such feedback to dysarthric speakers while playing a serious game. Important aspects of providing automatic feedback effectively in a gaming context are its precision and the timely presentation. Imprecise feedback should be mitigated as much as possible to not impede intervention efficacy and prevent frustrating players. Timely presentation of feedback is important to not impair the flow of a game and prevent players from getting impatient. Consequently, this places demands on the used ASR technology and models to be trained with respect to precision and computational efficiency. Dysarthric speech training corpora are generally small (see subsection 1.5.1). Additionally, here feedback should be provided on Dutch dysarthric speech. As Dutch is a low-resource language, it is additionally challenging to improve model precision. Chapters 3 and 4 researched ways to improve the recognition of dysarthric speech without having a large amount of data available. In chapter 3, a method was described that can moderately improve the recognition performance of the dysarthric speech by including neurologically healthy speech of Flemish, the southern variety of the Dutch language, in the training process. More importantly, retraining the models on that language variety increased the recognition improvement of the Flemish dysarthric speech. One would expect to observe similar or even greater improvements when the model that is trained on neurologically healthy speech of both language varieties is retrained and tested on Dutch dysarthric speech. However, chapter 4 did not show this improvement. An explanation can be that the models in that research were trained on speech from the northern and southern variety, but retrained on dysarthric speech, whereas in research reported in chapter 3, models were retrained on neurologically healthy speech. It is highly likely that training a background model on neurologically healthy speech of multiple language varieties and retraining them on dysarthric speech of one of these varieties causes a larger mismatch with that same dysarthric speech than when training a background model on neurologically healthy speech of only the language variety equal to that of the target dysarthric speech. Research using this method was continued by retraining the background models with target dysarthric speech or dysarthric speech combined with elderly speech, as elderly speech was believed to be relevant because the majority of dysarthric speakers are 50 years or older. Only retraining with dysarthric speech showed improvement, the set combined with elderly speech did not. This may suggest that elderly speech has different characteristics that do not match well with (elderly) dysarthric speech. To conclude this paragraph, increasing the size of the data set for training background models with neurologically healthy speech of language varieties does not necessarily improve recognition performance on dysarthric speech. Training on neurologically healthy speech of language varieties and retraining on the variety equal to that of the target dysarthric speech does improve the recognition performance. As such, future experiments could be run to investigate whether adding dysarthric speech to retraining with the language variety improves recognition performance further. Additionally, adding target dysarthric speech to both background model training and retraining also may improve recognition performance even more.

Since the research reported in chapter 3 and chapter 4, other data augmentation methods have been explored to increase the size of the training set. As described in subsection 1.5.1, traditional methods involving speech, tempo, and vocal tract length perturbations of the original speech signal have also been found effective to increase dysarthric speech recognition performance (Geng et al., 2020; Vachhani et al., 2018). These methods increase the variation and the quantity of dysarthric speech in the training set increasing the models' capacity for generalisation. Other data transformation methods tailored to dysarthric speech have also proven effective (Soleymanpour et al., 2021). While these transformations are a valuable addition, to what degree the resulting speech contributes to the way speech recognition models can cope with the high variability in dysarthric speech remains to be studied. Also, given that dysarthric speech corpora are commonly small and state-of-the-art models require large amounts of speech data, it remains to be studied to what extent the increase in the size of the training set due to the transformations improves recognition performance.

Other promising methods use deep learning technology to either transform neurologically healthy speech to dysarthric speech (Bhat et al., 2022; Jin et al., 2021) or train a multi-speaker text-to-speech system that can synthesize dysarthric speech

(Soleymanpour et al., 2022). In theory, these methods can generate a near infinite amount of dysarthric speech for model training. However, it is not entirely clear to what extent the generated speech represents the large variation in dysarthric speech. In retrospect, data augmentation techniques are one of the methods to improve the automatic recognition of dysarthric speech as highlighted in this discussion. As described in subsection 1.5.1, also other methods for improvement exist.

While the studies reported in chapters 3 and 4 were conducted in 2016 and 2017, their results remain relevant and contribute to the general research question on which speech data is suitable for training models that improve dysarthric speech recognition. The described methodologies can be used in future research that investigates the suitability of other speech data to be included for training.

After improving recognition performance, the resulting models were used in the first stage of mispronunciation detection that verified whether an attempt was made to actually speak the target utterance. This was part of improving the precision of automatic feedback such that it would not provide feedback in the second stage when it was not confident about the target utterance. Word-level confidence measures were calculated in the second stage that, evaluated against a predetermined threshold, resulted in the detection of either a correct word pronunciation or a mispronunciation. The decision for providing feedback on word level was made after discussions including clinical experts. It was believed that this would be cognitively less demanding to process for dysarthric speakers than more detailed feedback at subword level. The participants in the experiment described in chapter 7 indicated that they found it useful for adjusting their speech. Despite the limited number of participants, this indicates that word-level feedback is not too demanding to process. On the other hand, word-level feedback is less detailed than feedback at subword level. For example, when providing feedback at phoneme level, dysarthric speakers get to know which particular phonemes need to improve. Other methods like the Goodness of Pronunciation (GoP) described in subsection 1.5.2 can be used to provide such feedback. Perhaps future research can use these methods to investigate whether more detailed feedback is useful and not too demanding for dysarthric speakers to process. Research on automatic pronunciation evaluation for elderly dysarthric speech similar to that in the current thesis is limited. However, in the area of speech therapy for young children similar research exists and has also used feedback on word-level pronunciation (Hair et al., 2021; McLeod et al., 2023). When the recorded pronunciation was automatically judged to be correct, this was textually and visually displayed. As the speech training entailed single-word pronunciation exercises instead of the multi-word utterances from the current thesis, the researchers made use of template matching ASR technology to provide automatic pronunciation evaluation (Hair et al., 2021; McLeod et al., 2023). The advantage of template matching technology is that it requires less data for training. On the other hand, it is not designed for providing more detailed feedback. Interestingly, the same researchers found that the GoP algorithm, which is suited for more detailed feedback, performed almost similarly to the template matching system when trained on a large set of adult speech.

In the current thesis, the calculated word-level confidence measures were evaluated against a predetermined threshold. This threshold was fixed for all dysarthric speakers that participated in chapter 7's experiment as no speech recordings were available for fine-tuning. For this reason, the threshold was also configured on the safe side to limit the number of false rejects (i.e. detecting a mispronunciation when there was none). Future research to improve feedback on pronunciation should entail researching methods for personalizing thresholds. This way, feedback can be tailored to every dysarthric speaker's personal requirements. More detailed feedback at phoneme-level may also be included in future research. Providing such feedback while also playing a serious game may be too demanding, but an overview providing statistics on phoneme pronunciation after having completed a certain number of exercises could be useful.

Since the ASR research in the current thesis, new developments in the recognition of dysarthric speech have provided more efficient methods to adapt models for improving dysarthric speech recognition. For example, the availability of large mono- and cross-lingual acoustic models that are already pre-trained on tens of thousands (XLSR-53, Conneau et al., 2021) or even hundreds of thousands of hours (Whisper, Radford et al., 2022) of neurologically healthy speech potentially provide an efficient starting point for adaptation. They also showed better generalisation to different dysarthric severity groups (Wang & Van Hamme, 2023), which actually confirms our conclusion at the end of the first paragraph of this section: adding neurologically healthy speech of related language varieties to the train set improves dysarthric speech recognition performance. Also, Wang and Van Hamme (2023) show that even cross-lingual train sets of sufficient size can improve the robustness of models. Additionally, the large mono- and cross-lingual acoustic models enable robust self-supervised training/adaptation methods which saves quite some time manually transcribing speech recordings (Wang & Van Hamme, 2023). On the other hand, while training and adaptation may have been more efficient with these pre-trained models, they would have been more complex to integrate into a serious game as they would require more computational power in comparison to the Hybrid Deep-learning Neural Net approach used in the current thesis.

While the studies reported in chapters 6 and 7 have been conducted between 2016 and 2019, the research remains relevant as they have explored the efficacy of combining serious game design and automatic pronunciation evaluation to provide motivating speech training for elderly at home. Similar studies have only been found in the area of speech therapy for young children (Hair et al., 2021; McLeod et al., 2023). These are more recent studies and have used methodologies similar to those in the current thesis, which shows the relevancy of this thesis' research. Additionally, future research for elderly dysarthric speakers can use the current thesis' research results and methodologies to continue investigating automatic pronunciation evaluation and serious gaming for speech training.

To conclude this section, research question two of the current thesis on ways to effectively improve automatic recognition of dysarthric speech can be answered as follows. One of these ways is to improve the low resource conditions for training / adapting speech recognition models. In subsection 1.5.1 multiple methods have been described that improve these conditions by using language varieties, generating variants of existing dysarthric speech data, using out-of-domain data or generating speech data artificially. Consequently, these data augmentation methods improve the automatic recognition of dysarthric speech. Chapters 3 and 4 both contribute to the improvement of dysarthric speech recognition using out-of-domain data. Methods for improving mispronunciation detection using personalized confidence thresholds that include feedback at the more detailed phoneme level have also been discussed.

8.3. Serious game design for speech rehabilitation

When research in the CHASING project started, literature on designing serious games for speech training in older adults with dysarthria was limited. Since then, only a few publications have appeared on this topic. In related topics like the design of exergames for older adults and game design for this user group in general, research is often conducted by a multidisciplinary team of researchers. Most of these research efforts include the input of future users by following a user-centered design (UCD) methodology. CoDesign is one of such methodologies. The CoDesign approach taken in chapter 5 of the current thesis is a joint, creative process that supports multiple stakeholders in reaching their common goals. It can be considered a UCD methodology as it has many of the distinct characteristics: user profiling, iterative design process, and the inclusion of input from users in every design phase. What makes the CoDesign approach unique is the inclusion of knowledge from the creative industry. As a partner in the CHASING project, Waag Society provided knowledge of methods for identifying game principles users prefer. Subsequently, how to design and develop game concepts using these principles to test at an early stage what motivates users to play. In the case of the current thesis, this knowledge was also used to test how users can be triggered to speak while playing a game and to investigate physical and cognitive limitations. To conclude this paragraph, having a structured design methodology in place that includes a multidisciplinary team proves to be beneficial in designing a serious game for speech rehabilitation in dysarthric speakers.

The current thesis' serious game concept was designed as a two-player cooperative game. In this game, players needed to work together to find clues to reach a common goal like finding treasure or solving a mystery. Choosing a cooperative mode of play and having to work together towards a common goal was the result of user interviews and concept testing. This type of game may motivate a large group of older adults with dysarthria to train their speech. However, game concept preferences can differ per user on multiple dimensions. For example, due to personal interests, age, gender, gaming experience, physical and cognitive capabilities, cultural background, etc. Future research should entail studies of game concept preferences across larger user groups over multiple dimensions. This would provide a valuable starting point for game designers to develop additional serious games for speech rehabilitation. More serious games will not only cater the multitude of game concept preferences users may have, it will also help with the challenge of providing sufficient and varying content for training. The game in the current thesis was designed for an intervention of 16 training sessions over four weeks. Given that dysarthria due to progressive neurological disorders may further decrease speech intelligibility over time, serious games that enable lifelong training may prove effective in improving speech intelligibility or at the very least maintaining current levels. To research the effects of serious games over such a long period of time requires users to be motivated to play the game regularly. Providing sufficient suitable content will be an interesting challenge. A possible solution is to have multiple serious games available and let users play one after the other. However, this may complicate designs for efficacy experiments due to the potential effects of order, differences in motivation to play a game, differences in completion of games, etc. Perhaps, many of these effects can be controlled, but having only a single game would immediately remove these confounding variables. Another potential solution to obtain sufficient content is to develop a game with a level designer that can easily be used by a community around the game to design additional levels over time. It may also be possible to find a game concept that users find motivating to play and is suitable for automatic content generation. For example, the concept developed in the current thesis to find treasure together uses predefined maps and clues. When developing a map generator and a database of clues, such maps can be generated automatically providing a large amount of unique maps. On the other hand, such a generator alone would probably not be enough to keep users motivated as it may become quite monotonous to always play the same game only on different maps.

Designing feedback on speech in a gaming context also has its challenges. Users should be able to interpret and act on it. Oppositely, the feedback should not hinder users in playing the game. The current thesis' research found that providing indirect, near real-time feedback on loudness (measured as intensity in decibels) is a promising way of providing visual feedback. Such indirect feedback should change a user's view on the game world. Also, to reinforce the indirect feedback, a form of direct feedback was added using textual notifications on the height of dysarthric speakers' loudness and pitch levels. An earlier version of both direct and indirect feedback was difficult to process by dysarthric speakers. The feedback was visualized just above the user's view on the game world. Therefore, users had to quickly switch their focus back and forth between the feedback visualisation and view on the game world. That switching was found to be too difficult to cognitively process by dysarthric speakers and they stopped using the feedback. In addition to feedback on loudness and pitch, word-level feedback was also provided on pronunciation. The current thesis chose to provide the feedback visually by colouring the individual words that may contain a deviation. Other research found that additional auditory feedback is also beneficial to improving speech intelligibility. In the current thesis it was decided to not provide additional auditory feedback because it could hinder the tempo and flow of the game. However, future research should study how auditory feedback can be integrated in serious games for speech training minimizing those drawbacks.

In the area of speech therapy for young children with speech sound disorders, recent studies (Hair et al., 2021; McLeod et al., 2023) also included the use of a serious game similar to the ones in the current thesis. A Super Mario-like, single-player 2-dimensional platform game was designed that included single-word speech exercises and provided word-level automatic feedback on pronunciation. Those exercises were made part of the gameplay in that a speech exercise popup was triggered when the player touched a required star using the chosen game character. Collection of these stars was mandatory and the game briefly paused to perform the exercise in the popup. A similar concept was used in the current thesis' game design to unlock gates and bring players together. Also similar to the design in the current thesis, the popup disappeared when the number of pronunciation attempts was reached, also when all were incorrect. This way, the player is not frustrated by having to keep trying for a correct pronunciation before being allowed to continue. The popup provides direct feedback on pronunciation. Indirect feedback is also used by the game to motivate the player to keep playing and doing speech exercises. That is implemented by the concept of the game character's energy decreasing slowly over time. When it runs out it can only move slowly, making it more challenging to play the game. Energy levels can be increased by doing speech exercises in which correct attempts are rewarded with more energy then incorrect attempts. By collecting in-game coins, players can also buy items to personalize and customize their character and give them "superpowers". Applying gameplay elements like the ones previously described to the current thesis' game design may also be motivating for older adults. More research is needed to obtain additional insight on this topic.

Recent advances in generative AI models may provide additional opportunities for exploring new game concepts. Generative AI models are machine learning models that: "...are capable of generating seemingly new, meaningful content such as text, images, or audio from training data." (Feuerriegel et al., 2024, p. 111). Large Language Models (LLMs) are a specific application of generative AI that have been developed for specific tasks involving natural language generation and understanding. Currently, LLMs are often used in developing virtual assistants that can answer comprehensive questions by means of natural conversation (Casheekar et al., 2024). Such an assistant may enable new game concepts like a virtual coach that encourages and rewards the player for practising and correct pronunciation. For example, a more interactive version of the newsreader concept in section 5.4 or even a virtual speech therapist that engages in natural conversation and provides feedback on the pronunciation of the user. In addition to new game concepts, LLMs and other generative AI models may be able to assist in automatically generating game content to prolong speech training and keep players motivated. Generation of speech exercises given a set of rules and constraints to which the exercises must comply can also be a possibility. That would make the generation of such content more efficient and less time-consuming. It would also make the process of tailoring speech exercises to individual dysarthric speakers easier, which is potentially beneficial to the effects of the speech training. In summary, the potential benefits of generative AI and specifically LLM application in the design of a serious game for speech training warrants future research.

The research reported in chapter 5 was conducted between 2014-2018, but is still relevant. As that chapter has shown, research on serious game design for elderly in general exists. However, previous design research was not found for the unique combination of speech training for elderly with dysarthria due to a neurological disorder using serious games. This unique combination provides additional design challenges that are not covered by the literature on serious games design for neurologically healthy elderly. Recently, research has been reported on serious game design for speech rehabilitation independent of speech impairment and age group (Baranyi et al., 2024; Weber, 2025).

To answer research question three of the current thesis (see section 1.7), a serious game for speech rehabilitation in dysarthric speakers can be designed by following a CoDesign approach and involving disciplines from speech rehabilitation, creative design, automatic speech recognition, and software development. Users and their input are also included at every phase of the CoDesign approach. Providing feedback on loudness and pitch in a serious game is possible by using a combination of indirect and direct feedback which are visualized as part of the user's view on the game world. Feedback on pronunciation is possible by colouring individual words of the spoken utterance when a mispronunciation is detected.

8.4. Efficacy of game-based speech training

Studying the effects of training with the game on dysarthric speakers' intelligibility provides its own challenges. For example, how to measure intelligibility, what training program should dysarthric speakers follow, and which experimental design should be used to obtain validated results. Chapters 6 and 7 of the current thesis describe studies into the effects of the developed game when used in a speech training program. In chapter 6 the game-based speech training was compared to a non-game computer-based training from previous research. Ideally, such a comparison may show equal or even larger improvements in intelligibility for the game-based speech training. However, the outcomes of this study on the intelligibility of the dysarthric speakers were inconsistent and showed substantial variability. The number of participants that completed the two training programs was limited to five. Having such a limited and also heterogeneous group of participants may at least partially explain why no significant improvements were observed. On the positive side, no inexplicable decreases in intelligibility were observed either.

