Summary
Summary

The goal of this thesis was to develop and validate objective speech analysis techniques to assess speech quality of head and neck cancer (HNC) patients. Three pilot studies (chapters 2 – 4) and a validation study (chapter 5) of objective speech quality analyses were performed for development of an objective speech assessment protocol. Various aspects of speech sounds such as duration, spectral information, and feature analysis were investigated by acoustic-phonetic analyses and by using artificial neural network (ANN) techniques. In the pilot and validation studies we investigated speech sounds that are known to be difficult to produce for HNC patients. The speech feature voicing is among the most seriously affected speech characteristics of HNC patients and is a characteristic of voice-onset-time in stop consonants such as /p/, /t/ and /b/, /d/. Insufficient velar closure of the nasopharynx often results in hypernasality and in distorted velar consonants as /k/ and /x/. Further exploration of the artificial neural network as speech quality assessment technique is described in chapter 6. Finally, associations between voice quality and swallowing function were investigated (chapter 7). The results of this thesis contribute to the development of a multidimensional speech quality assessment tool.

Chapter 2 describes a pilot study on the role of objective acoustic-phonetic analyses in a multidimensional speech assessment protocol. Speech recordings of 51 patients treated for HNC and of 18 control speakers were subjectively evaluated by trained raters (speech therapists) regarding intelligibility, hypernasality, articulation and by patients themselves using the speech subscale of a questionnaire (European Organisation for Research and Treatment of Cancer (EORTC) QLQ-H&N35). Formant values of the vowels /a/, /i/, and /u/, size of the vowel space, and timing of air pressure release of /k/ and spectral slope of /x/ were measured. Values of the first and second formant are multiples of the frequency of the fundamental frequency (F0) and are of importance in identifying vowels. The value of the first formant is determined by vertical tongue position; the value of the second formant is determined by the horizontal tongue position. The spectral slope is perceived as loudness. A qualitatively good voice is characterized by...
Summary

strongly produced high frequencies and a gradual decrease of the spectral slope. A weaker voice has a steep decrease of the spectral slope. Treatment of HNC may result in less effective coordination of muscles and deviant anatomy and physiology of the oral-, nasal- and hypopharynx, resulting in deteriorated speech quality. Results of this study showed that intelligibility, hypernasality and articulation were best predicted by the vowel space and pressure release of /k/. Results of this study revealed that intelligibility, hypernasality and articulation were predicted best by the vowel space and pressure release of /k/. \( R^2 \) values varied from 45% to 74%. Objective acoustic-phonetic analyses distinguished between patients and controls: patients having a higher F1 of /i/ and a lower F2 of /i/ than controls. Within patients, pressure release of /k/ and spectral slope of /x/ differentiated patients regarding tumour site and stage: patients with smaller tumours have a longer pressure release of /k/ compared to patients with a larger tumour. Patients with an oropharyngeal tumour have a steeper spectral slope of /x/ than patients with an oral tumour. Objective speech parameters were not significantly related to speech problems as reported by patients. It was concluded that objective acoustic-phonetic analysis of speech of patients is feasible and contributes to further development of a speech assessment protocol. Further investigation is needed to obtain more insight into acoustic-phonetic analysis in combination with other speech sounds that are oftentimes distorted in patients treated for HNC such as hypernasality.

In chapter 3 the possibilities of a second objective method were described in a pilot study towards nasality. From previously performed research it is known that hypernasality is common among patients treated for HNC. They oftentimes experience incomplete closure of the velum causing airflow into the nasal cavity resulting in hypernasal speech. Analysis of the articulatory feature nasalance was performed by an Artificial Neural Network (ANN). ANN automatically calculates the amount of nasalance per time frame of .01 seconds. The articulatory feature nasalance was measured on the vowels /a/, /i/ and /u/ and on the entire stretch of speech. Speech recordings of the same cohort of 51 patients and 18 control speakers as in the previous study were subjectively evaluated regarding intelligibility, hypernasality, articulation and patients completed the EORTC QLQ-HN35 including the Speech
subscale. Results indicated that the feature nasalance as measured by ANN on /i/ and /a/ distinguishes between patients and controls. Within the patient group regarding tumour subsite or stage, no differences in nasalance were found. Nasalance in the vowels /a/ and /i/ predicted best intelligibility ($R^2 = 21.3\%$), while nasalance in the vowel /a/ predicted best articulation ($R^2 = 48.7\%$) and nasalance in the vowels /i/ and /u/ predicted best hypernasality ($R^2 = 24.9\%$). It was concluded that nasalance as assessed by ANN contributes moderately to the speech evaluation by trained raters. Further research with larger study samples and other speech features is needed.

