Irma M. Verdonck-de Leeuw

Voice characteristics following radiotherapy: the development of a protocol

IFOTT

voice quality

vocal function

vocal performance

STUDIES IN LANGUAGE AND LANGUAGE USE

33

Irma M. Verdonck-de Leeuw

Voice characteristics following radiotherapy: the development of a protocol

The series 'Studies in Language and Language Use' comprises doctoral dissertations by postgraduate students and monographs and collections produced by researchers of IFOTT.
The theoretical orientation in terms of which IFOTT approaches its research domain can be characterised under the broad term 'functional'. A functional approach to language and language use implies that the study of language must take account of its primary function as a medium of human communication.
The IFOTT research programme consists of the following:

Argumentation Theory
Augmentative and Alternative Communication
Corpus Linguistics
Creole Linguistics
Discourse Analysis
Functional Grammar
Language Acquisition
Language Change
Language Contact
Language Typology
Language Variation
Pathology of Language & Speech
Sign Language
Speech Processing
Speech Production & Perception

Cover: Harry Wildenberg

Voice characteristics following radiotherapy: the development of a protocol

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus prof. dr J.J.M. Franse

ten overstaan van een door het college van dekanen ingestelde commissie

in het openbaar te verdedigen in de Aula der Universiteit

op dinsdag 3 februari 1998 te 15.00 uur

door

Irma Maria Verdonck-de Leeuw

geboren te 's-Gravenhage

FACULTEIT DER GEESTESWETENSCHAPPEN (LETTEREN),
UNIVERSITEIT VAN AMSTERDAM

Promotor:	prof. dr ir L.C.W. Pols
Co-promotores:	prof. dr G.M.M. Bartelink
	dr F.J. Koopmans-van Beinum

Dit onderzoek is uitgevoerd in samenwerking met het
Nederlands Kanker Instituut/Antoni van Leeuwenhoek ziekenhuis
in het kader van het project:
A randomised trial on dose response in radiotherapy for early glottic cancer.

ISBN 90-74698-33-6

STUDIES IN LANGUAGE AND LANGUAGE USE
Uitgave IFOTT Amsterdam
Spuistraat 210 1012 VT Amsterdam

Printed in the Netherlands by ICG Printing B.V. - Dordrecht

Aan mijn ouders

Voorwoord

Het proefschrift is klaar! Het schrijven ervan is vooral een proces geweest van jaren discipline: elke avond achter de computer. Deze discipline was nodig omdat collega's van ander werk en de leden van mijn gezin mij overdag voor zich opeisten. Daar ben ik hen zeer dankbaar voor, daar ik anders mogelijk veranderd zou zijn in een asociaal mens. Toch is dit boekje niet door mij alleen tot stand gekomen. In dit voorwoord wil ik hen bedanken die mij geholpen hebben.

Ten eerste bedank ik de sprekers en luisteraars die aan dit onderzoek hebben meegewerkt. Zonder hen was er niets geweest om over te schrijven.

Graag wil ik de vakgroep Fonetische Wetenschappen van de Universiteit van Amsterdam bedanken voor de gastvrijheid gedurende een groot aantal jaren. Zowel tijdens de studie als tijdens het promotieonderzoek heb ik lief en leed kunnen delen met collega's binnen de vakgroep. Rustige en stormachtige perioden wisselden elkaar af, zowel voor mijzelf als voor de vakgroep. We hebben de periode goed doorstaan: de vakgroep bloeit volop en ik sluit een leerzame periode af met dit proefschrift.

Florien Koopmans-van Beinum en Louis Pols wil ik bedanken voor hun begeleiding door de jaren heen. Ik bedank Jeannette van der Stelt voor het mogen delen van 'haar' kamer. Met haar en met Paul Boersma, Louis ten Bosch, Ton Wempe en Astrid van Wieringen kon ik telkens weer van gedachten wisselen over diverse onderwerpen. Dirk Jaasma dank ik voor zijn nimmer aflatende behulpzaamheid. Willemien Eelzak bedank ik hartelijk voor haar zorgen en vriendschap.