Because of the inconsistent results on intelligibility and the limited number of participants that could be included, the design of the second efficacy study (chapter 7) was simplified. In that study the game-based speech training was compared to a period of no training. This way, the study researched whether a game-based speech training could have any significant effect on intelligibility. Additionally, dysarthric speakers were incentivized to join the study as they would have to spend practically half the time in comparison to the previous study. Still, the dysarthric speakers that could be included in the study remained limited to eight. The results of the study showed a positive significant effect of the game-based speech training, which leads to the conclusion that game-based speech training can positively affect speech intelligibility in patients with dysarthria. Identifying reasons why this second study did find a significant effect and the first one did not is challenging. The following reasons may have contributed: simpler experiment design testing only one type of speech training, more homogeneous participant group (only dysarthria due to Parkinson's Disease, PD), improved version of the game that included pronunciation exercises, less technical, and device control issues.

In recent research on the efficacy of serious games in speech therapy for young children, the same experiment design was used as the research reported in chapter 6 (Hair et al., 2021; McLeod et al., 2023). Two speech interventions, with either manual or automatic provision of feedback, were compared in a crossover repeated measures design. Similar to the studies in chapter 6 and chapter 7, Hair et al. (2021) experienced difficulties in including large numbers of participants. They included 11 children of which 10 completed both speech interventions. Both treatments caused significant improvements in participants' pronunciations and no significant differences between the effects of the treatments were found. Although a similar experiment design was used in this thesis' chapter 6, that chapter compared two different methods of speech interventions, both including automatic feedback. In Hair et al. (2021) the comparison was made at a different level. The study included the same speech intervention, but compared one with manual provision of feedback by caregivers and one with automatic feedback using ASR technology. Besides some differences, the similarity in the research of these more recent studies with that reported in chapter 6 already points to the relevancy of that chapter's research. Additionally, exploring the efficacy of a new speech intervention that has not been tested before is relevant research on the road to offer that intervention to patients in the future. Besides the efficacy studies for speech therapy in young children, no other literature was found studying the efficacy of a serious game providing automatic pronunciation evaluation for dysarthric speakers.

To summarize, research on the comparison of game-based with non-game based speech training, research question four (see section 1.7), provided inconsistent results due to a limited and heterogeneous group of participants. Future research including additional participants may find more consistent results. Apart from the previously mentioned comparison, the current thesis found that game-based speech training can positively affect speech intelligibility in patients with dysarthria. Consequently, research question five of the current thesis (see section 1.7) can be answered positively. The positive effects of the game-based training were observed in older adult male and female individuals with dysarthria due to PD. The number of participants to the studies remains limited. For that reason, future research should confirm these findings on a larger sample of the dysarthric speakers due to PD. Additional research should include older adults with dysarthria from other aetiology's to observe whether their speech intelligibility can also be positively affected.

8.5. User satisfaction and preference

In addition to the effects on speech intelligibility, the game-based speech training was also rated on user satisfaction. In both studies (section 6.3 and section 7.6), the same user satisfaction questionnaire was used with ratings on four dimensions. Observing the average of the four dimensions individually, the attractiveness and overall satisfaction dimensions were scored lower for the first version of the game. This is most likely due to the technical issues some participants experienced during the first study and another participant not being used to working with a tablet computer. On average user satisfaction with the game-based speech training was similar for both versions of the game and more than sufficient. However, the number of participants in both studies is limited which prevents robust conclusions.

Users' preferences for training were also compared, game-based versus non-game computer-based training (section 6.3) and game-based versus face-to-face training (section 7.6). In this comparison, different scenarios were used in which the hypothesized effects on speech intelligibility were modulated. Both studies show no consistent preference for either game-based speech training, non-game computer-based speech training or face-to-face training. The results indicate that the hypothesized effects seem to be the decisive factor rather than the type of speech training. Most participants preferred the type of speech training that resulted in the largest hypothesized effects, which is in line with previous research (Beijer, 2012, ch. 8).

While the research into user satisfaction and preference have been conducted between 2016 and 2018, it remains relevant as it showed what elderly with dysarthria due to a neurological disorder like and dislike about game-based speech training. Those results can be used to further improve game-based speech training and increase motivation for training. That same research also gave insight into what method of training elderly dysarthric speakers prefer under different conditions. These insights can assist in deciding when to use which method of training and for whom, moving towards a more personalized speech intervention.

In retrospect, user satisfaction ratings for game-based and non-game computer-based speech training were similar. No consistent preference for either game-based, non-game computer-based or face-to-face training was observed. To answer research question six (see section 1.7), game-based speech training is not preferred over face-to-face training in all scenarios. Users seem to prefer the type of training that results in the largest improvement of speech intelligibility. Despite the limited number of participants in both studies, dysarthric speakers do seem to positively receive game-based types of speech training as an alternative.

8.6. Conclusions

Game-based speech training has potential to provide intensive training that has beneficial effects on speech intelligibility of dysarthric speakers. As the game is played from the home environment, this saves travel time making it easier to fit multiple training sessions into dysarthric speakers' schedules. Consequently, this allows for intensification of the speech training, which has shown to be beneficial to improving intelligibility. Similar to web-based courseware systems, game-based systems can reduce speech therapists' burden to provide face-to-face training sessions too, allowing them additional focus on adjusting the training program to individual needs. As a result, speech therapists may be able to provide training to more dysarthric individuals simultaneously, reducing the healthcare capacity problem. This is also interesting from a financial perspective, making more efficient use of healthcare resources, improving the cost-effectiveness ratio. In addition to courseware systems, game-based systems provide challenging games to motivate dysarthric individuals to train intensively. Consequently, these challenging games' strong motivational power may also contribute to increase therapy adherence, a well-known issue in speech rehabilitation where dysarthric speakers become less motivated to train at home. Another advantage of game-based systems is that they can provide more interactive and ecologically valid training scenarios making it easier to apply what was trained into functional daily communication.

As the current thesis has made clear, providing reliable automatic feedback is of great importance when training at home. The current thesis researched automatic feedback using speech analysis and automatic speech recognition algorithms. It has shown that speech analysis algorithms can be used to provide automatic feedback on loudness and pitch using threshold levels adequately. Methods to improve automatic dysarthric speech recognition can be found in using out-of-domain data, adding language varieties, noisified copies of the original recordings, and artificially generating speech data. Using one of these methods, the current thesis shows that providing pronunciation feedback on word-level in a near real-time gaming context is technically feasible. Additionally, dysarthric speakers are able to interpret that feedback and adjust their pronunciation while playing the game. To limit dysarthric speakers' frustration and perception of unreliable feedback, it is important to calibrate the pronunciation feedback's ratio of false positives versus false rejects. That is because detecting a mispronunciation where there is none, a false reject, potentially frustrates the dysarthric speaker more than failing to detect a mispronunciation, a false accept.

The current thesis has also confirmed that a "one size fits all" approach to game-based speech training does not apply. Individual dysarthric speakers or at least groups of dysarthric speakers have individual preferences towards games that they would like to play, feedback thresholds that need to be personalized and pronunciation feedback that needs to be calibrated per individual. Consequently, doing a game-based speech training should always be supervised by a speech therapist such that the right speech exercises are included and thresholds are recalibrated when a dysarthric individual's speech changes over time.

From a clinical perspective, the current thesis has explored the potentials of game-based speech training as an alternative to face-to-face and web-based courseware systems. Generally, the game-based system was positively received by dysarthric speakers. Given its advantages over the already existing face-to-face sessions, videoconferencing and courseware systems, game-based systems are in the current thesis' context not regarded as an alternative but as an addition. For example, face-to-face and videoconferencing systems can be used to explain exercises, control of systems, and subjective evaluation of speech. Courseware systems can be used for providing drill-and-practice exercises at home, and game-based systems for providing more ecologically valid scenarios and prolonged training. When considering the above statements, it is clear that the current thesis regards all approaches to be needed. This way, the strengths of all types of training can be used to improve or stabilize speech intelligibility in patients with dysarthria.

Bibliography

Abur, D., Enos, N. M., & Stepp, C. E. (2019). Visual analog scale ratings and orthographic transcription measures of sentence intelligibility in parkinson's disease with variable listener exposure. *American Journal of Speech-Language Pathology*, *28*(3), 1222–1232. https://doi.org/10.1044/2019_AJSLP-18-0275

Ackermann, H., & Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *54*(12), 1093–1098. https://doi.org/10.1136/jnnp.54.12.1093

Albiol-Pérez, S., Gil-Gómez, J.-A., Muńoz-Tomás, M.-T., Gil-Gómez, H., Vial-Escolano, R., & Lozano-Quilis, J.-A. (2017). The effect of balance training on postural control in patients with Parkinson's disease using a virtual rehabilitation system. *Methods of Information in Medicine*, *56*(02), 138–144. https://doi.org/10.3414/ME16-02-0004

Almadhor, A., Irfan, R., Gao, J., Saleem, N., Rauf, H. T., & Kadry, S. (2023). E2e-dasr: End-to-end deep learning-based dysarthric automatic speech recognition. *Expert Systems with Applications*, *222*, 119797. https://doi.org/10.1016/j.eswa.2023.119797

Al-Rayes, S., Ali Al Yaqoub, F., Alfayez, A., Alsalman, D., Alanezi, F., Alyousef, S., AlNujaidi, H., Al-Saif, A. K., Attar, R., Aljabri, D., Al-Mubarak, S., Al-Juwair, M. M., Alrawiai, S., Saraireh, L., Saadah, A., Al-umran, A., & Alanzi, T. M. (2022). Gaming elements, applications, and challenges of gamification in healthcare. *Informatics in Medicine Unlocked*, *31*, 100974. https://doi.org/10.1016/j.imu.2022.100974

Amengual Alcover, E., Jaume-i-Capó, A., & Moyà-Alcover, B. (2018). PROGame: A process framework for serious game development for motor rehabilitation therapy. *PLOS ONE*, *13*(5), 1–18. https://doi.org/10.1371/journal.pone.0197383

Baker, J. (1975). The dragon system–an overview. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*(1), 24–29. https://doi.org/10.1109/TASSP.1975.1162650

Bakker, M., Beijer, L., & Rietveld, T. (2018). Considerations on Effective Feedback in Computerized Speech Training for Dysarthric Speakers. *Telemedicine and e-Health*, *25*(5), 351–358. https://doi.org/10.1089/tmj.2018.0050

Balzan, P., Palmer, R., & Tattersall, C. (2023). Speech and language therapists' management practices, perceived effectiveness of current treatments and interest in neuromuscular electrical stimulation for acquired dysarthria rehabilitation: An international perspective. *International Journal of Language & Communication Disorders*, 1–18. https://doi.org/10.1111/1460-6984.12963

Balzan, P., Tattersall, C., & Palmer, R. (2022). Non-invasive brain stimulation for treating neurogenic dysarthria: A systematic review. *Annals of Physical and Rehabilitation Medicine*, *65*(5), 101580. https://doi.org/10.1016/j.rehab.2021.101580

Baranyi, R., Czech, P., Hofstätter, S., Aigner, C., & Grechenig, T. (2020). Analysis, Design, and Prototypical Implementation of a Serious Game Reha@Stroke to Support Rehabilitation of Stroke Patients With the Help of a Mobile Phone. *IEEE Transactions on Games*, *12*(4), 341–350. https://doi.org/10.1109/TG.2020.3017817

Baranyi, R., Weber, L., Aigner, C., Hohenegger, V., Winkler, S., & Grechenig, T. (2024). Voice-Controlled Serious Game: Design Insights for a Speech Therapy Application. *2024 E-Health and Bioengineering Conference (EHB)*, 1–4. https://doi.org/10.1109/EHB64556.2024.10805613

Barreto, S. d. S., & Ortiz, K. Z. (2008). Intelligibility measurements in speech disorders: A critical review of the literature. *Pró-Fono Revista de Atualização Científica*, *20*(3), 201–206.

Barry, G., Galna, B., & Rochester, L. (2014). The role of exergaming in Parkinson's disease rehabilitation: a systematic review of the evidence. *Journal of neuroengineering and rehabilitation*, *11*(1), 33. https://doi.org/10.1186/1743-0003-11-33

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beijer, L. J. (2012). *E-learning based Speech Therapy (EST): Exploring the potentials of e-health for dysarthric speakers* [Doctoral dissertation, Radboud University Nijmegen]. https://hdl.handle.net/2066/101662

Beijer, L. J., Clapham, R., & Rietveld, A. (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), 17–35.

Beijer, L. J., & Rietveld, A. C. M. (2011). Potentials of Telehealth Devices for Speech Therapy in Parkinson's Disease, Diagnostics and Rehabilitation of Parkinson's Disease. *InTech*, 379–402. https://doi.org/10.5772/17865

Beijer, L. J., Rietveld, A. C. M., Ruiter, M. B., & Geurts, A. C. H. (2014). Preparing an E-learning-based Speech Therapy (EST) efficacy study: Identifying suitable outcome measures to detect within-subject changes of speech intelligibility in dysarthric speakers. *Clinical Linguistics & Phonetics*, *28*(12), 927–950. https://doi.org/10.3109/02699206.2014.936627

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*(4), 381–392. https://doi.org/10.1016/0167-6393(96)00026-X

Berisha, V., Utianski, R., & Liss, J. (2013). Towards a clinical tool for automatic intelligibility assessment. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2825–2828. https://doi.org/10.1109/ICASSP.2013.6638172

Beristain-Colorado, M. D. P., Ambros-Antemate, J. F., Vargas-Treviño, M., Gutiérrez-Gutiérrez, J., Moreno-Rodriguez, A., Hernández-Cruz, P. A., Gallegos-Velasco, I. B., & Torres-Rosas, R. (2021). Standardizing the Development of Serious Games for Physical Rehabilitation: Conceptual Framework Proposal. *JMIR Serious Games*, *9*(2), e25854. https://doi.org/10.2196/25854

Beverly, D., Cannito, M. P., Chorna, L., Wolf, T., Suiter, D. M., & Bene, E. R. (2010). Influence of stimulus sentence characteristics on speech intelligibility scores in hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology*, *18*(4), 9–14.

Bharati, P., Chandra, S., & Das Mandal, S. K. (2023). Automatic Deep Neural Network-Based Segmental Pronunciation Error Detection of L2 English Speech (L1 Bengali). *Proc. Interspeech 2023*, 3068–3072. https://doi.org/10.21437/Interspeech.2023-1481

Bhat, C., Panda, A., & Strik, H. (2022). Improved asr performance for dysarthric speech using two-stage dataaugmentation. *Proc. Interspeech 2022*, 46–50. https://doi.org/10.21437/Interspeech.2022-10335

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of Aphasia Therapy, Impact on Recovery. *Stroke*, *34*(4), 987–993.

Blaney, B., & Wilson, J. (2000). Acoustic variability in dysarthria and computer speech recognition. *Clinical Linguistics & Phonetics*, *14*(4), 307–327. https://doi.org/10.1080/02699200050024001

Bogost, I. (2007, June). *Persuasive Games: The Expressive Power of Videogames*. The MIT Press. https://doi.org/10.7551/mitpress/5334.001.0001

Bunnell, H. T., Yarrington, D., & Polikoff, J. B. (2000). Star: Articulation training for young children. *Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000, Beijing, China, October 16-20, 2000*, 85–88. http://www.isca-speech.org/archive/icslp_2000/i00_4085.html

Bunton, K., Kent, R. D., Kent, J. F., & Duffy, J. R. (2001). The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics & Phonetics*, *15*(3), 181–193. https://doi.org/10.1080/02699200010003378

Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical Linguistics & Phonetics*, *14*(1), 13–24. https://doi.org/10.1080/026992000298922

Burke, J. W., McNeill, M., Charles, D., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, *25*(12), 1085–1099. https://doi.org/10.1007/s00371-009-0387-4

Burke, J., McNeill, M., Charles, D., Morrow, P., Crosbie, J., & McDonough, S. (2010). Augmented reality games for upper-limb stroke rehabilitation. *2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, 75–78. https://doi.org/10.1109/VS-GAMES.2010.21

Caballero-Morales, S.-O., & Cox, S. J. (2009). Modelling errors in automatic speech recognition for dysarthric speakers. *EURASIP J. Adv. Signal Process*, *2009*, 2:1–2:14. https://doi.org/10.1155/2009/308340

Caballero-Morales, S.-O., & Trujillo-Romero, F. (2014). Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition [Methods and Applications of Artificial and Computational Intelligence]. *Expert Systems with Applications*, *41*(3), 841–852. https://doi.org/10.1016/j.eswa.2013.08.014

Cancela, J., Pastorino, M., Arredondo, M. T., & Vera-Muñoz, C. (2014). State of the art on games for health focus on parkinson's disease rehabilitation. In Y.-T. Zhang (Ed.), *The international conference on health informatics* (pp. 13–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-03005-0_4

Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic parkinson's disease. *Journal of Voice*, *26*(2), 214–219. https://doi.org/10.1016/j.jvoice.2011.08.014

Casheekar, A., Lahiri, A., Rath, K., Sanjay Prabhakar, K., & Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. *Computer Science Review*, *52*, 100632. https://doi.org/10.1016/j.cosrev.2024.100632

Chen, F., & Kostov, A. (1997). Optimization of dysarthric speech recognition. *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, *4*, 1436–1439 vol.4. https://doi.org/10.1109/iembs.1997.756975

Chen, Y.-P. P., Johnson, C., Lalbakhsh, P., Caelli, T., Deng, G., Tay, D., Erickson, S., Broadbridge, P., El Refaie, A., Doube, W., et al. (2016). Systematic review of virtual speech therapists for speech disorders. *Computer Speech & Language*, *37*, 98–128.