In chapter 4 acoustic-phonetic analyses and the automatic feature detection methods by Artificial Neural Networks (ANNs) were used to analyze the quality of stop consonants (in Dutch /b/, /d/, /p/, /t/). The VOT in stop consonants distinguishes voiced and voiceless stops. Patients treated for HNC may have difficulty with adequate coordination of motor function of articulatory speech structures and vocal fold vibration. Building up oral pressure necessary for stop consonants in combination and synchronously with ceasing vocal fold vibration in case of the voiceless stop consonants may be especially problematic. For patients it seems problematic to quickly stop the activity of the glottis so that no voicing is produced. Because this action is often difficult to perform for HNC patients it is therefore hypothesized that the duration of VOT preceding the burst in voiced stops in patients is longer compared to controls and that the silence period preceding the burst in voiceless stops show more voicing in patients compared to controls. In the present study, stop consonants /p, t, b, d/ were extracted from speech samples of the same cohort of 51 patients and 18 controls as was used in chapter 2–3. Acoustic-phonetic analyses were performed to investigate the duration of VOT and of the burst. The amount of the articulatory feature ‘voicing’ in VOT and in the burst was measured using ANN. Results revealed that objective acoustic-phonetic analysis and feature ‘voicing’ analysis for /b, d, p/ distinguish between patients and controls. Within patients, /t, d/ distinguish for tumour location and tumour stage: patients with larger tumours had significantly less voicing during VOT compared to patients with smaller tumours and during the burst in the voiced consonant /d/. Regarding tumour location, patients with a tumour originating in the oral cavity had a shorter burst in the
voiceless consonant /t/ compared to patients treated for oropharyngeal cancer. Measurements of the phonological feature voicing in almost all consonants were significantly correlated with articulation and intelligibility, but not with self-evaluations of speech problems in daily life (EORTC QLQ-HN35 Speech Subscale). It was concluded that objective acoustic–phonetic and feature analyses of stop consonants are feasible and contribute to further development of a multidimensional speech quality assessment protocol.

Chapter 5 describes an external validation study of the speech analyses techniques previously investigated in the pilot studies (chapters 2–4). In these previously performed studies we tested the objective acoustic–phonetic and ANN analyses separately (chapters 2 and 3) and then we tested these two objective methods combined onto a selection of speech sounds (chapter 4). This study is aimed at multivariate validation of these objective speech analyses methods together with all previously used speech sounds onto the previously used patient cohort of 51 patients six months after treatment as well as onto a new patient cohort (external validation). This second patient cohort was composed of 64 patients, six months to nine years after treatment for HNC. Speech quality was subjectively evaluated for intelligibility, articulation, hypernasality and by self-evaluations of patients (the speech–subscala of the European Organisation for Research and Treatment of Cancer (EORTC) QLQ–H&N35). Acoustic–phonetic analysis were performed on vowels /a, i, u/, stop consonants /k, p, b, d, t/ and fricative /x/. ANN analysis of the feature ‘nasalance’ was performed on vowels /a, i, u/ and the entire stretch of running speech; ANN analysis of the feature ‘voicing’ was performed on consonants /p, b, d, t/.

In patient cohort 1 subjective intelligibility was predicted by acoustic–phonetic analysis of /p/ and vowel space and by ANN analysis of /d/. Articulation was predicted by acoustic–phonetic analysis of vowel space and by ANN analysis of the feature ‘voicing’ of /b/. Hypernasality was predicted by acoustic–phonetic analysis of /a/, /x/ and /b/. Self-evaluations by patients were predicted by acoustic–phonetic analysis of /i/ and /k/ and by ANN analysis of /p/. The amount of explained variance varied between moderate (52.0% for hypernasality, 37.7% for intelligibility and 36.2% for articulation) to poor (21.1% for self-evaluations by patients).
In cohort 2 intelligibility was predicted by acoustic–phonetic analysis of /a/, /i/ and /x/. Articulation was predicted by acoustic–phonetic analysis of vowel space and by ANN analysis of the feature ‘voicing’ of /p/. Hypernasality was predicted by acoustic–phonetic analysis of /p/ and /t/. Self-evaluations by patients were predicted by acoustic–phonetic analysis of /u/ and /t/ and by ANN analysis of /d/. The amount of explained variance varied between moderate (51.9% for self-evaluations by patients and 41.3% for intelligibility) to poor (21.8% for hypernasality and 20.9% for articulation).

The conclusion is that the combination of previously used analysis techniques and speech material in both cohorts lead to two different predictive models, that both are moderately predictive but are not better that previously tested models.