Van het Nederlands Kanker Instituut / Antoni van Leeuwenhoek ziekenhuis wil ik de afdelingen Radiotherapie en Keel- Neus- en Oorheelkunde bedanken voor het vertrouwen in een 'buitenstaander'. Ik denk met veel plezier terug aan de samenwerking met Gertrude Baris; groot was dan ook het verlies toen zij plotseling overleed. Harry Bartelink en Ronald Keus wil ik bedanken voor het feit dat zij haar project hebben voortgezet. Naar aanleiding van de discussies over de volgorde van publiceren heb ik op mijn 'eigen wijze' gehandeld: het proefschrift ligt er dus eerst; de afronding van artikelen zal snel volgen. Ronald Keus wil ik ook bedanken voor het maken van de illustraties in hoofdstuk 2.4. Frans Hilgers bedank ik voor zijn enthousiasme en stimulerende samenwerking. Corina van As dank ik voor de prettige samenwerking, eerst als 'mijn' studente en inmiddels als collega.

Collega's van de afdeling Logopedie en Foniatrie van het Academisch Ziekenhuis van de Vrije Universiteit wil ik bedanken voor hun gastvrijheid; altijd was ik welkom om mijn materiaal te verzamelen, ook als het hen wel eens niet zo goed uitkwam. Met name An Greven wil ik bedanken voor het maken van het grootste deel van de video-opnames.

In het Limburgse dank ik dr. J.M.A. de Jong van het Radiotherapeutisch Instituut Limburg voor zijn inzet. Van het Academisch Ziekenhuis Maastricht dank ik Gerard Vreebrug voor het maken van een deel van de video-opnames en Joyce Uiterwijck-Winkel voor het mij wegwijs maken op de afdeling Stem-Spraak- en Taalpathologie.

Prof. dr. P.H. Dejonkere, prof. dr. H.F. Mahieu, dr. R. van Bezooijen en dr. G. de Krom wil ik bedanken voor de opbouwende kritiek op de leesversie van dit proefschrift. Renée bedank ik ook voor haar leerzame en altijd doeltreffende commentaar op delen van eerdere versies.

Collega's van de afdeling Keel- Neus- en Oorheelkunde van het Academisch Ziekenhuis Leiden wil ik bedanken voor een zeer prettige werksfeer gedurende bijna anderhalf jaar. In deze periode heb ik veel praktische ervaring opgedaan. Robert André dank ik voor zijn commentaar op de 'Summary'. Marika Voerman wil ik bedanken voor haar mentale steun en vriendschap.

Irma de Jong dank ik voor haar commentaar op de Nederlandse samenvatting.

Pap en mam, er zijn niet genoeg woorden om jullie te bedanken, daarom draag ik dit proefschrift aan jullie op.

Frank en Pieter, mijn grote broers, ik ben blij dat jullie als paranimf willen optreden.

Lieve Maarten, alhoewel we beiden zeer energiek zijn en dat ook terug zien in onze kinders, is het nu tijd voor een rustiger periode.

Lieve Juriaan en Casper, ik hoop jullie bij te brengen dat je veel kunt bereiken als je datgene doet wat je het meest aan het hart ligt.

Table of Contents

1

GENERAL INTRODUCTION

Abstract

The aim of the present study was twofold: the development of a protocol to analyse voice characteristics and the use of this protocol to investigate the late effect of different doses of radiotherapy on voice. In this first chapter the need of the development of such a protocol is described from a clinical as well as from a theoretical point of view. An overview of the relevant literature on voice characteristics before and after radiotherapy is given, and the general aim and topics of the present study are described in more detail. Finally, the outline of this thesis is presented.