Christensen, H., Aniol, M. B., Bell, P., Green, P. D., Hain, T., King, S., & Swietojanski, P. (2013). Combining in-domain and out-of-domain speech data for automatic recognition of disordered speech. *Proc. Interspeech 2013*, 3642–3645. https://doi.org/10.21437/Interspeech.2013-324

Christensen, H., Cunningham, S., Fox, C., Green, P., & Hain, T. (2012). A comparative study of adaptive, automatic recognition of disordered speech. *Proc. Interspeech 2012*, 1776–1779. https://doi.org/10.21437/Interspeech.2012-484

Cieri, C., Miller, D., & Walker, K. (2004). The fisher corpus: A resource for the next generations of speech-to-text. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04).*

Cikajlo, I., Hukić, A., Dolinšek, I., Zajc, D., Vesel, M., Krizmanič, T., Potisk, K. P., Blažica, B., Biasizzo, A., & Novak, F. (2017). Telerehabilitation of upper extremities with target based games for persons with Parkinson's disease. *2017 International Conference on Virtual Rehabilitation (ICVR)*, 1–2. https://doi.org/10.1109/ICVR.2017.8007466

Colado, I. J. P., Colado, V. M. P., Morata, A. C., Píriz, R. S. C., & Manjón, B. F. (2023). Using New AI-Driven Techniques to Ease Serious Games Authoring. *2023 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE58773.2023.10343021

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2021). Unsupervised cross-lingual representation learning for speech recognition. *Interspeech 2021*, 2426–2430. https://doi.org/10.21437/Interspeech.2021-329

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication*, *55*(4), 572–585. https://doi.org/10.1016/j.specom.2013.01.001

Creative Care Lab, Waag Society. (2014). CHASING project.

Cucchiarini, C., Driesen, J., Van Hamme, H., & Sanders, E. (2008). Recording Speech of Children, Non-Natives and Elderly People for HLT Applications: the JASMIN-CGN Corpus. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.), *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*. European Language Resources Association (ELRA).

Cucchiarini, C., & Strik, H. (2017). Automatic speech recognition for second language pronunciation training. In O. Kang, R. Thomson, & J. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 556–569). Routledge.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268. https://doi.org/10.1121/1.2946707

Dagenais, P. A., & Wilson, A. F. (2012). Acceptability and intelligibility of moderately dysarthric speech by four types of listeners. In *Investigations in clinical phonetics and linguistics* (pp. 379–388). Psychology Press.

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2011). Large vocabulary continuous speech recognition with context-dependent dbn-hmms. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4688–4691. https://doi.org/10.1109/ICASSP.2011.5947401

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Trans-*

actions on Audio, Speech, and Language Processing, *20*(1), 30–42. https://doi.org/10.1109/TASL.2011.2134090

Damaševičius, R., Maskeliūnas, R., & Blažauskas, T. (2023). Serious games and gamification in healthcare: A meta-review. *Information*, *14*(2). https://doi.org/10.3390/info14020105

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, *12*(3), 462–496. https://doi.org/10.1044/jshr.1203.462

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

De Bodt, M., Guns, C., & Van Nuffelen, G. (2006). *NSVO: Handleiding* (tech. rep.). Vlaamse Vereniging voor Logopedie: Herentals.

De Bodt, M., Hernandez-Diaz, H. M., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, *35*(3), 283–292.

De Cock, E., Batens, K., Hemelsoet, D., Boon, P., Oostra, K., & De Herdt, V. (2020). Dysphagia, dysarthria and aphasia following a first acute ischaemic stroke: Incidence and associated factors. *European Journal of Neurology*, *27*(10), 2014–2021. https://doi.org/10.1111/ene.14385

De Swart, B., Willemse, S., Maassen, B., & Horstink, M. (2003). Improvement of voicing in patients with parkinson's disease by speech therapy. *Neurology*, *60*(3), 498–500. https://doi.org/10.1212/01.WNL.0000044480.95458.56

Deller Jr., J. R., Hsu, D., & Ferrier, L. J. (1991). On the use of hidden markov modelling for recognition of dysarthric speech. *Computer Methods and Programs in Biomedicine*, *35*(2), 125–139. https://doi.org/10.1016/0169-2607(91)90071-z

de Oliveira, L. C., Mendes, L. C., de Lopes, R. A., Carneiro, J. A. S., Cardoso, A., Júnior, E. A. L., & de Oliveira Andrade, A. (2021). A systematic review of serious games used for rehabilitation of individuals with parkinson's disease. *Research on Biomedical Engineering*, *37*(4), 849–865. https://doi.org/10.1007/s42600-021-00162-3

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics*, *20*(9), 659–675. https://doi.org/10.1080/02699200500224272

Duffy, J. R. (2019, October). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th). Mosby.

Elffers, B., Van Bael, C., & Strik, H. (2005). ADAPT: Algorithm for dynamic alignment of phonetic transcriptions. *Radboud University, Nijmegen, The Netherlands, Tech. Rep.*

España-Bonet, C., & Fonollosa, J. A. R. (2016). Automatic speech recognition with deep neural networks for impaired speech. *Advances in Speech and Language Technologies for Iberian Languages*, 97–107. https://doi.org/10.1007/978-3-319-49169-1_10

Eyal, N. (2014). *Hooked*. Penguin Books Ltd.

Feenaughty, L., Tjaden, K., & Sussman, J. (2014). Relationship between acoustic measures and judgments of intelligibility in parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics*, *28*(11), 857–878. https://doi.org/10.3109/02699206.2014.921839

Feigin, V. L., Vos, T., Nichols, E., Owolabi, M. O., Carroll, W. M., Dichgans, M., Deuschl, G., Parmar, P., Brainin, M., & Murray, C. (2020). The global burden of neurological disorders: Translating evidence into policy. *The Lancet Neurology*, *19*(3), 255–265. https://doi.org/10.1016/S1474-4422(19)30411-9

Ferguson, C. J. (2007). The good, the bad and the ugly: A meta-analytic review of positive and negative effects of violent video games. *Psychiatric quarterly*, *78*(4), 309–316. https://doi.org/10.1007/s11126-007-9056-9

Ferguson, S., Johnston, A., Ballard, K., Tan, C. T., & Perera-Schulz, D. (2012). Visual feedback of acoustic data for speech therapy: Model and design parameters. *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound*, 135–140. https://doi.org/10.1145/2371456.2371478

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Finizia, C., Lindström, J., & Dotevall, H. (1998). Intelligibility and perceptual ratings after treatment for laryngeal cancer: Laryngectomy versus radiotherapy. *The Laryngoscope*, *108*(1), 138–143. https://doi.org/10.1097/00005537-199801000-00027

Frieg, H., Muehlhaus, J., Ritterfeld, U., & Bilda, K. (2017). ISi-Speech: A Digital Training System for Acquired Dysarthria. *242*, 330–334. https://doi.org/10.3233/978-1-61499-798-6-330

Fritsch, J., & Magimai-Doss, M. (2021). Utterance verification-based dysarthric speech intelligibility assessment using phonetic posterior features. *IEEE Signal Processing Letters*, *28*, 224–228. https://doi.org/10.1109/LSP.2021.3050362

Gamito, P., Oliveira, J., Coelho, C., Morais, D., Lopes, P., Pacheco, J., Brito, R., Soares, F., Santos, N., & Barata, A. F. (2017). Cognitive training on stroke patients via virtual reality-based serious games. *Disability and Rehabilitation*, *39*(4), 385–388. https://doi.org/10.3109/09638288.2014.934925

Gandhi, P., Tobin, S., Vongphakdi, M., Copley, A., & Watter, K. (2020). A scoping review of interventions for adults with dysarthria following traumatic brain injury. *Brain Injury*, *34*(4), 466–479. https://doi.org/10.1080/02699052.2020.1725844

Ganzeboom, M., Yılmaz, E., Cucchiarini, C., & Strik, H. (2016). On the development of an ASR-based multimedia game for speech therapy: Preliminary results. *Proceedings of the 2016 ACM workshop on multimedia for personal health and health care*, 3–8.

García, C., Nickolai, D., & Jones, L. (2020). Traditional versus asr-based pronunciation instruction: An empirical study. *CALICO Journal*, *37*(3), pp. 213–232.

Garcia-Agundez, A., Folkerts, A.-K., Konrad, R., Caseman, P., Göbel, S., & Kalbe, E. (2017). PDDanceCity: an exergame for patients with idiopathic Parkinson's disease and cognitive impairment. *Mensch und Computer 2017-Tagungsband*, 381–386. https://doi.org/10.18420/muc2017-mci-0334

Geng, M., Xie, X., Liu, S., Yu, J., Hu, S., Liu, X., & Meng, H. (2020). Investigation of data augmentation techniques for disordered speech recognition. *Proc. Interspeech 2020*, 696–700. https://doi.org/10.21437/Interspeech.2020-1161

Ghanouni, P., Jarus, T., Collette, D., & Pringle, R. (2017). Using virtual reality gaming platforms to improve balance in rehabilitation of stroke survivors. *2017 International Conference on Virtual Rehabilitation (ICVR)*, 1–2. https://doi.org/10.1109/ICVR.2017.8007465

Gibbon, D., Moore, R., & Winski, R. (Eds.). (1998). *Handbook of standards and resources for spoken language systems*. De Gruyter Mouton. https://doi.org/10.1515/9783110809817

Goršič, M., Darzi, A., & Novak, D. (2017). Comparison of two difficulty adaptation strategies for competitive arm rehabilitation exercises. *2017 International Conference on Rehabilitation Robotics (ICORR)*, 640–645. https://doi.org/10.1109/ICORR.2017.8009320

Green, P., Carmichael, J., Hatzis, A., Enderby, P., Hawley, M. S., & Parker, M. (2003). Automatic speech recognition with sparse training data for dysarthric speakers. *INTERSPEECH*. https://doi.org/10.21437/Eurospeech.2003-384

Haderlein, T., Nöth, E., Batliner, A., Eysholdt, U., & Rosanowski, F. (2011). Automatic intelligibility assessment of pathologic speech over the telephone. *Logopedics Phoniatrics Vocology*, *36*, 175–181. https://doi.org/10.3109/14015439.2011.607470

Hahm, S., Heitzman, D., & Wang, J. (2015). Recognizing dysarthric speech due to amyotrophic lateral sclerosis with across-speaker articulatory normalization. *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 47–54. https://doi.org/10.18653/v1/W15-5109

Hair, A., Ballard, K. J., Markoulli, C., Monroe, P., Mckechnie, J., Ahmed, B., & Gutierrez-Osuna, R. (2021). A longitudinal evaluation of tablet-based child speech therapy with apraxia world. *ACM Transactions on Accessible Computing*, *14*(1). https://doi.org/10.1145/3433607

Hajesmaeel-Gohari, S., Goharinejad, S., Shafiei, E., & Bahaadinbeigy, K. (2023). Digital games for rehabilitation of speech disorders: A scoping review. *Health science reports*, *6*(6). https://doi.org/10.1002/hsr2.1308

Halpern, A. E., Ramig, L. O., Matos, C. E. C., Petska-Cable, J. A., Spielman, J. L., Pogoda, J. M., Gilley, P. M., Sapir, S., Bennett, J. K., & McFarland, D. H. (2012). Innovative technology for the assisted delivery of intensive voice treatment (lsvt loud) for parkinson disease. *American Journal of Speech-Language*

Pathology, *21*(4), 354–367. https://doi.org/10.1044/1058-0360(2012/11-0125)

Halpern, B. M., Fritsch, J., Hermann, E., Van Son, R., Scharenborg, O., & Magimai-Doss, M. (2021). An objective evaluation framework for pathological speech synthesis. *Speech Communication; 14th ITG Conference*, 1–5.

Hardy, J. (1967). Suggestions for physiological research in dysarthria. *Cortex*, *3*(1), 128–156. https://doi.org/10.1016/S0010-9452(67)80009-1

Hawley, M. S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J., Parker, M., Hatzis, A., O'Neill, P., & Palmer, R. (2006). A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering & Physics*, *29*(5), 586–593. https://doi.org/10.1016/j.medengphy.2006.06.009

Hawley, M. S., Enderby, P., Green, P., Cunningham, S., & Palmer, R. (2006). Development of a voice-input voice-output communication aid (vivoca) for people with severe dysarthria. In K. Miesenberger, J. Klaus, W. Zagler, & A. Karshmer (Eds.), *Computers helping people with special needs* (pp. 882–885, Vol. 4061). Springer Berlin Heidelberg. https://doi.org/10.1007/11788713_128

Hawley, M. S., Green, P., Enderby, P., Cunningham, S., & Moore, R. K. (2005). Speech technology for e-inclusion of people with physical disabilities and disordered speech. *INTERSPEECH*. http://www.isca-speech.org/archive/interspeech_2005/i05_0445.html

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Neural Networks, 1989. IJCNN., International Joint Conference on*, 593–605 vol.1.

Hermann, E., & Magimai-Doss, M. (2020). Dysarthric speech recognition with lattice-free mmi. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6109–6113. https://doi.org/10.1109/ICASSP40776.2020.9053549

Hinton, G. E. (2010). *A Practical Guide to Training Restricted Boltzmann Machines* (tech. rep. No. UTML TR 2010003). Department of Computer Science, University of Toronto.

Hinton, G. E., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. https://doi.org/10.1109/MSP.2012.2205597

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Holmes, R. J., Oates, J. M., Phyland, D. J., & Hughes, A. J. (2000). Voice characteristics in the progression of parkinson's disease. *International Journal of Language & Communication Disorders*, *35*(3), 407–418.

Hu, S., Xie, X., Cui, M., Deng, J., Liu, S., Yu, J., Geng, M., Liu, X., & Meng, H. (2022). Neural architecture search for lf-mmi trained time delay neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 1093–1107. https://doi.org/10.1109/TASLP.2022.3153253

Huang, J.-T., Li, J., Yu, D., Deng, L., & Gong, Y. (2013). Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. *Proc. ICASSP*, 7304–7308. https://doi.org/10.1109/ICASSP.2013.6639081

Hustad, K. C. (2006). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica*, *58*(3), 217–228. https://doi.org/10.1159/000091735

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*(6), 306–317. https://doi.org/10.1159/000108337

Hustad, K. C. (2008). The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria. *Journal of Speech, Language, and Hearing Research*, *51*(3), 562–573. https://doi.org/10.1044/1092-4388(2008/040)

Hutchins, S. (1992). Say & see articulation therapy software. *Proceedings of the Johns Hopkins National Search for Computing Applications to Assist Persons with Disabilities*, 37–40. https://doi.org/10.1109/CAAPWD.1992.217399

Ishikawa, K., Boyce, S., Kelchner, L., Powell, M. G., Schieve, H., de Alarcon, A., & Khosla, S. (2017). The effect of background noise on intelligibility of dysphonic speech. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1919–1929. https://doi.org/10.1044/2017_JSLHR-S-16-0012

Ishikawa, K., Webster, J., & Ketring, C. (2021). Agreement between transcription- and rating-based intelligibility measurements for evaluation of dysphonic speech in noise. *Clinical Linguistics & Phonetics*, *35*(10), 983–995. https://doi.org/10.1080/02699206.2020.1852602

Jaddoh, A., Loizides, F., & Rana, O. (2023). Interaction between people with dysarthria and speech recognition systems: A review. *Assistive Technology*, *35*(4), 330–338. https://doi.org/10.1080/10400435.2022.2061085

Janbakhshi, P., Kodrasi, I., & Bourlard, H. (2019). Pathological speech intelligibility assessment based on the short-time objective intelligibility measure. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6405–6409. https://doi.org/10.1109/ICASSP.2019.8683741

Jayaram, G., & Abdelhamied, K. (1995). Experiments in dysarthric speech recognition using artificial neural networks. *Journal of rehabilitation research and development*, *32*, 162–162.

Jin, Z., Geng, M., Xie, X., Yu, J., Liu, S., Liu, X., & Meng, H. (2021). Adversarial data augmentation for disordered speech recognition. *Proc. Interspeech 2021*, 4803–4807. https://doi.org/10.21437/Interspeech.2021-168

Johansson, I.-L., Samuelsson, C., & Müller, N. (2023). Consonant articulation acoustics and intelligibility in swedish speakers with parkinson's disease: A pilot study. *Clinical Linguistics & Phonetics*, *37*(9), 845–865. https://doi.org/10.1080/02699206.2022.2095926

Joshy, A. A., & Rajan, R. (2022). Automated dysarthria severity classification: A study on acoustic features and deep learning techniques. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 1147–1157. https://doi.org/10.1109/TNSRE.2022.3169814

Joy, N. M., & Umesh, S. (2018). Improving acoustic models in torgo dysarthric speech database. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(3), 637–645. https://doi.org/10.1109/TNSRE.2018.2802914

Joy, N. M., Umesh, S., & Abraham, B. (2017). On improving acoustic models for torgo dysarthric speech database. *Proc. Interspeech 2017*, 2695–2699. https://doi.org/10.21437/Interspeech.2017-878

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd). Retrieved August 20, 2024, from https://web.stanford.edu/~jurafsky/slp3/

Kalf, J., de Swart, B., Bloem, B., & M., M. (2008). Guidelines for speech-language therapy in Parkinson's disease, Abstracts of The Movement Disorder Society's Twelfth International Congress of Parkinson's Disease and Movement Disorders. *Movement Disorders*, *23*(S1), S328. https://doi.org/10.1002/mds.22133

Kalf, J., de Swart, B., Bonnier, B., Hofman, M., Kanters, J., Kocken, J., Miltenburg, M., Bloem, B., & M., M. (2011). *Guidelines for speech-language therapy in Parkinson's disease*. Nijmegen, The Netherlands / Miami (FL), U.S.A.: ParkinsonNet/NPF. Retrieved September 17, 2025, from http://www.parkinsonnet.info/media/11927204/guidelines_for_speech-language_therapy_in_parkinson_s_disease.pdf

Kari, T. (2017). Promoting physical activity and fitness with exergames: Updated systematic review of systematic reviews. *Transforming gaming and computer simulation technologies across industries*, 225–245. https://doi.org/10.4018/978-1-5225-1817-4.ch013

Kempler, D., & Lancker, D. V. (2002). Effect of speech task on intelligibility in dysarthria: A case study of parkinson's disease. *Brain and Language*, *80*(3), 449–464. https://doi.org/10.1006/brln.2001.2602

Kent, R. D. (1992). *Intelligibility in speech disorders: Theory, measurement and management* (Vol. 1). John Benjamins Publishing.