In chapter 6, further investigation described the possibilities of the Artificial Neural Network (ANN) analyses to investigate the speech quality of patients treated for HNC. In chapters 3, 4 and 5 two specific articulatory features were investigated (‘nasal’ and ‘voicing’) known to be possible features affected in HNC patients. In the present chapter all 28 articulatory features of Dutch were investigated. From the results of the present study was revealed that the features nasalance, voicing and labio-dental appeared to be the most relevant speech features in HNC patients: the speech of patients was significantly different from control speakers for these speech features. The results for nasalance and voicing were in accordance with previously performed studies, as was described in chapters 3 and 4: the difference in voicing between patients and controls speakers was on average 0.16 and the averaged delay between patients and control speakers was more than 0.005 seconds. For the feature nasalance, 8 out of 51 patients appear to have an average nasality that is more than two standard deviations away from the control’s mean. For the feature labio–dental –a feature that is place-bounded– it was seen that four patients had trajectories (from vowel to labiodental) different from the controls. These trajectories were all related to the transitions between vowels and the voiced labio–dental, strongly suggesting that this effect is actually a side–effect related to the feature plosive. That the feature labio–dental is of importance is considered new information and was not known from previously performed research. This result could be possibly explained because
part of the patients underwent (maxillofacial) surgery that was of influence on production of labiodentals speech sounds. Oppositely of the previously performed pilot studies, in the present study no distinction was found between patients and control speakers concerning the speech feature velar (as measured by ANN), while in the velar speech sound \(/k/\) as measured by acoustic–phonetic analysis a difference was found between patients and control speakers (chapter 2). This difference could be explained because ANN calculates the amount of the speech feature velar in running speech, while the acoustic–phonetic method measures the duration of air pressure release as a percentage of the duration of the specific speech sound \(/k/\).

Patients treated for HNC oftentimes experience, next to speech difficulty, difficulty with swallowing. Because in literature a relation between swallowing problems and voice quality was assumed, in chapter 7 a study is described investigating a possible association between voice– and swallowing parameters in patients treated for HNC. Acoustic variables of voice that were measured were fundamental frequency (F0), jitter (percentage), shimmer (percentage), harmonics–to–noise ratio (HNR) and intensity. Jitter is the temporal deviation of cycles produced by the vocal folds (perturbation or disturbance in F0). A high jitter means large frequency perturbation of consecutive cycles in frequency. Shimmer means a disturbance of the amplitude cycles. HNR represents the harmonicity of the wave–like shape of consecutive voicing cycles and describes the amount of harmonicity relative to the amount of white noise in the signal. Vocal intensity was measured in decibel (dB) and represents the sound pressure level. These acoustic parameters were measured in the vowels \(/a/\), \(/i/\) and \(/u/\) and compared to swallowing function parameters as assessed via videofluoroscopy (oral, oropharyngeal and total transit time, estimated percent of oral, oropharyngeal, and total residue, oropharyngeal swallowing efficiency (OPSE) and the Penetration–Aspiration (PA–) scale). Results revealed that intensity in all three vowels \(/a/\), \(/i/\) and \(/u/\) was significantly associated with OPSE and the score on the PA–scale: a worse swallowing function is correlated with louder voice. A possible explanation may be found in overcompensation by increased laryngeal muscular strength resulting in increased intensity and pitch during phonation. However, more research is needed to examine this explanation.
In the general discussion (chapter 8) of this dissertation the main findings, methodological considerations and clinical implications are described, followed by recommendations for future research.

The main goal of the present research was to develop and validate objective speech analyses techniques to evaluate speech quality among patients treated for HNC. The applied methods (acoustic–phonetic and ANN analyses), as well as a variety of phonemes contributed to this goal. However, correlations with subjective assessments by listeners or by patients themselves were limited. There is a number of remarks to make concerning the studies. Speech of relatively small cohorts were used (51 patients (cohort 1), 64 patients (cohort 2) and 18 control speakers). Considerations concerning speech material include reading aloud text of which reading skills could have influenced speech production. A drawback of running speech is the presence of coarticulation and assimilation of speech sounds in which neighboring speech sounds influence the target phoneme. Concerning the two objective measuring methods, the used ANN is of relatively simple origin and was trained on speech of only two speakers. In the future this technique possibly may be improved by using a larger amount of test speakers.

For clinical application the further development of ANN seems to be a better choice than the further development of acoustic–phonetic analyses, although segmenting of target sounds from running speech can be performed automatically through ‘forced alignment’. In forced alignment the signal is lined up onto a sequence of acoustic models that were trained preceding the alignment. In a successful alignment it is expected that 80% of all found phone boundaries are located within a boundary of 20 milliseconds of human-annotated-boundaries. For development of an application that objectively measures voice intensity as an indication of swallowing problems in the oropharyngeal stage, more research is needed to confirm our results and to establish a threshold as a criterion for referral to the clinic setting for further swallowing assessment. Also more research is needed to investigate the (causal) relation between louder voice and worse swallowing.

Easily accessible tools for screening of speech–, voice– and swallowing problems are relevant for clinical practice. Possible applications in the
future are the development of a speech test through the telephone. Patients record speech through the telephone where after the speech recording immediately and automatically is processed by, p.e., an Artificial Neural Network. However, from the present research is known that further pilot studies are needed for validation of objective speech analysis methods. For further research it was advised to use more speech material of larger bodies of speakers –both patients and control speakers– and to take into account the differences in speaker style and demographic and clinical variables such as tumour location, tumour stage and treatment modality. The final conclusion of this dissertation is that objective analysis of speech of patients treated for HNC through acoustic–phonetic analyses and Artificial Neural Network analyses is feasible and valid. Using these findings, medical sciences and speech technology can perform further research that finally may lead to a multidimensional speech evaluation protocol that is usable in clinical practice.