1.1 Introduction

Approximately 700 new cases of laryngeal cancer are detected every year in the Netherlands, including subglottic, glottic and supraglottic carcinoma (Coebergh et al., 1995). The major symptom of the disease is hoarseness, especially for glottic carcinoma. When diagnosed early, glottic cancer can be treated by (laser) surgery or by radiotherapy, both treatments preserving the voice. Since prognosis is good with cure rates of 70-90%, other criteria than survival statistics alone should be taken into account in the choice of treatment of early glottic carcinoma. Voice quality is one of these criteria. Although most patients do not seem to care about the quality of their voice in the first year after diagnosis and treatment (remarks like "Fortunately, 'it' didn't come back" are often made), after some time, questions are asked whether voice quality will become better or vocal functions like the ability to call or to sing will return.

Although the choice of treatment depends on the preference and/or specialty of the physician (Hirano & Mori, 1996), agreement has been reached, at least in Europe and large parts of the United States, that radiotherapy is the treatment of choice for early glottic carcinoma. Radiotherapy has the advantage of a better voice and fewer complications over laser surgery or hemilaryngectomy (Hirano et al., 1994; Million et al., 1994). However, there is still uncertainty about the optimal radiation dose. Clinical insight in voice quality after radiotherapy indicates that voices are more hoarse after treatment compared to normal voices, probably due to the effects of radiation on normal tissue. As is known from literature, side-effects include early reactions such as mucositis and tissue oedema and late reactions like fibrosis and necrosis (Hill, 1990; Keane, 1994). It can be anticipated that a lower radiation dose will mitigate these side-effects and therefore improve voice quality.

Subjects in the research project which is described in this thesis, are patients who were diagnosed with early glottic cancer and who were treated with radiotherapy. At the Netherlands Cancer Institute/Antoni van Leeuwenhoekhuis, a prospective trial study was set up to investigate the possibility of obtaining the same tumour control rate and improved voice quality with a lower radiation dose than the conventional radiation dose (Baris et al., 1986). An overview of cancer trial studies in general and a description of this particular trial study is given in section 2.5.

The investigation of voice characteristics as part of this trial is the subject of the present thesis. The main aim of the present study is to develop a protocol to analyse voice quality in detail. This protocol is used to investigate late effects of radiotherapy on voice characteristics. In the present thesis, a clear distinction between voice quality and vocal function is made. Voice quality is considered to be a perceptual characteristic of the acoustic voice output. Vocal function describes habitual laryngeal behaviour and capabilities. In order to be able to interpret the results on voice quality, vocal function analyses are taken into account as well.

The clinical need of the development of a protocol concerning detailed analyses of voice quality stems from the fact that contradictory results are found in the literature concerning voices before and after radiotherapy. Few studies have been undertaken, and frequently factors that can influence results on voice quality have not been controlled for, like type of patients (male or female, younger or older), malignancy (size of the tumour), radiation dose (higher or lower), type of speech material (running speech or sustained vowels), voice quality parameters, and so on. An overview of the literature on voice characteristics before and after radiotherapy is given in section 1.2.

From a theoretical point of view, the development of a detailed protocol to measure voice quality can give insight into several aspects. As mentioned, the underlying point of view taken in the present work is that voice quality is a perceptual phenomenon. In other words, we see perceptual analyses as the standard measurements of voice quality. Still, there are uncertainties about the outcome of perceptual analyses. Are there differences between trained and naive raters? Are there differences between perceptual aspects on running speech and on sustained vowels? How do speakers themselves judge their voices compared to other judges? What happens if voices contain aspects like breathiness, roughness and tension at the same time, for instance when patients try to compensate for their breathy voice by phonating with a lot of tension? Can listeners make distinctions between separate voice quality aspects, like breathiness, roughness, tension and so on, or do they just judge on overall deviancy? Acoustical analyses are then expected to give additional insight into perceptual aspects of voice quality. Roughness, for instance is known to be related to perturbation aspects in the signal, whereas breathiness is more related to noise (Hammarberg & Gauffin, 1995). Furthermore, clinical analyses like electroglottography (EGG) and phonetography can help to interpret perceptual and acoustical analyses. Electroglottographic analyses give information about the source of the speech signal, i.e. vocal fold activity, that is more precise than acoustical analyses, which are influenced by the vocal tract. The phonetogram gives information about the frequency and amplitude range of a speaker's voice. When the average fundamental frequency of running speech is investigated, can we explain differences between speakers on the frequency range that they have? These, and other aspects are described in section 1.3 where the general aim and topics of the present study are given. The outline of the thesis is presented in section 1.4.