Kent, R. D., & Kim, Y. J. (2003). Toward an acoustic typology of motor speech disorders*. *Clinical Linguistics & Phonetics*, *17*(6), 427–445. https://doi.org/10.1080/0269920031000086248

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, *54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482

Kent, R., & Kim, Y. J. (2011). The assessment of intelligibility in motor speech disorders. In *Assessment of motor speech disorders* (pp. 21–37). Plural San Diego, CA.

Kim, H., & Gurevich, N. (2023). Positional asymmetries in consonant production and intelligibility in dysarthric speech. *Clinical Linguistics & Phonetics*, *37*(2), 125–142. https://doi.org/10.1080/02699206.2021.2019312

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T. S., Watkin, K., & Frame, S. (2008). Dysarthric speech database for universal access research. *INTERSPEECH*, 1741–1744. http://www.isca-speech.org/archive/interspeech_2008/i08_1741.html

Kim, M. J., Kim, Y., & Kim, H. (2015). Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *23*(4), 694–704. https://doi.org/10.1109/TASLP.2015.2403619

Kim, M., Wang, J., & Kim, H. (2016). Dysarthric speech recognition using kullback-leibler divergence-based hidden markov model. *Proc. Interspeech 2016*, 2671–2675. https://doi.org/10.21437/Interspeech.2016-776

Koivisto, J., & Malik, A. (2022). Gamification for older adults: A systematic literature review. *The Gerontologist*, *61*(7), e360–e372. https://doi.org/10.1093/geront/gnaa047

Krause, M., Smeddinck, J., & Meyer, R. (2013). A Digital Game to Support Voice Treatment for Parkinson's Disease. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 445–450. https://doi.org/10.1145/2468356.2468435

Kwakkel, G. (2006). Impact of intensity of practice after stroke: issues for consideration. *Disability and Rehabilitation*, *28*((13-14)), 823–830.

Laamarti, F., Eid, M., & El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*. https://doi.org/10.1155/2014/358152

Laborde, V., Pellegrini, T., Fontan, L., Mauclair, J., Sahraoui, H., & Farinas, J. (2016). Pronunciation assessment of Japanese learners of French with GOP scores and phonetic information. *Annual conference Interspeech (INTERSPEECH 2016)*, 2686–2690. https://doi.org/10.21437/Interspeech.2016-513

Laine, T. H., Normark, J., Lindvall, H., Lindqvist, A.-K., & Rutberg, S. (2020). A distributed multiplayer game to promote active transport at workplaces: User-centered design, implementation, and lessons learned. *IEEE Transactions on Games*, *12*(4), 386–397. https://doi.org/10.1109/TG.2020.3021728

Laurent, M., Monnier, S., Huguenin, A., Monaco, P.-B., & Jaccard, D. (2022). Design Principles for Serious Games Authoring Tools. *International Journal of Serious Games*, *9*(4), 63–87. https://doi.org/10.17083/ijsg.v9i4.458

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*(5), 1148–1156. https://doi.org/10.1044/jslhr.4205.1148

Laver, K. E., Lange, B., George, S., Deutsch, J. E., Saposnik, G., & Crotty, M. (2018). Virtual Reality for Stroke Rehabilitation. *Stroke*, *49*(4), e160–e161. https://doi.org/10.1161/STROKEAHA.117.020275

Leblong, E., Fraudet, B., Dandois, M., Nicolas, B., & Gallien, P. (2017). A 4 weeks home training program using a biofeedback serious game and sensors for parkinson's disease: A pilot study on a new and completely autonomous solution. *2017 International Conference on Virtual Rehabilitation (ICVR)*, 1–2. https://doi.org/10.1109/ICVR.2017.8007460

Lee, S., Oh, H., Shi, C.-K., & Doh, Y. Y. (2021). Mobile Game Design Guide to Improve Gaming Experience for the Middle-Aged and Older Adult Population: User-Centered Design Approach. *JMIR Serious Games*, *9*(2), e24449. https://doi.org/10.2196/24449

Lee, T., Liu, Y., Huang, P.-W., Chien, J.-T., Lam, W. K., Yeung, Y. T., Law, T. K. T., Lee, K. Y., Kong, A. P.-H., & Law, S.-P. (2016). Automatic speech recognition for acoustical analysis and assessment of Cantonese pathological voice and speech. *Proceedings of ICASSP*, 6475–6479.

Lehiste, I., Tikofsky, R. S., & Tikofsky, R. P. (1961). An acoustic description of dysarthric speech. *The Journal of the Acoustical Society of America*, *33*(11), 1677–1677. https://doi.org/10.1121/1.1936729

Lehner, K., & Ziegler, W. (2021). The impact of lexical and articulatory factors in the automatic selection of test materials for a web-based assessment of intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2196–2212. https://doi.org/10.1044/2020_JSLHR-20-00267

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

Levis, J. M., & Silpachai, A. O. (2022). Speech intelligibility. In *The routledge handbook of second language acquisition and speaking* (pp. 160–173). Routledge.

Lewis, G. N., Woods, C., Rosie, J. A., & Mcpherson, K. M. (2011). Virtual reality games for rehabilitation of people with stroke: Perspectives from the users. *Disability and Rehabilitation: Assistive Technology*, *6*(5), 453–463.

Lewis, Z. H., Swartz, M. C., & Lyons, E. J. (2016). What's the point?: A review of reward systems implemented in gamification interventions. *Games for health journal*, *5*(2), 93–99.

Li, C.-J., Yeo, E., Choi, K., Pérez-Toro, P. A., Someki, M., Das, R. K., Yue, Z., Orozco-Arroyave, J. R., Nöth, E., & Mortensen, D. R. (2025). Towards Inclusive ASR: Investigating Voice Conversion for Dysarthric Speech Recognition in Low-Resource Languages. https://doi.org/10.48550/arXiv.2505.14874

Li, J. (2022). Recent advances in end-to-end automatic speech recognition. *APSIPA Transactions on Signal and Information Processing*, *11*(1), 1–64. https://doi.org/10.1561/116.00000050

LimeSurvey Project Team and C. Schmitz. (2015). *LimeSurvey: An open source survey tool*. LimeSurvey Project, Hamburg, Germany. http://www.limesurvey.org

Lin, Y., Wang, L., Dang, J., Li, S., & Ding, C. (2020). End-to-end articulatory modeling for dysarthric articulatory attribute detection. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7349–7353. https://doi.org/10.1109/ICASSP40776.2020.9054233

Lin, Y., Wang, L., Li, S., Dang, J., & Ding, C. (2020). Staged Knowledge Distillation for End-to-End Dysarthric Speech Recognition and Speech Attribute Transcription. *Proc. Interspeech 2020*, 4791–4795. https://doi.org/10.21437/Interspeech.2020-1755

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? gamification in health and fitness apps. *JMIR Serious Games*, *2*(2), e9. https://doi.org/10.2196/games.3413

Liu, S., Hu, S., Wang, Y., Yu, J., Su, R., Liu, X., & Meng, H. (2019). Exploiting visual features using bayesian gated neural networks for disordered speech recognition. *Proc. Interspeech 2019*, 4120–4124. https://doi.org/10.21437/Interspeech.2019-1536

López-Rodríguez, D., & García-Linares, A. (2013). Spare: Spatial rehabilitation with learning, recommendations and gamification. *ICERI2013 Proceedings*, 5923–5931.

Lounis, M., Dendani, B., & Bahi, H. (2024). Mispronunciation detection and diagnosis using deep neural networks: A systematic review. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17899-x

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of Motor Learning in Treatment of Motor Speech Disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Machado, M. d. C., Ferreira, R. L. R., & Ishitani, L. (2018). Heuristics and Recommendations for the Design of Mobile Serious Games for Older Adults. *International Journal of Computer Games Technology*, *2018*(1), 6757151. https://doi.org/10.1155/2018/6757151

Magnuson, T., & Blomberg, M. (2000). Acoustic analysis of dysarthric speech with some implications for automatic speech recognition. *Quarterly Progress and Status Report - Royal Institute of Technology, Department of Speech, Music and Hearing*, *41*(1), 19–30. http://www.speech.kth.se/prod/publications/files/qpsr/2000/2000_41_1_019-030.pdf

Mathew, J. B., Jacob, J., Sajeev, K., Joy, J., & Rajan, R. (2018). Significance of feature selection for acoustic modeling in dysarthric speech recognition. *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 1–4. https://doi.org/10.1109/WiSPNET.2018.8538531

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., O'Beirne, G. A., & Anderson, T. (2017). Effect of Dysarthria Type, Speaking Condition, and Listener Age on Speech Intelligibility. *American Journal of Speech-Language Pathology*, *26*(1), 113–123. https://doi.org/10.1044/2016_AJSLP-15-0182

McConnell, S. (1996). Evolutionary Prototyping. In *Rapid development: Taming wild software schedules* (1st). Microsoft Press.

McLeod, S., Kelly, G., Ahmed, B., & Ballard, K. J. (2023). Equitable access to speech practice for rural australian children using the saybananas! mobile game. *International Journal of Speech-Language Pathology*, *25*(3), 388–402. https://doi.org/10.1080/17549507.2023.2205057

McNaney, R., Othman, M., Richardson, D., Dunphy, P., Amaral, T., Miller, N., Stringer, H., Olivier, P., & Vines, J. (2016). Speeching: Mobile Crowdsourced Speech Assessment to Support Self-Monitoring and Management for People with Parkinson's. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4464–4476. https://doi.org/10.1145/2858036.2858321

Mendoza Ramos, V., Vasquez-Correa, J. C., Cremers, R., Van Den Steen, L., Nöth, E., de Bodt, M., & Van Nuffelen, G. (2021). Automatic boost articulation therapy in adults with dysarthria: Acceptability, usability and user interaction. *International Journal of Language & Communication Disorders*, *56*(5), 892–906. https://doi.org/10.1111/1460-6984.12647

Menéndez-Pidal, X., Polikoff, J. B., Peters, S. M., Leonzio, J. E., & Bunnell, H. T. (1996). The nemours database of dysarthric speech. *Proceedings on the 4th international conference on Spoken Language Processing*, *3*, 1962–1966. http://www.asel.udel.edu/speech/Research-dys.html

Mengistu, K. T., & Rudzicz, F. (2011). Adapting acoustic and lexical models to dysarthric speech. *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 4924–4927. https://doi.org/10.1109/icassp.2011.5947460

Middag, C. (2012). *Automatic analysis of pathological speech* [Doctoral dissertation, Ghent University]. Ghent University, Department of Electronics; information systems.

Middag, C., Bocklet, T., Martens, J.-P., & Nöth, E. (2011). Combining phonological and acoustic asr-free features for pathological speech intelligibility assessment. *INTERSPEECH*, 3005–3008. http://www.isca-speech.org/archive/interspeech_2011/i11_3005.html

Middag, C., Clapham, R., Van Son, R., & Martens, J.-P. (2014). Robust automatic intelligibility assessment techniques evaluated on speakers treated for head and neck cancer. *Computer Speech & Language*, *28*(2), 467–482. https://doi.org/10.1016/j.csl.2012.10.007

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, *2009*(1). https://doi.org/10.1155/2009/629030

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, *48*(6), 601–612. https://doi.org/10.1111/1460-6984.12061

Miyamoto, C., Komai, Y., Takiguchi, T., Ariki, Y., & Li, I. (2010). Multimodal speech recognition of a person with articulation disorders using aam and maf. *2010 IEEE International Workshop on Multimedia Signal Processing*, 517–520. https://doi.org/10.1109/MMSP.2010.5662075

Mohamed, A.-r., Dahl, G., & Hinton, G. E. (2009). Deep belief networks for phone recognition. *NIPS 22 workshop on deep learning for speech recognition*, 1–9. http://www.cs.utoronto.ca/~gdahl/papers/dbnPhoneRec.pdf

Moya-Galé, G., Goudarzi, A., Bayés, A., McAuliffe, M., Bulté, B., & Levy, E. S. (2018). The effects of intensive speech treatment on conversational intelligibility in spanish speakers with parkinson's disease. *American Journal of Speech-Language Pathology*, *27*(1), 154–165. https://doi.org/10.1044/2017_AJSLP-17-0032

Moya-Galé, G., & Levy, E. S. (2019). Parkinson's disease-associated dysarthria: Prevalence, impact and management strategies. *Research and Reviews in Parkinsonism*, *9*, 9–16. https://doi.org/10.2147/JPRLS.S168090

Mubin, O., Alnajjar, F., Al Mahmud, A., Jishtu, N., & Alsinglawi, B. (2022). Exploring serious games for stroke rehabilitation: A scoping review. *Disability and Rehabilitation: Assistive Technology*, *17*(2), 159–165. https://doi.org/10.1080/17483107.2020.1768309

Mühlhaus, J., Frieg, H., Bilda, K., & Ritterfeld, U. (2017). Game-based speech rehabilitation for people with parkinson's disease. *Universal Access in Human–Computer Interaction. Human and Technological Environments: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part III 11*, 76–85. https://doi.org/10.1007/978-3-319-58700-4_7

Mulfari, D., Celesti, A., & Villari, M. (2022). Exploring ai-based speaker dependent methods in dysarthric speech recognition. *2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 958–964. https://doi.org/10.1109/CCGrid54584.2022.00117

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Mustafa, M. B., Rosdi, F., Salim, S. S., & Mughal, M. U. (2015). Exploring the influence of general and specific factors on the recognition accuracy of an asr system for dysarthric speaker. *Expert Systems with Applications*, *42*(8), 3924–3932. https://doi.org/10.1016/j.eswa.2015.01.033

Muzakki Bashori, H. S., Roeland van Hout, & Cucchiarini, C. (2024). I Can Speak: improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *0*(0), 1–19. https://doi.org/10.1080/17501229.2024.2315101

Nakashika, T., Yoshioka, T., Takiguchi, T., Ariki, Y., Duffner, S., & Garcia, C. (2014). Dysarthric speech recognition using a convolutive bottleneck network. *2014 12th International Conference on Signal Processing (ICSP)*, 505–509. https://doi.org/10.1109/ICOSP.2014.7015056

Nasiri, N., Shirmohammadi, S., & Rashed, A. (2017). A serious game for children with speech disorders and hearing problems. *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–7. https://doi.org/10.1109/SeGAH.2017.7939296

Nøhr, C., & Aarts, J. (2010). Use of "serious health games" in health care: A review. *Information Technology in Health Care: Socio-Technical Approaches 2010: from safe Systems to Patient Safety*, *157*, 160–166.

NUFFIC. (2022). The Dutch organisation for internationalisation and education. Grading systems. [online]. https://www.nuffic.nl/en/education-systems/netherlands/grading-systems

Olmstead, A. J., Lee, J., & Viswanathan, N. (2020). The role of the speaker, the listener, and their joint contributions during communicative interactions: A tripartite view of intelligibility in individuals with dysarthria. *Journal of Speech, Language, and Hearing Research*, *63*(4), 1106–1114. https://doi.org/10.1044/2020_JSLHR-19-00233

Ong, D. S. M., Weibin, M. Z., & Vallabhajosyula, R. (2021). Serious games as rehabilitation tools in neurological conditions: A comprehensive review. *Technology and Health Care*, *29*(1), 15–31. https://doi.org/10.3233/THC-202333

Oostdijk, N. (2000). The spoken Dutch corpus. overview and first evaluation. In M. Gavrilidou, G. Carayannis, S. Markantonatou, S. Piperidis, & G. Stainhauer (Eds.), *Proceedings of the second international conference on language resources and evaluation (LREC'00)*. European Language Resources Association (ELRA). https://aclanthology.org/L00-1083/

Oostdijk, N., Goedertier, W., van Eynde, F., Boves, L., Martens, J.-P., Moortgat, M., & Baayen, H. (2002). Experiences from the spoken Dutch corpus project. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*.