1.2 Voice characteristics following radiotherapy

Little is known about voice characteristics after radiotherapy. Recent studies report contradictory results ranging from voice improvement to a normal or near-normal level for at least 70% of the patients (Hoyt et al., 1992; Harrison et al., 1990; Miller et al., 1990; Karim et al., 1983; Colton et al., 1978; Stoicheff, 1975; Murry, 1974; Werner-Kukuk, 1968) to abnormal postradiation voices (Dagli et al., 1997; Heeneman et al., 1994; Hirano et al., 1994; McGuirt et al., 1994; Lehman et al., 1986; Stoicheff, 1983). In order to get some insight into these contradictory results, an overview of relevant studies on voice quality before and after radiotherapy is given in table 1.1. Eleven categories are taken into account to compare these studies: sex and age of the speakers, number of speakers in the study, control speakers (were there matched control speakers in the study?), size of malignancy (in terms of T classification, see Chapter 2), radiation dose, time schedule of voice quality analyses (time after radiotherapy), speech material, and voice quality parameters: perceptual, acoustical or clinical analyses. If a study comprised comparison of different types of malignancies or treatments (for instance. glottic versus subglottic carcinoma, or laser versus radiotherapy), only the data concerning early glottic carcinoma treated by means of radiotherapy were taken into account.

Table 1.1 shows that from these 19 studies on voice characteristics before and after radiotherapy, 11 studies were on male voices and 7 on male and female voices; in one study, sex was not indicated. It is known that laryngeal cancer occurs mainly in older men (Coebergh et al., 1995) and that it may be due to smoking and drinking habits (for further information on the epidemiology of laryngeal cancer, see Chapter 2). It is also known that speech and voice characteristics differ between men and women, which might be explained biologically or socio-culturally (Tielen, 1992; Van Bezooijen, 1995). The most obvious aspect is, of course, fundamental frequency, which may be influenced by socio-cultural preferences, but which is physically determined by the size of the larynx and the length of the vocal folds. Also, voice quality aspects differ between male and female voices: breathiness for instance, is an accepted voice characteristic for female voices (Henton & Bladon, 1985; Södersten & Lindestad, 1990). Therefore, when speaker groups are to be compared on voice quality, sex is an important matching factor. Furthermore, age of the speaker is an important matching criterion. Voice characteristics change with ageing of the vocal fold tissue (Mori et al., 1988; Hirano et al., 1989; Murty et al., 1991; Sato & Hirano, 1995). In as far as age was indicated in these studies, it ranged from 40 to 89 years, with a mean age of about 60 years. In 6 of the 19 studies reviewed in table 1.1, control speakers were included. In the remaining 13 studies, the speakers in the study group were compared with themselves on various moments before and after radiotherapy or mean results of the study group were compared with mean normative data from the literature.

Patients involved in the reviewed studies were diagnosed with glottic cancer varying from carcinoma in situ to T4 tumours; they were treated with radiotherapy varying from 50 Gy in 20 fractions to 70 Gy in 35 fractions. Investigation of voice characteristics varied from before and during radiotherapy, to shortly after radiotherapy up to 10 years after radiotherapy. Readers unfamiliar with cancer diagnosis and treatment are referred to Chapter 2 for more information. For the moment it can be stated that investigation of voice characteristics of patients with glottic cancer is dependent on the size of the tumour and thereby on the dose of radiation, and on acute and late effects of radiotherapy reflected in the time schedule. Based on the results from the literature reviewed, it can be stated that voice characteristics are deviant before, during and shortly after radiotherapy, but improve from about 3 months after radiotherapy (Heeneman et al., 1994; Hirano, 1994; Hoyt et al., 1992; Harrison et al, 1990; Miller et al. 1990; Stoicheff et al., 1983; Colton et al., 1978; Murry, 1974; Werner et al., 1968).