Orozco-Arroyave, J. R., Vásquez-Correa, J. C., Klumpp, P., Pérez-Toro, P. A., Escobar-Grisales, D., Roth, N., Ríos-Urrego, C. D., Strauss, M., Carvajal-Castaño, H. A., Bayerl, S. P., Castrillón-Osorio, L. R., Arias-Vergara, T., Künderle, A., López-Pabón, F. O., Parra-Gallego, L. F., Eskofier, B., Gómez-Gómez, L. F., Schuster, M., & Nöth, E. (2020). Apkinson: The smartphone application for telemonitoring parkinson's patients through speech, gait and hands movement. *Neurodegenerative disease management*, *10*(3), 137–157. https://doi.org/10.2217/nmt-2019-0037

O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. *Computer Speech & Language*, *83*, 101538. https://doi.org/10.1016/j.csl.2023.101538

Pacheco-Velazquez, E., Rodes-Paragarino, V., Rabago-Mayer, M., & Bester, A. (2023). How to Create Serious Games? Proposal for a Participatory Methodology. *International Journal of Serious Games*, *10*(4), 55–73. https://doi.org/10.17083/ijsg.v10i4.642

Pachoulakis, I., & Papadopoulos, N. (2016). Exergames for Parkinson's Disease patients: The balloon goon game. *2016 International Conference on Telecommunications and Multimedia (TEMU)*, 1–6. https://doi.org/10.1109/TEMU.2016.7551908

Pakarinen, A., & Salanterä, S. (2020). The use of gaming in healthcare. In A. Charalambous (Ed.), *Developing and utilizing digital technology in healthcare for assessment and monitoring* (pp. 115–125). Springer International Publishing. https://doi.org/10.1007/978-3-030-60697-8_9

Palmer, R., Enderby, P., & Hawley, M. (2007). Addressing the needs of speakers with longstanding dysarthria: Computerized and traditional therapy compared. *International Journal of Language & Communication Disorders*, *42*(sup1), 61–79. https://doi.org/10.1080/13682820601173296

Patel, R., Usher, N., Kember, H., Russell, S., & Laures-Gore, J. (2014). The influence of speaker and listener variables on intelligibility of dysarthric speech [Motor Speech Disorders]. *Journal of Communication Disorders*, *51*, 13–18. https://doi.org/10.1016/j.jcomdis.2014.06.006

Pellegrini, T., Fontan, L., Mauclair, J., Farinas, J., Alazard-Guiu, C., Robert, M., & Gatignol, P. (2015). Automatic assessment of speech capability loss in disordered speech. *ACM Trans. Access. Comput.*, *6*(3). https://doi.org/10.1145/2739051

Pellegrini, T., Fontan, L., Mauclair, J., Farinas, J., & Robert, M. (2014). The goodness of pronunciation algorithm applied to disordered speech. *Proceedings of the 15th international conference of the International Speech Communication Association*, 1463–1467. https://doi.org/10.21437/Interspeech.2014-357

Pennington, L., & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clinical Linguistics & Phonetics*, *21*(5), 393–403. https://doi.org/10.1080/02699200701276675

Perero-Codosero, J. M., Espinoza-Cuadros, F. M., & Hernández-Gomez, L. A. (2022). A comparison of hybrid and end-to-end asr systems for the iberspeech-rtve 2020 speech-to-text transcription challenge. *Applied Sciences*, *12*(2). https://doi.org/10.3390/app12020903

Pietrowicz, M., & Karahalios, K. G. (2014). Visualizing vocal expression. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 1369–1374. https://doi.org/10.1145/2559206.2581331

Pophale, C., & Chavan, S. (2023). A comprehensive survey of automatic dysarthric speech recognition. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(9s), 24–30. https://doi.org/10.17762/ijritcc.v11i9s.7392

Popovici, D. V., & Buică-Belciu, C. (2012). Professional challenges in computer-assisted speech therapy. *Procedia - Social and Behavioral Sciences*, *33*, 518–522. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.01.175

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi speech recognition toolkit. *Proc. ASRU*.

Prabhavalkar, R., Hori, T., Sainath, T. N., Schlüter, R., & Watanabe, S. (2023). End-to-end speech recognition: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *32*, 325–351. https://doi.org/10.1109/TASLP.2023.3328283

Psychology Software Tools. (2012). E-PRIME 2 (Version 2.0.8.90) [Computer software].

Pyae, A., Joelsson, T., Saarenpää, T., Mika, L., Kattimeri, C., Pitkäkangas, P., Granholm, P., & Smed, J. (2017). Lessons Learned from Two Usability Studies of Digital Skiing Game with Elderly People in Finland and Japan. *International Journal of Serious Games*, *4*(4). https://doi.org/10.17083/ijsg.v4i4.183

Qian, Z., Xiao, K., & Yu, C. (2023). A survey of technologies for automatic dysarthric speech recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2023*(48). https://doi.org/10.1186/s13636-023-00318-2

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. https://arxiv.org/abs/2212.04356

Raghavendra, P., Rosengren, E., & Hunnicutt, S. (2001). An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems. *Augmentative and Alternative Communication*, *17*(4), 265–275. https://doi.org/10.1080/aac.17.4.265.275

Ramig, L., Sapir, S., Countryman, S., Pawlas, A., O'Brien, C., Hoehn, M., & Thompson, L. (2001). Intensive voice treatment (LSVT®) for patients with parkinson's disease: A 2 year follow up. *J. Neurol. Neurosur. Ps.*, *71*(4), 493–498. https://doi.org/10.1136/jnnp.71.4.493

Ramírez, E. R. R., Duncan, W., Brebner, S., & Chan, K. (2017). The Design Process and Usability Assessment of an Exergame System to Facilitate Strength for Task Training for Lower Limb Stroke Rehabilitation. In M. B. Alonso & E. Ozcan (Eds.), *Proceedings of the conference on design and semantics of form and movement - sense and sensitivity, desform 2017*. IntechOpen. https://doi.org/10.5772/intechopen.71119

Ratnanather, J. T., Bhattacharya, M. B., Rohitand Heston, Song, J., Fernandez, L. R., Lim, H. S., Lee, S.-W., Tam, E., Yoo, S., Bae, S.-H., Lam, I., Jeon, H. W., Chang, S. A., & Koo, J.-W. (2021). An mhealth app (speech banana) for auditory training: App design and development study. *JMIR Mhealth Uhealth*, *9*(3), e20890. https://doi.org/10.2196/20890

Reddy, D., Erman, L., & Neely, R. (1973). A model and a system for machine recognition of speech. *IEEE Transactions on Audio and Electroacoustics*, *21*(3), 229–238. https://doi.org/10.1109/TAU.1973.1162456

Rietveld, T. (2021). *Human Measurement Techniques in Speech and Language Pathology: Methods for Research and Clinical Practice* (1st). Routledge.

Rietveld, T., & Van Heuven, V. (2016). *Algemene fonetiek [general phonetics]* (4th). Coutinho.

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*, 275–280.

Ritterfeld, U., Muehlhaus, J., Frieg, H., & Bilda, K. (2016). Developing a technology-based speech intervention for acquired dysarthria: A psychological approach. *International Conference on Computers Helping People with Special Needs*, 93–100.

Rosen, M. J. (1999). Telerehabilitation. *NeuroRehabilitation*, *12*(1), 11–26. https://doi.org/10.3233/NRE-1999-12103

Rudzicz, F. (2007). Comparing speaker-dependent and speaker-adaptive acoustic models for recognizing dysarthric speech. *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 255–256. https://doi.org/10.1145/1296843.1296899

Rudzicz, F. (2011). *Production knowledge in the recognition of dysarthric speech* [Doctoral dissertation, University of Toronto] [AAINR77904]. University of Toronto. http://hdl.handle.net/1807/29854

Rudzicz, F., Namasivayam, A., & Wolff, T. (2012). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation*, *46*(4), 523–541. https://doi.org/10.1007/s10579-011-9145-0

Sainath, T. N., Mohamed, A.-r., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for lvcsr. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 8614–8618. https://doi.org/10.1109/ICASSP.2013.6639347

Sanders, E. B.-N., & Stappers, P. J. (2014). Probes, toolkits and prototypes: Three approaches to making in codesigning. *CoDesign*, *10*(1), 5–14. https://doi.org/10.1080/15710882.2014.888183

Sanders, E., Ruiter, M. B., Beijer, L. J., & Strik, H. (2002). Automatic recognition of dutch dysarthric speech: A pilot study. In J. H. L. Hansen & B. L. Pellom (Eds.), *Interspeech*. ISCA. https://doi.org/10.21437/ICSLP.2002-217

Sapir, S., Spielman, J. L., Ramig, L. O., Story, B. H., & 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*(4), 899–912. https://doi.org/10.1044/1092-4388(2007/064)

Schaefer, R. S., Beijer, L. J., Seuskens, W., Rietveld, T. C., & Sadakata, M. (2016). Intuitive visualizations of pitch and loudness in speech. *Psychonomic bulletin & review*, *23*(2), 548–555. https://doi.org/10.3758/s13423-015-0934-0

Scheffé, H. (1952). An analysis of variance for paired comparisons. *Journal of the American Statistical Organization*, *47*(259), 381–400. https://doi.org/10.1080/01621459.1952.10501179

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In *Intelligibility in speech disorders: Theory, measurement and management* (pp. 11–34). Amsterdam/Philadelphia. https://doi.org/10.1075/sspcl.1.02sch

Schuster, M., Maier, A., Haderlein, T., Nkenke, E., Wohlleben, U., Rosanowski, F., Eysholdt, U., & Nöth, E. (2006). Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition. *International Journal of Pediatric Otorhinolaryngology*, *70*(10), 1741–1747. https://doi.org/10.1016/j.ijporl.2006.05.016

Scott, A. M., Clark, J., Cardona, M., Atkins, T., Peiris, R., Greenwood, H., Wenke, R., Cardell, E., & Glasziou, P. (2024). Telehealth versus face-to-face delivery of speech language pathology services: A systematic review and meta-analysis. *Journal of Telemedicine and Telecare*. https://doi.org/10.1177/1357633X241272976

Seide, F., Li, G., & Yu, D. (2011). Conversational speech transcription using context-dependent deep neural networks. *Proc. Interspeech 2011*, 437–440. https://doi.org/10.21437/Interspeech.2011-169

Seong, W., Park, J., & Kim, H. (2012). Dysarthric speech recognition error correction using weighted finite state transducers based on context-dependent pronunciation variation. In *Computers helping people with special needs* (pp. 475–482, Vol. 7383). https://doi.org/10.1007/978-3-642-3

Shahamiri, S. R., Lal, V., & Shah, D. (2023). Dysarthric speech transformer: A sequence-to-sequence dysarthric speech recognition system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 3407–3416. https://doi.org/10.1109/TNSRE.2023.3307020

Shahamiri, S. R., & Salim, S. S. B. (2014). Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of mfcc parameters and studying a speaker-independent approach. *Advanced Engineering Informatics*, *28*(1), 102–110. https://doi.org/10.1016/j.aei.2014.01.001

Sharma, H. V., Hasegawa-Johnson, M., Gunderson, J., & Perlman, A. (2009). Universal access: Speech recognition for talkers with spastic dysarthria. *Proceedings of INTERSPEECH*, 1451–1454. https://doi.org/10.21437/Interspeech.2009-444

Shtern, M., Haworth, M. B., Yunusova, Y., Baljko, M., & Faloutsos, P. (2012). A game system for speech rehabilitation. In M. Kallmann & K. Bekris (Eds.), *Motion in games* (pp. 43–54). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34710-8_5

Sidi Yakoub, M., Selouani, S.-a., Zaidi, B.-F., & Bouchair, A. (2020). Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network. *EURASIP Journal on Audio, Speech, and Music Processing*, *2020*(1). https://doi.org/10.1186/s13636-019-0169-5

Smiljanic, R. (2021). Clear speech perception. In *The handbook of speech perception* (pp. 177–205). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119184096.ch7

Soleymanpour, M., Johnson, M. T., & Berry, J. (2021). Dysarthric Speech Augmentation Using Prosodic Transformation and Masking for Subword End-to-end ASR. *2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 42–46. https://doi.org/10.1109/SpeD53181.2021.9587372

Soleymanpour, M., Johnson, M. T., Soleymanpour, R., & Berry, J. (2022). Synthesizing Dysarthric Speech Using Multi-Speaker TTS For Dysarthric Speech Recognition. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7382–7386. https://doi.org/10.1109/ICASSP43922.2022.9746585

Soleymanpour, M., Johnson, M. T., Soleymanpour, R., & Berry, J. (2024). Accurate synthesis of dysarthric speech for asr data augmentation. *Speech Communication*, *164*, 103112. https://doi.org/10.1016/j.specom.2024.103112

Speaks, C., Parker, B., Harris, C., & Kuhl, P. (1972). Intelligibility of connected discourse. *Journal of Speech and Hearing Research*, *15*(3), 590–602. https://doi.org/10.1044/jshr.1503.590

Sriranjani, R., Umesh, S., & Reddy, M. R. (2015). Pronunciation adaptation for disordered speech recognition using state-specific vectors of phone-cluster adaptive training. *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 72–78. https://doi.org/10.18653/v1/W15-5113

Standen, P., Threapleton, K., Richardson, A., Connell, L., Brown, D., Battersby, S., Platts, F., & Burton, A. (2017). A low cost virtual reality system for home based rehabilitation of the arm following stroke: A randomised controlled feasibility trial. *Clinical Rehabilitation*, *31*(3), 340–350. https://doi.org/10.1177/0269215516640320

Stanfield, M. (2008). *Proceedings of the 2nd European Conference on Games Based Learning: ECGBL*. Academic Publishing International. https://books.google.nl/books?id=r6mPyaObITEC

Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, *59*(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271

Strik, H. (2014). Project CHASING: CHAllenging Speech training In Neurological patients by interactive Gaming. https://www.helmer-strik.nl/chasing/

Strik, H., & Cucchiarini, C. (1999). Modeling pronunciation variation for asr: A survey of the literature. *Speech Communication*, *29*(2-4), 225–246. https://doi.org/10.1016/S0167-6393(99)00038-2

Swietojanski, P., Ghoshal, A., & Renals, S. (2012). Unsupervised cross-lingual knowledge transfer in dnn-based lvcsr. *2012 IEEE Spoken Language Technology Workshop (SLT)*, 246–251. https://doi.org/10.1109/SLT.2012.6424230

Swinnen, B. E., Lotfalla, V., Scholten, M. N., Prins, R. H., Goes, K. M., de Vries, S., Geytenbeek, J. J., Dijk, J. M., Odekerken, V. J., Bot, M., van den Munckhof, P., Schuurman, P. R., de Bie, R. M., & Beudel, M. (2023). Programming algorithm for the management of speech impairment in subthalamic nucleus deep brain stimulation for parkinson's disease. *Neuromodulation: Technology at the Neural Interface*. https://doi.org/10.1016/j.neurom.2023.05.002

Tabak, M., de Vette, F., van Dijk, H., & Vollenbroek-Hutten, M. (2020). A game-based, physical activity coaching application for older adults: Design approach and user experience in daily life. *Games for Health Journal*, *9*(3), 215–226. https://doi.org/10.1089/g4h.2018.0163

Takashima, Y., Nakashika, T., Takiguchi, T., & Ariki, Y. (2015). Feature extraction using pre-trained convolutive bottleneck nets for dysarthric speech recognition. *2015 23rd European Signal Processing Conference (EUSIPCO)*, 1411–1415. https://doi.org/10.1109/EUSIPCO.2015.7362616

Takashima, Y., Takashima, R., Takiguchi, T., & Ariki, Y. (2019). Knowledge transferability between the speech data of persons with dysarthria speaking different languages for dysarthric speech recognition. *IEEE Access*, *7*, 164320–164326. https://doi.org/10.1109/ACCESS.2019.2951856

Takashima, Y., Takiguchi, T., & Ariki, Y. (2019). End-to-end dysarthric speech recognition using multiple databases. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6395–6399. https://doi.org/10.1109/ICASSP.2019.8683803

Tamayo-Serrano, P., Garbaya, S., & Blazevic, P. (2018). Gamified in-home rehabilitation for stroke survivors: Analytical review. *International Journal of Serious Games*, *5*(1), 2384–8766. https://doi.org/10.17083/ijsg.v5i1.224

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in esl/efl pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Tikofsky, R. S. (1970). A revised list for the estimation of dysarthric single word intelligibility. *Journal of Speech and Hearing Research*, *13*(1), 59–64. https://doi.org/10.1044/jshr.1301.59

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*(4), 358. https://doi.org/10.1037/h0076640

Tjaden, K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics*, *9*(2), 139–154. https://doi.org/10.3109/02699209508985329

Tjaden, K., & Wilding, G. (2010). The impact of rate reduction and increased loudness on fundamental frequency characteristics in dysarthria. *Folia Phoniatrica et Logopaedica*, *63*(4), 178–186. https://doi.org/10.1159/000316315

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria. *Journal of Speech, Language, and Hearing Research*, *47*(4), 766–783. https://doi.org/10.1044/1092-4388(2004/058)

Tjaden, K., & Wilding, G. E. (2011). Effects of speaking task on intelligibility in parkinson's disease. *Clinical Linguistics & Phonetics*, *25*(2), 155–168. https://doi.org/10.3109/02699206.2010.520185

Tomassi, N. E., Castro, M. E., Timmons Sund, L., Díaz-Cádiz, M. E., Buckley, D. P., & Stepp, C. E. (2023). Effects of sidetone amplification on vocal function during telecommunication. *Journal of Voice*, *37*(4), 553–560. https://doi.org/10.1016/j.jvoice.2021.03.027

Torres, A., Kapralos, B., & Dubrowski, A. (2025). Examining the Usability of the Moirai Serious Game Authoring Platform. *IEEE Transactions on Games*, *17*(1), 235–241. https://doi.org/10.1109/TG.2024.3360918

Türkbey, T. A., Kutlay, S., & Gök, H. (2016). Clinical feasibility of Xbox KinectTM training for stroke rehabilitation: A single-blind randomized controlled pilot study. *Journal of Rehabilitation Medicine*, *49*(1), 22–29. https://doi.org/10.2340/16501977-2183

Vachhani, B., Bhat, C., Das, B., & Kopparapu, S. K. (2017). Deep autoencoder based speech features for improved dysarthric speech recognition. *Proc. Interspeech 2017*, 1854–1858. https://doi.org/10.21437/Interspeech.2017-1318

Vachhani, B., Bhat, C., & Kopparapu, S. K. (2018). Data augmentation using healthy speech for dysarthric speech recognition. *Proc. Interspeech 2018*, 471–475. https://doi.org/10.21437/Interspeech.2018-1751

Van de Weijer, J., & Slis, L. (1991). Nasaliteitsmeting met de nasometer [measuring nasality using the nasometer]. *Logopedie & Foniatrie*, (63), 97–101. https://hdl.handle.net/2066/323177

Van der Bruggen, S., De Letter, M., & Rietveld, T. (2023). Effects of near-monotonous speech of persons with parkinson's disease on listening effort and intelligibility. *Clinical Linguistics & Phonetics*, *0*(0), 1–14. https://doi.org/10.1080/02699206.2023.2272032

van de Weijer, S. C., Duits, A. A., Bloem, B. R., Kessels, R. P., Jansen, J. F., Köhler, S., Tissingh, G., & Kuijf, M. L. (2016). The Parkin'Play study: protocol of a phase II randomized controlled trial to assess the effects of a health game on cognition in Parkinson's disease. *BMC neurology*, *16*(1), 209. https://doi.org/10.1186/s12883-016-0731-z

Van Dijk, D., Kresin, F., Reitenbach, M., Rennen, E., & Wildevuur, S. (2011). *Users as designers a hands-on approach to creative research*. https://waag.org/sites/waag/files/Publicaties/Users_as_Designers.pdf

Van Doremalen, J., Cucchiarini, C., & Strik, H. (2010). Optimizing Automatic Speech Recognition for Low-Proficient Non-Native Speakers. *EURASIP Jour-*

nal on Audio, Speech, and Music Processing, *973954*. https://doi.org/10.1155/2010/973954

Van Doremalen, J., Cucchiarini, C., & Strik, H. (2013). Automatic pronunciation error detection in non-native speech: The case of vowel errors in dutch. *The Journal of the Acoustical Society of America*, *134*(2), 1336–1347. https://doi.org/10.1121/1.4813304

Van Nuffelen, G., de Bodt, M., Vanderwegen, J., van de Heyning, P., & Wuyts, F. (2010). Effect of rate control on speech production and intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *62*(3), 110–119. https://doi.org/10.1159/000287209

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 & Communication Disorders*, *44*(5), 716–730. https://doi.org/10.1080/13682820802342062

Van Vilsteren, C., Gerritsen, E., Pols, H., de Jong, L., Hogendoorn, P., Schikan, H., & van Meeteren, N. (2019). *Knowledge and Innovation Agenda 2020-2023 Health and Care*. https://www.health-holland.com/sites/default/files/downloads/Knowledge-and-Innovation-Agenda-2020-2023-health-and-care.pdf

Wagner, D., Baumann, I., Engert, N., Lee, S., Nöth, E., Riedhammer, K., & Bocklet, T. (2025). Personalized Fine-Tuning with Controllable Synthetic Speech from LLM-Generated Transcripts for Dysarthric Speech Recognition. https://doi.org/10.48550/arXiv.2505.12991

Wallach, D., & Scholz, S. C. (2012). User-Centered Design: Why and How to Put Users First in Software Development. In A. Maedche, A. Botzenhardt, & L. Neer (Eds.), *Software for people: Fundamentals, trends and best practices* (pp. 11–38). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31371-4_2

Walshe, M., & Miller, N. (2011). Living with acquired dysarthria: The speaker's perspective. *Disability and rehabilitation*, *33*(3), 195–215. https://doi.org/10.3109/09638288.2010.511685

Walshe, M., Miller, N., Leahy, M., & Murray, A. (2008). Intelligibility of dysarthric speech: Perceptions of speakers and listeners. *International Journal of Language & Communication Disorders*, *43*(6), 633–648. https://doi.org/10.1080/13682820801887117

Wan, V., & Carmichael, J. (2005). Polynomial dynamic time warping kernel support vector machines for dysarthric speech recognition with sparse training data. *INTERSPEECH*, 3321–3324. https://doi.org/10.21437/Interspeech.2005-853

Wang, P., & Van Hamme, H. (2023). Benefits of pre-trained mono- and cross-lingual speech representations for spoken language understanding of dutch dysarthric speech. *EURASIP Journal on Audio, Speech, and Music Processing*, *15*. https://doi.org/10.1186/s13636-023-00280-z

Wanick, V., & Bitelo, C. (2020). Exploring the use of participatory design in game design: a Brazilian perspective. *International Journal of Serious Games*, *7*(3), 3–20. https://doi.org/10.17083/ijsg.v7i3.358

Weber, L. (2025, January). *Analysis, design and prototypical implementation of a serious game to aid in speech rehabilitation.* [Master's thesis]. Technische Universität Wien. https://doi.org/10.34726/hss.2025.123480

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2000). Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatrica et Logopaedica*, *53*(1), 1–18. https://doi.org/10.1159/000052649

Weismer, G., & Laures, J. S. (2002). Direct magnitude estimates of speech intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, *45*(3), 421–433. https://doi.org/10.1044/1092-4388(2002/033)

Wenke, R. J., Theodoros, D., & Cornwell, P. (2008). The short- and long-term effectiveness of the lsvt®for dysarthria following tbi and stroke. *Brain Injury*, *22*(4), 339–352. https://doi.org/10.1080/02699050801960987

Wight, S., & Miller, N. (2015). Lee silverman voice treatment for people with parkinson's: Audit of outcomes in a routine clinic. *International Journal of Language & Communication Disorders*, *50*(2), 215–225. https://doi.org/10.1111/1460-6984.12132

Witt, S. M. (1999). *Use of speech recognition in computer-assisted language learning* [Doctoral dissertation, Newnham College, University of Cambridge]. https://www.repository.cam.ac.uk/handle/1810/251707

World Health Organization, W. (2001). *International Classification of Functioning, Disability and Health* (1st ed.). WHO Press.

World Health Organization, W. (2005). *Resolution WHA58.28: eHealth.* http://extranet.who.int/iris/bitstream/10665/20378/1/WHA58_28-en.pdf?ua=1

World Health Organization, W. (2022, October). *Ageing and health [online]*. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

Xiong, F., Barker, J., & Christensen, H. (2019). Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised Dysarthric Speech Recognition. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5836–5840. https://doi.org/10.1109/ICASSP.2019.8683091

Xiong, F., Barker, J., Yue, Z., & Christensen, H. (2020). Source Domain Data Selection for Improved Transfer Learning Targeting Dysarthric Speech Recognition. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7424–7428. https://doi.org/10.1109/ICASSP40776.2020.9054694

Xue, W., Van Hout, R., Boogmans, F., Ganzeboom, M., Cucchiarini, C., & Strik, H. (2021). Speech Intelligibility of Dysarthric Speech: Human Scores and Acoustic-Phonetic Features. *Proc. Interspeech 2021*, 2911–2915. https://doi.org/10.21437/Interspeech.2021-1189

Xue, W., Van Hout, R., Catia, C., & Strik, H. (2023). Assessing speech intelligibility of pathological speech in sentences and word lists: The contribution of phoneme-level measures. *Journal of Communication Disorders*, *102*, 106301. https://doi.org/10.1016/j.jcomdis.2023.106301

Xue, W., Van Hout, R., Cucchiarini, C., & Strik, H. (2021). Assessing speech intelligibility of pathological speech: Test types, ratings and transcription measures. *Clinical Linguistics & Phonetics*, *37*(1), 52–76. https://doi.org/10.1080/02699206.2021.2009918

Yeo, E. J., Choi, K., Kim, S., & Chung, M. (2023). Speech intelligibility assessment of dysarthric speech by using goodness of pronunciation with uncertainty quantification. *Proc. Interspeech 2023*, 166–170. https://doi.org/10.21437/Interspeech.2023-173

Yin, S.-C., Rose, R. C., Saz, O., & Lleida, E. (2009). A study of pronunciation verification in a speech therapy application. *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 4609–4612. https://doi.org/10.1109/icassp.2009.4960657

Yılmaz, E., Ganzeboom, M., Beijer, L., Cucchiarini, C., & Strik, H. (2016). A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. https://aclanthology.org/L16-1127/

Yılmaz, E., Mitra, V., Bartels, C., & Franco, H. (2018). Articulatory features for asr of pathological speech. *Proc. Interspeech 2018*, 2958–2962. https://doi.org/10.21437/Interspeech.2018-67

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, *11*(6), 499–512. https://doi.org/10.1016/0021-9924(78)90024-2

Yu, D., & Deng, L. (2015). *Automatic speech recognition: A deep learning approach.* Springer-Verlag London. https://doi.org/10.1007/978-1-4471-5779-3

Yu, J., Xie, X., Liu, S., Hu, S., Lam, M. W. Y., Wu, X., Wong, K. H., Liu, X., & Meng, H. (2018). Development of the CUHK Dysarthric Speech Recognition System for the UA Speech Corpus. *Proc. Interspeech 2018*, 2938–2942. https://doi.org/10.21437/Interspeech.2018-1541

Yue, Z., Loweimi, E., Christensen, H., Barker, J., & Cvetkovic, Z. (2022). Acoustic Modelling From Raw Source and Filter Components for Dysarthric Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 2968–2980. https://doi.org/10.1109/TASLP.2022.3205766

Yue, Z., Loweimi, E., Cvetkovic, Z., Christensen, H., & Barker, J. (2022). Multi-Modal Acoustic-Articulatory Feature Fusion For Dysarthric Speech Recognition. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7372–7376. https://doi.org/10.1109/ICASSP43922.2022.9746855

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *J Speech*

Lang Hear Res., *48*(6), 1294–1310. https://doi.org/10.1044/1092-4388(2005/090)

Yunusova, Y., Kearney, E., Kulkarni, M., Haworth, B., Baljko, M., & Faloutsos, P. (2017). Game-based augmented visual feedback for enlarging speech movements in parkinson's disease. *Journal of Speech, Language, and Hearing Research*, *60*(6S), 1818–1825. https://doi.org/10.1044/2017_JSLHR-S-16-0233

Zaidi, B. F., Selouani, S. A., Boudraa, M., & Sidi Yakoub, M. (2021). Deep neural network architectures for dysarthric speech analysis and recognition. *Neural Computing and Applications*, *33*(15), 9089–9108. https://doi.org/10.1007/s00521-020-05672-2

Zavaliagkos, G., Austin, S., Makhoul, J., & Schwartz, R. (1993). A hybrid continuous speech recognition system using segmental neural nets with hidden markov models. *International Journal of Pattern Recognition and Artificial Intelligence*, *07*(04), 949–963. https://doi.org/10.1142/S0218001493000480

Zheng, W.-Z., Han, J.-Y., Chen, C.-Y., Chang, Y.-J., & Lai, Y.-H. (2023). Improving the Efficiency of Dysarthria Voice Conversion System Based on Data Augmentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 4613–4623. https://doi.org/10.1109/TNSRE.2023.3331524

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. https://doi.org/10.1016/j.jcomdis.2008.05.001

Research Data Management

Existing and collected data

Except for chapter I and chapter 8, existing and/or collected data was used in the research described in every other chapter. In chapter 2, listener judgements were collected through an online survey on the basis of dysarthric speech recordings from an existing collection (Beijer, 2012). Pseudonymized data, mother tongue, gender, age, and familiarity with dysarthric speech, was collected regarding the non-expert listeners.

In chapter 3 and chapter 4 data from existing speech corpora were used to train and adapt acoustic models for dysarthric speech recognition. Neurologically healthy speech from the Spoken Dutch (Oostdijk et al., 2002) and JASMIN-CGN (Cucchiarini et al., 2008) corpora were selected and used for those purposes. Speech recordings from the Dutch Dysarthric Speech database (Yılmaz et al., 2016) were used to retrain or adapt acoustic models for the recognition of dysarthric speech. That database was published as part of the same project as the current thesis. The existing COPAS corpus (Middag, 2012) was also used for testing trained acoustic models' dysarthric speech recognition performance. For more details on which particular parts of the corpora were used, see chapter 3 and chapter 4 of the current thesis.

Research on the design of a serious game was described in chapter 5. Contact data of speech therapists and dysarthric speakers were collected for the runtime of the CHASING research project to make an appointment for participating in design activities and game prototype test sessions.

As part of the research in chapter 6 and chapter 7, two versions of the serious game were used in speech interventions to study the effects of game-based speech therapy. The already existing E-learning based Speech Therapy (EST) system (Beijer, 2012) was also used for comparison. Continuous speech of all enrolled dysarthric speakers and their neurologically healthy coplayers was collected during all practice sessions with the serious game and EST. Speech was also recorded at multiple measurement intervals given predefined words, utterances, but also spontaneous speech. The following personal data was collected of all participants for effect analysis purposes: age, diagnosis, time since diagnosis, mobility limitations, perceived impact on daily communication, gender, and experience with computers.

For the analysis of intervention effects, two listening experiments were used in which the recordings of the multiple measurement intervals were used to judge the speech intelligibility. From these non-expert listeners the following personal data were collected: mother tongue, gender, age, and familiarity with dysarthric speech.

After having completed the speech intervention, dysarthric speakers were asked about how satisfied they were practising with the serious game and EST system. As part of chapter 6 a brief online survey on user satisfaction was completed as well as a short experiment on preference for game-based and EST-based speech interventions. A similar survey and experiment on user satisfaction and preference, respectively, were held as part of the research in chapter 7. However, this survey contained additional open questions on the provided automatic feedback and preference for game-based speech therapy versus face-to-face sessions.

Informed consent

Dysarthric speakers were recruited via speech pathologists, patient internet fora and Facebook groups. They were given information about our research and assured they could withdraw at any time. They signed consent forms for participation and the collection of speech recordings. Ethics approval from Radboud University Medical Centre Research Ethics Committee was not required as stated by this committee (2015-1707).

For the listening experiments, participants were recruited among the students of the Radboud University and HAN University of Applied Sciences in Nijmegen. Participants were asked to come to a class room in which multiple PCs or laptops were set up with the listening experiment. Before starting the experiment, participants were asked to provide multiple characteristics, required for valid statistical analysis, and their consent for data collection through multiple forms. These forms followed the informed consent procedure established by the Ethics Committee at the Radboud University Medical Centre.

Protection of participants' privacy

Protecting the privacy of the individuals participating in a scientific experiment is important to prevent misuse of the data they provided. This is even more important when working with speech recordings recorded in a rehabilitation setting, as the voice of an individual is often considered as personally identifyable information (PII). Where applicable, speech recordings were made in a pseudonymized way by using a key in the form of the participant number. Any PII of recorded participants was stored in a separate file. The file holding the personal data and the key linking to participant names was registered in accordance with the university's safety regulations for sensitive and critical data and is kept for a minimum of 15 years. The judgements of the participants in the listening experiment were collected anonymously. Any characteristics of those participants, required for valid statistical analysis, were stored in a separate file.

Data storage in the context of scientific integrity

Data were collected and stored in multiple contexts. During the practice sessions of the game-based speech interventions, the dysarthric speakers and coplayer were both recorded locally on their own Apple iPads running the game. The recordings were moved to the secure and versioned storage of the Science Faculty at Radboud University after the speech intervention ended. While practising with the web-based speech intervention using Radboud University supplied laptops, the dysarthric speakers were recorded online via the server that ran the web-based speech intervention. This was a secure server running at the server park of the Science Faculty at Radboud University. The connection to the server was secured and encrypted using SSL technology. After the speech intervention was completed, recordings were moved to the same secure storage facility as the game-based recordings.

At the pre and post measurement times, dysarthric speakers were recorded using a separate course in the web-based speech intervention. At a later point in that session, during the user satisfaction questionnaire, they were also recorded when asked to provide additional comments, if any, on their experiences using the game-based or web-based speech intervention. All of the above recordings were moved to the secure storage facility of the Science Faculty after the session was completed.

The judgements collected in the listening experiments were initially stored locally on the secured PC or laptop alongside the OpenSesame (Mathôt et al., 2012) listening experiment file. They were moved to the secure storage facility of the Science Faculty after the session was completed.

Selected speech recordings used from the Spoken Dutch (CGN) and JASMIN corpora were also stored at the secure storage facility of the Science Faculty. They were used for training Automatic Speech Recognition models via a network connection that was secured and encrypted using OpenSSH[1] technology.

Giving access to the data

Research in the CHASING project, which includes the current thesis, was funded by the NWO research grant with Ref. no. 314-99-101 (CHASING). Although no real requirements for sharing of data were provided, NWO encourages sharing of data for transparency and reproducibility of research results. Additionally, sharing of data enables its re-use in future endeavours. For those reasons, research data were shared via the Radboud Data Repository (https://data.ru.nl). Data collected in chapter 6 are available under DOI: https://doi.org/10.34973/0w4g-bb09. Those from chapter 7 are available under DOI: https://doi.org/10.34973/h19s-c122. Both datasets are available under restricted access due to the private nature of recordings

[1] For more information see: https://www.openssh.com. Last accessed on February 23, 2025.

made in a speech rehabilitation context. Access can be obtained by using or applying for one of the supported authentication methods. Authorization is then given by one of the collection owners. See https://data.ru.nl/login for more details.