The conclusion that voice characteristics remain deviant or that voices become normal again several months after radiotherapy, seems to depend on the choice of speech material and voice parameters. From the 19 studies reviewed, 10 studies involved perceptual analysis of voice quality. Self-ratings were used in two studies, a panel of experienced raters was used in four studies, one experienced rater was used in two studies, and in two studies self-ratings and experienced raters were combined. In the studies that involved experienced raters, voice characteristics were investigated on sustained vowels and on running speech. In 15 of the 19 studies reviewed, various acoustical analyses were used to measure voice changes before and after radiotherapy, mostly on sustained vowels. Clinical analyses were performed in 9 of the 19 studies, like evaluations of stroboscopic recordings, phonetogram, maximum phonation time and so on.

Because so many parameters are used in the studies reviewed, only the results of the studies that involved control speakers, which is an element that we consider to be central, are discussed in the paragraph below, in order to get some insight in voice characteristics several months after radiation (Dagli et al., 1997; Lehman et al., 1986; Stoicheff et al., 1983; Colton et al., 1978). Colton et al. found that voices of 5 patients 1 month after radiation had steeper spectral slopes due to lower spectral levels in the high frequency region (above 5 kHz) compared to 12 control voices, but were in the normal range 13 months after radiation. Stoicheff et al. showed that voices one year after radiotherapy were rated less deviant than voices before radiation, but that their dysphonia was still significantly higher than that of the controls.

Table 1.1. Overview of studies on voice quality before and after radiotherapy. To compare these studies, eleven factors are reviewed. The first seven factors are given here: sex, age (range or mean (m) in years) and number (n) of speakers, involvement of matched control speakers, size of the tumour (T-stage), radiation dose (Gy/fractions), time schedule of the voice analyses (prior, during, and/or after radiotherapy). In a few studies no information was available on some factors (n.i.). This table is continued on the next page with the next four factors.

study	sex	age	n	control	tumour size	radiation schedule	time schedule
Werner, 1968	male	77	1	no	n.i.	62/21	prior, during, after: 4 mths
Murry, 1974	male	49	1	no	T1	58/29	prior, during, after: 2 mths
Stoicheff, 1975 part I	male/ female	59 (m) 52 (m)	203 24	no	n.i.	n.i.	after
Stoicheff, 1975 part II	male	50-70	22	29	n.i.	n.i.	n.i.
Colton, 1978	male	65 (m)	5	12	T1	60-66/30-33	prior, during, after: 1-13 mths
Karim, 1983	n.i.	n.i.	110	no	T1, T2, T3, T4	n.i.	after: > 2 yr.
Stoicheff, 1983	male	30-89	46	58	n.i.	50/20	prior, after: 1 yr.
Lehman, 1986	male	55-80	20	30	T1	66/33	after: 1-7 yr.
Van Wijng. 1988	male	60 (m)	23	no	T1	n.i.	after: 7-117 mths,
Harrison, 1990	male	45-84	18	no	T1, T2	66/33 66-70/ 33	prior, during, after: 1,2,3,6 and 9 mths
Feijoo, 1990	male/ female	40-87	56 1	64	T1, T2 T3, T4		prior
Miller, 1990	male	45-84	20	no	T1, T2	66/33 66-70/ 33-35	prior, during, after: 1,2,3,6 and 9 mths
Hoyt, 1992	male	n.i.	10	no	T1, T2	65/37	prior, after: 6 mths
Ott, 1992	male/ female	64 (m)	13	no	T1, T2	n.i.	n.i.
Benninger, 1994	male/ female	43-81	51	no	T1, T2	60-70 /6-8 wk	after: >2 yr.
Heeneman, 1994	male female	37-85 45-75	37 8	no no	T1a	60/n.i.	prior, after: 3-6-9-12 mths
Hirano, 1994	male/ female	n.i.	24	no	Tis, T1a,b	60/n.i.	prior, after
McGuirt, 1994	male	n.i.	13	no	T1	63/28	after: > 6 mths
Dagli 1997	male female	43-86 57-87	16 4	16 4	T1a,b	57.5-70/ 23-35	after: 1-13 yr.
Present study	male	47-81	60	20	T1a,b	66/33, 60/30, 60/25	prior, after: 6 mths, 2 yr. to 10 yr.

Table 1.1 (continued). Overview of relevant studies for the last four factors: speech material (vowels, read-aloud sentences or text (longer than 30 seconds)), perceptual analyses (self-ratings by the speakers and/or ratings by experienced raters like speech pathologists, physicians (SP)), acoustical analyses (F0-measurements, perturbation (jitter, shimmer), noise (SNR, NNE), percentage voicing) and clinical analyses (stroboscopic evaluation, phonetogram, electroglottogram (EGG), phonation quotient (PQ), maximum phonation time (MPT), etc.). In a few studies factors were not indicated (n.i.).

study	material	perceptual analyses	acoustical analyses	clinical analyses
Werner, 1968	vowels	none	spectral analyses	cinematography, airflow rate
Murry, 1974	vowels	none	mean F0, F0 range	flow rate, intraoral pres.
Stoicheff, 1975 part I	none	self-ratings	none	none
Stoicheff, 1975 part II	reading	none	mean F0	none
Colton, 1978	sentence	none	LTAS	none
Karim, 1983	n.i.	1 SP on excellent -- unsatisfactory	none	none
Stoicheff, 1983	sentence	8 SP on dysphonia, rough, breathy, hoarse, strained	none	none
Lehman, 1986	vowels, sentence	self-ratings	jitter, shimmer, SNR	subgl. press., MPT, strob., F0-range, airflow
Van Wijng. 1988	none	none	none	phon.gram, strob. phon. flow, MPT
Feijoo, 1990	vowel	grade, breathiness	jitter, shimmer, SNR, cepstr., spectr. dist.	none
Harrison, 1990	/a/	none	%voicing, Breath. Index, Strain Index	none
Miller, 1990	/a/	none	% voicing	none
Hoyt, 1992	/a/, wordlist	1 SP % intelligibility	mean F0, % voicing, jitter	none
Ott, 1992	30 s. speech	7 SP on hoarseness	SNR	stroboscopy
Benninger, 1994	vowels	self-ratings and 1 SP on good /poor	mean F0, jitter, shimm., SNR (n=5)	none
Heeneman, 1994	vowels, sentence	10 SP on roughness	modal F0, F0 range, jitter, shimm., SNR	stroboscopy, MPT
Hirano, 1994	n.i.	none	F0 /SPL range, SNR jitter, shim. (n=6)	MPT, airflow rate
McGuirt, 1994	vowels, sentence	average rating of patients and SP's on dysphonia	mean F0, jitter, intensity, spectral distortion, % voicing	airway resist., stroboscopy, MPT
Dagli, 1997	vowels	none	mean F0, jitter, shimmer, NNE	phonetogram, MPT, rate
Present study	vowels, read text	3 trained, 20 naive raters , self-ratings	mean F0, F0 range, HNR, jitter, shimm.	phon.gram, MPT, PQ, strob., EGG

The dysphonia was characterised as breathy and strained before radiotherapy, but tends to become rough after radiation. Lehman et al. showed that voices of patients one to seven years after radiation had worse jitter, shimmer and signal-to-noise ratio values than control voices. Results from clinical analyses indicated a poor vibratory pattern of patients compared to controls. Also, Dagli et al. found that male voices one to 10 years after radiotherapy had worse jitter, mean F0, maximum intensity and intensity range values compared to control voices; they did not find an effect of stage of the tumour (T1a compared to T1b). Although in the study by Benninger et al., no control speakers were involved, their findings are of interest since they found that voice changes may be present after radiotherapy in those patients with associated risk factors like smoking after treatment, stripping or excision rather than biopsy for initial diagnosis, and complications of radiotherapy (like oedema), but that voices are normal in those patients without these risk factors.