Summary

As populations around the world are ageing, the group of patients with acquired neurological disorders such as Parkinson's disease (PD), Cerebral Vascular Accident (CVA or stroke) and traumatic brain injury (TBI) is growing. Many of these individuals are affected by dysarthria, a neurologic speech disorder that reflects abnormalities in one or more aspects of speech production. Due to these abnormalities, their speech becomes less intelligible. The field of speech rehabilitation aims to improve intelligibility by restoring as much as possible of the affected aspects of speech production.

Speech rehabilitation is normally provided by speech therapists, but the need for intensive therapy and the growing number of patients make it difficult for therapists to cope with the demand. Therefore, there is a growing attention for telerehabilitation: the delivery of healthcare services remotely. Attempts have been made at finding telerehabilitation solutions to provide speech training in patients' home environment. Promising solutions include automatic feedback on pronunciation using speech technology. Patients that participated in corresponding effect studies described being able to follow the training program, but having increasing difficulty adhering to its 'drill-and-practice' nature and not experiencing beneficial effects in daily-life communication. Exploring different approaches to remote and independent speech training potentially provides valuable insights on these topics. Interactive gaming is one such approach with the goal to leverage its entertainment and enjoyment properties for a 'serious' purpose, hence the term 'serious gaming'. For that reason, the current thesis explores the use of serious gaming including methods of automatic feedback for the purpose of remote and independent speech training.

A general introduction to the three involved research topics, dysarthria and measurement of intelligibility, serious gaming, and speech technology is provided in **chapter I**. It sheds some light on past and more recent research into these topics. The CHASING research project, that stands for 'Challenging speech training in neurological patients using serious gaming', is also introduced, in addition to the research questions addressed by the current thesis.

In **chapter 2**, research on objective measurements for the intelligibility of dysarthric speech is described. Measuring the intelligibility of disordered speech is a common practice in both clinical and research contexts. It enables the measurement of the effects of a particular speech intervention. Over the years various methods have been proposed and studied, but many of these methods measure speech intelligibility at the speaker or utterance level. While this may be satisfactory for some purposes, more detailed evaluations might be required in other cases such as diagnosis and measuring or comparing the outcomes of different types of therapy (by humans or computer programs). In **chapter 2**, intelligibility ratings are investigated at three different levels of granularity: utterance, word, and subword level. In a web experiment 50 speech fragments produced by seven dysarthric speakers were rated by 36 listeners in three ways: a score per utterance on a Visual Analogue and a Likert scale, and an orthographic transcription. The latter was used to obtain word and subword (grapheme and phoneme) level ratings using automatic alignment and conversion methods. The implemented phoneme scoring method proved feasible, reliable, and provided a more sensitive and informative measure of intelligibility.

Research into automatic recognition of dysarthric speech, a technology commonly used in providing feedback on pronunciation during speech training, is described in **chapter 3**. The impact of combining speech data from different varieties of the Dutch language in training deep neural network (DNN)-based acoustic models is investigated. Flemish is chosen as the target variety for testing the acoustic models, since a Flemish database of pathological speech was available. Non-pathological speech data from the northern Dutch and Flemish varieties were used and speaker-independent recognition was performed using the DNN-HMM system trained on the combined data. The results showed that this system provided improved recognition of pathological Flemish speech compared to a baseline system trained only on Flemish data.

Improving automatic recognition of dysarthric speech is further investigated in **chapter 4**. A multi-stage deep neural network (DNN) training scheme aimed at obtaining better modeling of dysarthric speech by using only a small amount of in-domain training data is researched. The results show that the system employing the proposed training scheme considerably improves the recognition of Dutch dysarthric speech compared to a baseline system with single-stage training only on a large amount of normal speech or a small amount of in-domain data.

In **chapter 5**, research into the design and development of two versions of a serious game for speech training in older adults is described. Using a CoDesign approach, two tablet games, called Treasure Hunters 1 and 2, were designed. The games provided automatic feedback on users' speech and enabled training speech at home. Valuable lessons were learned in multiple design areas. The most notable ones were that target users preferred a cooperative style of play and freedom of movement within a game. Also, the need for sharing information between players was found to motivate speech production. When designing the feedback on speech, an indirect approach combined with a direct approach showed the most potential. These and other valuable lessons are described in more detail in chapter 5.

In Treasure Hunters 1 two players interact verbally to find the way to the treasure, while receiving automatic feedback on voice loudness and pitch. Participants played with the game in several sessions and generally appreciated it, hinting at its potential for speech training in elderly patients. In **chapter 6**, a within-subjects experiment with five dysarthric patients is described in which Treasure Hunters 1 was compared to a non-game computer-based speech training system: E-learning-based Speech Therapy (EST). The research focused on three variables: speech intelligibility, user

satisfaction, and user preference. Substantial variability between participants was observed, in the outcomes of these three variables and their relations. This showed that a ’one size that fits all’ does not apply to computer-based speech training, but a personalised approach is needed.

In **chapter 7**, research into the effects of Treasure Hunters 2 on speech intelligibility and patient satisfaction are described. Treasure Hunters 2 is an improved game-based speech training that provides automatic feedback on loudness, pitch and pronunciation. Eight adult dysarthric speakers with Parkinson’s disease completed a four-week game-based speech training in their home environment. A significant, positive effect was found on speech intelligibility of the game-based speech training period in comparison to a period of no training. Additionally, patients generally seemed satisfied with the game. They also generally agreed with the automatic feedback and could use it to positively change the way they spoke.

Using the previously described research as input, **chapter 8** provides a general discussion reflecting the three main topics of the current thesis, dysarthria and measurement of intelligibility, serious gaming, and speech technology. The research questions defined in the first chapter are revisited. Discussions on the limitations of the current thesis and future research opportunities are also included.

As the current thesis has made clear, providing reliable automatic feedback is of great importance when training at home. The current thesis researched automatic feedback using speech analysis and automatic speech recognition algorithms as part of game-based speech training. It has shown that speech analysis algorithms can be used to provide automatic feedback on loudness and pitch using threshold levels adequately. It has also shown that providing pronunciation feedback on word-level in a near real-time gaming context is technically feasible and can have positive effects on speech intelligibility. Additionally, dysarthric speakers are able to interpret that feedback and adjust their pronunciation while playing the game. However, the current thesis has also shown that a "one size fits all" approach to game-based speech training does not apply. Individual dysarthric speakers or at least groups of dysarthric speakers have individual preferences towards games that they would like to play, feedback thresholds that need to be personalized and pronunciation feedback that needs to be calibrated per individual. From a clinical perspective, the current thesis has explored the potentials of game-based speech training as an alternative to face-to-face and web-based courseware systems. Although the game-based speech training was received well, individual speech training scenarios may require a different set of speech exercises that is better suited for a face-to-face or web-based approach to speech training. For that reason, the current thesis regards all approaches to be needed. This way, the strengths of all types of training can be used to improve or stabilize speech intelligibility in patients with dysarthria.

Samenvatting

Naarmate de wereldbevolking vergrijst, groeit de groep patiënten met verworven neurologische aandoeningen zoals de ziekte van Parkinson (PD), cerebraal vasculair accident (CVA of beroerte) en traumatisch hersenletsel (TBI). Veel van deze mensen hebben last van dysartrie, een neurologische spraakstoornis die afwijkingen veroorzaakt in een of meer aspecten van de spraakproductie. Door deze afwijkingen wordt hun spraak minder verstaanbaar. Het doel van spraakrevalidatie is om de verstaanbaarheid te verbeteren door de aangetaste aspecten van de spraakproductie zoveel mogelijk te herstellen.

Spraakrevalidatie wordt normaal gesproken gegeven door logopedisten, maar de noodzaak voor intensieve therapie en het groeiende aantal patiënten maken het moeilijk voor therapeuten om aan de vraag te voldoen. Daarom is er steeds meer aandacht voor telerevalidatie: het leveren van zorg op afstand. Er is onderzoek gedaan naar telerevalidatieoplossingen voor het geven van spraaktraining in de thuisomgeving van patiënten. Veelbelovende oplossingen maken gebruik van automatische feedback op uitspraak met behulp van spraaktechnologie. Patiënten in corresponderende effectstudies beschreven dat ze in staat waren om het trainingsprogramma te volgen, maar dat ze steeds meer moeite hadden om zich aan het programma te houden door het ’drill-and-practice’-karakter van de oefeningen. Tevens ondervonden ze geen positieve effecten van het trainingsprogramma in de normale, dagelijkse communicatie. Het verkennen van verschillende benaderingen om zelfstandig spraak trainen op afstand te realiseren, kan tot waardevolle inzichten leiden. Interactief gamen is één zo’n benadering, met het doel om de entertainment- en pleziereigenschappen van gaming te benutten voor een ’serieus’ doel, vandaar ook wel de term ’serious gaming’. Om die reden onderzoekt dit proefschrift het gebruik van serious gaming, met inbegrip van methoden voor automatische feedback, met als doel het realiseren van zelfstandige spraaktraining op afstand.

In **hoofdstuk 1** wordt een algemene inleiding gegeven op de drie betrokken onderzoeksonderwerpen, dysartrie en het meten van verstaanbaarheid, serious gaming en spraaktechnologie. Het geeft een beeld van onderzoek uit het verleden, maar ook recenter onderzoek naar deze onderwerpen. Het CHASING onderzoeksproject, dat staat voor ’Challenging speech training in neurological patients using serious gaming’, wordt ook geïntroduceerd in hoofdstuk 1, naast de onderzoeksvragen die in dit proefschrift aan de orde komen.

In **hoofdstuk 2** wordt onderzoek naar objectieve metingen voor de verstaanbaarheid van dysartrische spraak beschreven. Het meten van de verstaanbaarheid van spraakstoornissen is gebruikelijk in zowel klinische als onderzoekscontexten. Het maakt het mogelijk om de effecten van een bepaalde spraakinterventie te meten. In de loop der jaren zijn verschillende methoden voorgesteld en bestudeerd, maar veel van deze

methoden meten spraakverstaanbaarheid op spreker- of uitingsniveau. Hoewel dit voor sommige doeleinden kan volstaan, zijn er in andere gevallen meer gedetailleerde evaluaties nodig, zoals bij diagnose en het meten of vergelijken van de resultaten van verschillende soorten therapie (door mensen gegeven of computerprogramma's). In hoofdstuk 2 worden beoordelingen van verstaanbaarheid onderzocht op drie verschillende granulariteitsniveaus: uitingsniveau, het woord- en het subwoordniveau. In een webexperiment werden 50 spraakfragmenten van zeven dysartrische sprekers door 36 luisteraars op drie manieren beoordeeld: een score per uiting op een Visueel Analoge en een Likertschaal, en een orthografische transcriptie. Deze laatste werd gebruikt om woord- en subwoordscores (grafeem en foneem) te verkrijgen met behulp van automatische uitlijnings- en conversiemethoden. De geïmplementeerde foneemscoremethode bleek haalbaar en betrouwbaar en leverde een gevoeligere en informatievere maat voor verstaanbaarheid op.

Onderzoek naar automatische herkenning van dysartrische spraak, een technologie die vaak wordt gebruikt bij het geven van feedback op uitspraak tijdens spraaktraining, wordt beschreven in **hoofdstuk 3**. De impact van het combineren van spraakdata van verschillende varianten van de Nederlandse taal in het trainen van op deep neural networks (DNN) gebaseerde akoestische modellen wordt onderzocht. Vlaams is gekozen als variant voor het testen van de akoestische modellen, omdat er een Vlaamse database met pathologische spraak beschikbaar was. Niet-pathologische spraakgegevens van de Noord-Nederlandse en Vlaamse varianten werden gebruikt en sprekeronafhankelijke herkenning werd uitgevoerd met behulp van het DNN-HMM-systeem dat getraind werd op de gecombineerde spraakgegevens. De resultaten toonden aan dat dit systeem verbeterde herkenning van pathologische Vlaamse spraak bood in vergelijking met een basissysteem dat alleen op Vlaamse data was getraind.

Het verbeteren van de automatische herkenning van dysartrische spraak wordt verder onderzocht in **hoofdstuk 4**. Een multi-stage deep neural network (DNN) trainingsschema gericht op het verkrijgen van betere modellering van dysartrische spraak door gebruik te maken van slechts een kleine hoeveelheid in-domain trainingsdata wordt onderzocht. De resultaten laten zien dat het systeem dat gebruik maakt van het voorgestelde trainingsschema de herkenning van Nederlandse dysartrische spraak aanzienlijk verbetert in vergelijking met een basissysteem met single-stage training op alleen een grote hoeveelheid normale spraak of een kleine hoeveelheid in-domain data.

In **hoofdstuk 5** wordt onderzoek beschreven naar het ontwerp en de ontwikkeling van twee versies van een serious game voor spraaktraining bij oudere volwassenen. Met behulp van een CoDesign-aanpak werden twee tablet games, genaamd Treasure Hunters 1 en 2, ontworpen. De games gaven automatische feedback over de spraak van gebruikers en maakten het mogelijk om thuis spraak te trainen. Er werden waardevolle lessen geleerd op meerdere ontwerpgebieden. De meest opmerkelijke waren dat doelgebruikers de voorkeur gaven aan een coöperatieve speelstijl en bewegingsvrijheid binnen een spel. Ook bleek de behoefte aan het delen van informatie tussen spelers de spraakproductie te motiveren. Bij het ontwerpen van de feedback op spraak bleek een indirecte benadering in combinatie met een directe benadering het meeste potentieel te bieden. Deze en andere waardevolle lessen worden in meer detail beschreven in hoofdstuk 5.

In Treasure Hunters 1 interacteren twee spelers verbaal om de weg naar de schat te vinden, terwijl ze automatische feedback krijgen over de luidheid en toonhoogte van hun stem. Deelnemers speelden met de game in verschillende sessies en konden het over het algemeen waarderen, wat wijst op de mogelijkheden voor spraaktraining bij oudere patiënten. In **hoofdstuk 6** wordt een within-subjects experiment met vijf dysartrische patiënten beschreven waarin Treasure Hunters 1 werd vergeleken met een op een computer gebaseerd spraaktrainingssysteem zonder game elementen: E-learning-gebaseerde Spraaktherapie (EST). Het onderzoek richtte zich op drie variabelen: spraakverstaanbaarheid, gebruikerstevredenheid en gebruikersvoorkeur. Er werd een aanzienlijke variabiliteit tussen de deelnemers waargenomen in de uitkomsten van deze drie variabelen en hun relaties. Dit toonde aan dat een 'one size that fits' all niet van toepassing is op computergebaseerde spraaktraining, maar dat een gepersonaliseerde aanpak nodig is.

In **hoofdstuk 7** wordt onderzoek beschreven naar de effecten van Treasure Hunters 2 op spraakverstaanbaarheid en patiënttevredenheid. Treasure Hunters 2 is een verbeterde, spelgebaseerde spraaktraining die automatische feedback geeft op luidheid, toonhoogte en uitspraak. Acht volwassen dysartrische sprekers met de ziekte van Parkinson volgden een vier weken durende spraaktraining in hun thuisomgeving. Er werd een significant, positief effect gevonden op de spraakverstaanbaarheid van de trainingsperiode in vergelijking met een periode zonder training. Bovendien leken de patiënten over het algemeen tevreden met het spel. Ze waren het ook over het algemeen eens met de automatische feedback en konden deze gebruiken om hun manier van spreken positief te veranderen.

Met het eerder beschreven onderzoek als input, bevat **hoofdstuk 8** de algemene discussie met als leidraad de drie hoofdonderwerpen van het huidige proefschrift, dysartrie en het meten van verstaanbaarheid, serious gaming en spraaktechnologie. De onderzoeksvragen uit het eerste hoofdstuk worden opnieuw besproken en beantwoord met behulp van de resultaten uit de eerdere hoofdstukken. Verder worden de beperkingen van het huidige proefschrift en suggesties voor toekomstig onderzoek bediscussieerd.

Zoals uit het huidige proefschrift blijkt, is het geven van betrouwbare automatische feedback van groot belang voor het geven van spraaktraining op afstand bij patiënten thuis. In het huidige proefschrift is onderzoek gedaan naar automatische feedback met behulp van spraakanalyse en automatische spraakherkenningsalgoritmen als onderdeel van een spelgebaseerde spraaktraining. Het heeft aangetoond dat spraakanalysealgoritmen gebruikt kunnen worden om automatische feedback te geven over luidheid en toonhoogte door adequaat gebruik te maken van drempelniveaus. Er is ook aangetoond dat het geven van uitspraakfeedback op woordniveau in een bijna realtime gamecontext technisch haalbaar is en positieve effecten kan hebben op de spraakverstaanbaarheid. Bovendien zijn dysartrische sprekers in staat om die feedback te interpreteren en hun uitspraak aan te passen tijdens het spelen van het spel. Het huidige proefschrift heeft echter ook aangetoond dat een “one size fits all”-benadering van een op een spelgebaseerde spraaktraining niet van toepassing is. Individuele dysartrische sprekers of in ieder geval groepen dysartrische sprekers hebben individuele voorkeuren voor games die ze willen spelen, feedbackdrempels die gepersonaliseerd moeten worden en feedback op uitspraak dat per individu gekalibreerd moet worden. Vanuit een klinisch perspectief heeft het huidige proefschrift de mogelijkheden onderzocht van een spelgebaseerde spraaktraining als alternatief voor face-to-face en webgebaseerde trainingssystemen. Hoewel de spelgebaseerde spraaktraining goed werd ontvangen, kunnen individuele spraakoefenscenario’s een andere set spraakoefeningen vereisen dat beter geschikt is voor een face-to-face of webgebaseerde aanpak van spraaktraining. Om die reden beschouwt het huidige proefschrift alle benaderingen als noodzakelijk. Op deze manier kunnen de sterke punten van alle soorten spraaktraining gebruikt worden om de spraakverstaanbaarheid bij patiënten met dysartrie te verbeteren of te stabiliseren.

Deze samenvatting is tot stand gekomen door automatische vertaling van de Engelse samenvatting via DeepL’s translate text functie[2]. Het resultaat is nadien bewerkt door de auteur van dit proefschrift.

[2] DeepL’s translate text functie: https://www.deepl.com/en/translator. Laatst gebruikt op 1 juli 2025.

Contributions

Chapter 2

Chapter 2 is based on an original research article featuring the PhD candidate as first author. The article was co-authored by dr. Marjoke Bakker and dr. Catia Cucchiarini, and promotor dr. Helmer Strik (co-promotor at the time of writing).

PhD candidate
The candidate conducted a literature review, converted all previously collected data, calculated the intelligibility scores, assisted in the analysis and interpretation of the scores, and wrote the draft version of the research article.

Co-authors
The co-authors assisted in the analysis and interpretation of the scores and provided extensive feedback on the draft version of the research article.

Chapters 3 and 4

Chapters 3 and 4 are based on original research articles featuring the PhD candidate as second author. The articles first author is dr. Emre Yılmaz. Other co-authors were dr. Catia Cucchiarini, and dr. Helmer Strik.

PhD candidate
The candidate conducted a literature review, assisted in the experiment design, analysis and interpretation of the experiment results, and wrote the draft version of the research article.

First author
The first author proposed the experiment design, selected the data sets to be used in the experiments, ran the experiments, led the analysis and interpretation of the experiment results and co-authored the research article.

Remaining co-authors
The remaining co-authors assisted in the analysis and interpretation of experiment results and provided extensive feedback on the draft version of the research article.

Chapter 5

Chapter 5 is based on an original research article featuring the PhD candidate as first author. The article was co-authored by promotor dr. Helmer Strik (co-promotor at the time of writing), and co-promotor prof. dr. Toni Rietveld (promotor at the time of writing). Additionally, the article was based on a technical report written by Douwe-Sjoerd Boschman of Waag, a creative partner in the research project. He led the game concept development process.

PhD candidate
The candidate worked closely with Douwe-Sjoerd Boschman to provide ideas and input on the topics of dysarthria, dysarthric speakers and their capabilities, and dysarthric speech technology during the game concept development process. Also, the candidate conducted a literature review, analysed the lessons learned, and wrote the draft version of the research article.

Co-authors
The co-authors also worked closely with Douwe-Sjoerd Boschman to provide ideas and input on the same topics as the PhD candidate. Additionally, they provided extensive feedback on the draft version of the research article.

Chapters 6 and 7

Chapters 6, and 7 are based on original research articles featuring the PhD candidate as first author. The articles were co-authored by dr. Marjoke Bakker, dr. Lilian Beijer, promotor dr. Helmer Strik (co-promotor at the time of writing), and co-promotor prof. dr. Toni Rietveld (promotor at the time of writing).

PhD candidate
The candidate proposed the experiment designs, assisted in coordinating and running the experiments with dysarthric speakers, set up the speech technology and corresponding infrastructure, manage the recording data set, create and run the listening experiments, assist in the analysis and interpretation of the results, and wrote the draft version of the research article.

Co-authors
The co-authors assisted in designing the experiments, in coordinating and running the experiments with dysarthric speakers, led the analysis and interpretation of the results, and provided extensive feedback on the draft version of the research article.

Curriculum Vitae

Mario Ganzeboom was born on August 5th, 1984 in Olst, The Netherlands. After completing secondary education (Mavo in 2000 and Havo in 2002), he started his software engineering education at the Saxion University of Applied Sciences in Enschede. In 2006, he graduated and continued his studies as a master student in Human Media Interaction at the University of Twente, also in Enschede. He obtained his Master of Science degree from the University of Twente in 2010. Afterwards, he took on a position as a scientific programmer supporting PhD students in linguistics and syntax. During this period his interest for research grew even more.

Late 2013 he came across the PhD position on the CHASING project at the Radboud University of Nijmegen and started there in the beginning of 2014. His PhD project focused on computer-assisted speech therapy using serious gaming for elderly whose speech was affected by a neurological disorder. He was also involved in several other language learning and therapy projects. Besides his research, Mario supervised several Master's students and student projects.

After his regular PhD contract ended late 2018 he started working at Telecats B.V. as a Natural Language Software Engineer and continued his PhD in his spare time. After two acquisitions, Telecats has transformed into Concentrix Netherlands CRM Services B.V., a part of the global Concentrix company. Mario continues his work there focusing on speech and language technology and Generative AI with Large Language Models in particular.

List of publications

Journal articles

Ganzeboom, M.S., Bakker, M., Beijer, L.J., Rietveld, A.C.M. & Strik, H. (2018). Speech training for neurological patients using a serious game. *British Journal of Educational Technology, 49* (4), 761-774. doi: 10.1111/bjet.12640

Ganzeboom, M.S., Bakker, M., Beijer, L.J., Strik, H. & Rietveld, A.C.M. (2022). A serious game for speech training in dysarthric speakers with Parkinson's disease. Exploring therapeutic efficacy and patient satisfaction. *International Journal of Language and Communication Disorders, 2022* (57), 808-821. doi: 10.1111/1460-6984.12722

Ganzeboom, M.S., Strik, H., Rietveld, T. (2025). Lessons learned from designing and developing a serious game for speech training in older adults. [Manuscript submitted for publication]. Centre for Language Studies, Radboud University Nijmegen.

Conference papers

Ganzeboom, M.S., Yilmaz, E., Cucchiarini, C. & Strik, H. (2016). An ASR-based Interactive Game for Speech Therapy. In *Proceedings of the 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)* (pp. 63-68). San Francisco, CA, USA: ISCA

Ganzeboom, M.S., Bakker, M., Cucchiarini, C. & Strik, H. (2016). Intelligibility of Disordered Speech: Global and Detailed Scores. In *Proceedings of Interspeech 2016* (pp. 2503-2507). San Franciso, CA, USA: ISCA doi: 10.21437/Interspeech.2016-1448

Ganzeboom, M.S., Yilmaz, E., Cucchiarini, C. & Strik, H. (2016). On the Development of an ASR-based Multimedia Game for Speech Therapy: Preliminary Results. In *Proceedings of the International Workshop on Multimedia for Personal Health and Health Care (MM Health)* (pp. electr.). Amsterdam, Netherlands: ACM

Yilmaz, E., Ganzeboom, M.S., Beijer, L.J., Cucchiarini, C. & Strik, H. (2016). A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC) 2016* (pp. 792-795). Portoroz, Slovenia: European Language Resources Association

Yilmaz, E., Ganzeboom, M.S., Cucchiarini, C. & Strik, H. (2016). Combining Non-pathological Data of Different Language Varieties to Improve DNN-HMM Performance on Pathological Speech. In *Proceedings of Interspeech 2016* (pp. 218-222). San Francisco, CA, USA: ISCA doi: 10.21437/Interspeech.2016-109

Yilmaz, E., Ganzeboom, M.S., Cucchiarini, C. & Strik, H. (2017). Multi-stage DNN training for Automatic Recognition of Dysarthric Speech. In *Proceedings of Interspeech 2017* (pp. 2685-2689). Baixas: ISCA doi: 10.21437/Interspeech.2017-303

Poster presentations

Ganzeboom, M.S., Yilmaz, E., Cucchiarini, C. & Strik, H. (2016). Prototype ASR-based Multimedia Game for Speech Therapy. In *International Workshop on Multimedia for Personal Health and Health Care (MM Health).* Amsterdam, Netherlands

Ganzeboom, M.S. (2016). Valorisatie: Een op Automatische Spraakherkenning gebaseerde game voor spraak training. Demo tijdens de studiedag 'ICT in de zorg' van de Hogeschool van Arnhem en Nijmegen. In

Yilmaz, E., Ganzeboom, M.S., Bakker, M., Boschman, D., Loos, L., Ongering, J., Beijer, L.J., Rietveld, A.C.M., Cucchiarini, C. & Strik, H. (2016). A Serious Game for Speech Training in Neurological Patients. In *Show & Tell, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Shanghai, China

Oral presentations

Ganzeboom, M.S. (2014, November 24). *Advanced speech technology in a serious game for speech training: an introduction to the CHASING project.* KU Leuven, Oral presentation at the closing symposium 'Assistive Speech and Signal Processing' of the ALADIN project.

Ganzeboom, M.S. (2015, September 8). *Dysarthric speech: acoustic-phonetic features and intelligibility.* Chania, Griekenland, The International Symposium on Mono- and Bilingual Speech.

Ganzeboom, M.S. (2017, October 3). *Speech therapy through interactive gaming with automatic speech recognition.* Eindhoven, The Netherlands, Games for Health Europe Conference.

Research lectures

Ganzeboom, M.S., Strik, H., Cucchiarini, C. & Bakker, M. (2016). Valorisation: Lab session: Beter leren spreken door te gamen. Drongo Festival: Utrecht, The

Netherlands (2016, October 1).

Ganzeboom, M.S., Strik, H. & Cucchiarini, C. (2017). Lab session: CHASING: Motiverende spraaktherapie in de vorm van een game. Drongo Festival: Utrecht, The Netherlands (2017, September 29 - 2017, September 30).

Datasets

Ganzeboom, M.S., Bakker, M., Beijer, L.J., Strik, H. & Rietveld, A.C.M. (2025). Treasure Hunters: speech training for neurological patients using a serious game. Radboud Data Repository [Dataset]. doi: 10.34973/0w4g-bb09

Ganzeboom, M.S., Bakker, M., Beijer, L.J., Strik, H. & Rietveld, A.C.M. (2022). Treasure Hunters 2: exploration of speech training efficacy. Radboud Data Repository [Dataset]. doi: 10.34973/h19s-c122.

Dankwoord

Het is alweer 'even' geleden, dat ik aan mijn promotietraject begon. Januari 2014 was het, dat ik startte op de 8e verdieping van het Erasmusgebouw in Nijmegen. Na zo'n twee en een half jaar als wetenschappelijk programmeur aan de Rijksuniversiteit Groningen te hebben gewerkt en promotiestudenten te hebben ondersteund, begon het bij mij zelf toch ook te kriebelen. Alleen, vier jaar onderzoek doen naar één onderwerp of onderzoeksgebied? Ja, dat leek mij wel wat. Helemaal toen de vacature op het CHASING-project mij passeerde waarin de onderzoeksonderwerpen uit dit proefschrift werden omschreven. Kort door de bocht gezegd, werken met/aan games, natuurlijke interactietechnologie met ook nog eens een evidente maatschappelijke relevantie? Voor mij de ideale combinatie voor langdurig onderzoek: waar kon ik tekenen!? Er was natuurlijk wel een sollicitatieprocedure te doorlopen, maar aan het eind bleek ik toch het predicaat 'PhD candidate' te mogen gaan voeren!

Nu, een prachtig aantal jaren (zullen we ze maar niet tellen?) en drie kinderen verder, breng ik de periode van wetenschappelijk onderzoek doen ten einde. In dit gedeelte van mijn proefschrift wil ik dan ook menigeen danken die tijdens die periode direct of indirect hebben bijgedragen.

Allereerst bedank ik mijn huidige promotor, associate professor dr. Strik. Helmer, dank voor je begeleiding en supervisie al die jaren. Je ruime kennis en ervaring op het gebied van spraaktechnologie voor spraakpathologie en taalverwerving is bewonderenswaardig. Tijdens onze besprekingen, die ik altijd als prettig heb ervaren, kon ik daar altijd op rekenen. Je stuurde me tijdens die besprekingen ook altijd bij wanneer ik teveel de diepte in ging op een specifiek gedeelte van het onderzoek. Echter, wist je ook wanneer ik de behoefte had een onderwerp verder zelf uit te zoeken. Op persoonlijk vlak hebben we ook het één en ander van elkaar meegemaakt, zoals mijn bruiloft, dat ik vader werd, maar ook jouw 60-feest en pensionering waren memorabele momenten. Door mijn, zogezegde, 'lange staat van dienst' moest je in het afgelopen jaar zelfs mijn promotor worden zodat ik überhaupt nog kon promoveren. Tot mijn geluk kon en wilde je dat, waarvoor heel veel dank!

Enorme dank gaat ook uit naar mijn huidige copromotor professor dr. Rietveld. Toni, ook jij hebt het lang met mij uitgehouden. Voor meer dan 10 jaar ben jij mij promotor geweest, zelfs nadat je met emeritaat ging. Helaas kon het ius promovendi niet (nog een keer) verlengd worden en slaagde ik er niet in het proefschrift binnen die termijn af te ronden. Desondanks, bleef jij 'am Bal', zoals ze colloquiaal in het Duits zouden uitdrukken, want ook in het laatste jaar als copromotor begeleidde je mij in het afronden van het proefschrift. Je grote kennis van fonetiek, methodologie, en statistiek waren van enorme waarde in het opzetten en uitvoeren van mijn gebruikers- en luisterexperimenten. Op persoonlijk vlak hebben we allebei iets met 'die deutsche Sprache'. Samen met collega Henk werd er, tot mijn genoegen, nog wel eens een klein woordje over de Duitse grens gesproken. Ook jij hebt mijn bruiloft en dat ik vader werd meegemaakt, naast dat ik dat heb wat betreft jouw emeritaat en verhuizing naar Harlingen. Mooi, hoe je jouw grote liefde voor het nautische daar verder hebt op kunnen pakken, maar hoe zat het ook alweer? Had je nu een boot of een schip? Volgens mij een scheepje[3]. Toni, dank voor al die jaren supervisie en je geduld in de lange afrondende fase van het proefschrift!

Mijn directe collega's in het CHASING-project in Nijmegen wil ik ook danken. Lilian Beijer, voor je inhoudelijke bijdrage vanuit de klinische wereld van de spraakpathologie en het meedenken over het opzetten van de experimenten met patiënten. Marjoke Bakker, voor je organisatietalent en kennis van statistiek. Catia Cucchiarini, voor je kennis op het gebied van toepassingsgericht inzetten van spraaktechnologie, en Emre Yilmaz, voor je kennis en enthousiasme over automatische spraakherkenning.

In Amsterdam zaten mijn andere CHASING-collega's, gelieerd aan Waag's Creative Care Lab: Sabine Wildevuur, Paulien Melis, en, Jurre Ongering, dank voor jullie project management vaardigheden met betrekking tot het ontwikkelen van de serious games. Douwe-Sjoerd Boschman voor je kennis en kunde op het gebied van game concept development en Lodewijk Loos voor je allround softwareontwikkelingsvaardigheden.

Met z'n allen hebben we een flink aantal goede, maar ook stevige sessies gehad om de multidisciplinaire aandachtsgebieden bij elkaar te brengen.

Natuurlijk vergeet ik niet mijn collega's van de 8e verdieping! Sara, Joop, Nelleke, Hans, Hella, Bert, Henk, Eric, Wessel, Micha, Erwin, Marina, Pepi, Remy, Roeland, Ferdy, Claire, Stephen, Bart, Joost, Wei, en alle anderen, dank voor jullie steun in de vorm van de befaamde koffiepauzes, lunches (de vrijdagse pannenkoek met appel), tot aan de toneelstukjes bij de promotiefeesten! Ik voelde me vanaf dag één direct op mijn plek.

Mijn huidige werkgever Telecats, inmiddels opgegaan in het internationale Concentrix wil ik tevens danken voor de tijd die ik in het begin van mijn functie kreeg om verder aan mijn proefschrift te werken. Tijdens werktijd heb ik zo mijn laatste resultaten kunnen verwerken in een publicatie. Daarnaast ook mijn collega's van de Speechlab-afdeling, Nadine, Gregor, Martin, Lluvia en Bertine. Dank voor het meeleven, maar met name jullie humor en relativerende woorden wat betreft de duur van het afronden van mijn proefschrift! Het duurde, maar het is nu toch mooi gelukt!

[3] Binnenvaarttaal, https://www.debinnenvaart.nl/binnenvaarttaal/index.php?scheepstype=boot, laatst geopend op 2 september, 2025.

Paranimfen, jullie rol is misschien niet ingewikkeld, maar het idee alleen al twee mede-ICT'ers naast mij te hebben staan gaf deze burger moed. Isaac en Wessel, enorm dank dat jullie mij bijstaan tijdens mijn verdediging!

Lieve papa, mama en grote zus Natalie, dank voor jullie belangstelling en steun al die jaren! Tevens in de laatste jaren voor het geloven, en meermaals de overtuiging uitspreken, dat ik mijn proefschrift af zou krijgen. Ook mijn schoonfamilie dank ik voor al die jaren steun.

Lieve (Dr.) Suzanne, jij weet als geen ander hoe het is om aan een proefschrift te werken en te promoveren. Jij hebt mij de laatste jaren scherp gehouden door geregeld mijn klankbord te zijn. Ook gaf jij mij vaak ideeën om zaken omtrent het promoveren efficiënter aan te pakken. Je hebt me vaak genoeg 's avonds aan de keukentafel achter mijn laptop zien tikken, dat zal nu voorgoed voorbij zijn. Tijdens mijn promotietraject zijn we getrouwd en hebben we ook een mooi gezin gekregen. Drie prachtige dochters die de broodnodige afleiding hebben gegeven om op het volgende moment weer hernieuwd verder te gaan. Heel veel dank voor je steun al die jaren! Het is me gelukt om het proefschrift af te krijgen en ben dan ook erg blij dat wij deze dag, na die van jou, samen mogen meemaken!

9 789465 152097 >

Radboud Universiteit