Although it is hard to compare results from the 19 studies reviewed, it can be concluded that an acute effect of radiation on voice characteristics has been shown but that late effects are still obscure. In order to describe late effects of different doses of radiotherapy, the present study comprises perceptual, acoustical, *and* clinical analyses of voice quality and vocal function of patients diagnosed with early glottic cancer and of matched control speakers. In the next section, the general aim and topics of the present study are described in more detail.

1.3 General aim and topics of the study

As mentioned in the introduction, the general aim of the present study was twofold: the development of a protocol to analyse voice characteristics of patients diagnosed with early glottic cancer, and the use of this protocol to describe late effects of different doses of radiotherapy on voice characteristics. The need of a thorough description of voice characteristics following radiotherapy has become clear in the previous section. In the present study, voice changes at various moments after radiotherapy ranging from 6 months to 10 years will be described, compared to voices of the patients before radiotherapy and compared to normal voices. The grouping of patients in these time stages after radiotherapy is also used to develop the protocol to analyse voice characteristics, by determining which voice parameters can differentiate speaker groups best.

The starting point of investigation in the present study is the perceptual analysis of voice quality. Trained and naive raters were asked to judge voice quality. The task of the trained raters was to provide an analytic and precise description of the voice quality. Descriptions by naive raters were used to find out how 'ordinary' people evaluate the voice quality. Next to the trained and naive raters, the speakers themselves and their partners were asked to evaluate the voice of the speaker. The purpose of these self-ratings was to get some insight in quality of life aspects related to voice characteristics.

The trained and untrained raters were asked to evaluate voice quality of read-aloud text and of sustained /a/ produced by the speakers. What are the important cues for listeners: the stable voiced portions of vowels or the more dynamic transients or voice onsets in the signal? De Krom (1994) found already that reliability of perceptual ratings can by improved by adding the onset of vowels. And what about the speakers? Is it easier for voice patients to produce a "normal" sustained vowel, while their running speech is deviant? Or is it just the other way around and can they compensate more easily for their poor voice quality during running speech? Analyses of sustained vowels are common practice in clinical settings (for instance voice range profile, phonation flow); sustained vowels are also more suitable for acoustical analyses, since most techniques require more or less stable signals. Therefore, sustained vowels are included as speech material in our project. In order to assess the practical relevance of voice changes in the patients' home environment, fragments of running speech are used as well, because these are more representative of conversational speech.

Perceptual measures are probably superior to acoustic analysis methods, at least for perturbation-based measures (Rabinov et al., 1995). Still, acoustical measurements can help us to obtain more insight in the perception and production process of voice quality. Furthermore, in the present study, an attempt is made to determine whether recent technological development has improved perceptual-acoustical correlations, and to determine whether acoustical analyses are useful for clinical practice. If so, this would mean that voice quality can be objectively investigated in a relative easy and quick way by means of acoustical analyses rather than by the time-consuming perceptual analyses. For clinical purposes, this would be an important advantage.

Once decisions have been made on the choice of analysis methods of voice quality, the main clinical question of this thesis is addressed: what is the effect of radiotherapy on voice quality? The effect of different radiation doses is investigated together with influencing factors on voice quality like stage of early glottic cancer (one or both vocal folds), stripping or biopsying for initial diagnosis, age of the speaker, and smoking after treatment.

In order to get some insight into the vocal function of patients and control speakers, clinical measurements were taken into account. Stroboscopic video-recordings were made in order to evaluate vocal fold vibration and closure directly. Phonetograms were made in order to determine pitch and amplitude range of the voices. Aerodynamic measures were applied in order to determine glottal efficiency. Electroglottographic recordings were made of the same speech material used for perceptual and acoustical analyses in order to measure vocal fold activity directly. Finally, questionnaires related to voice characteristics were presented to the speakers in this study and to their partners, in order to get some insight in the effect of possible voice changes on normal daily and social life situations.

1.4 Outline of the thesis

In order to understand vocal fold changes due to cancer and cancer treatment, a brief description of the anatomy and physiology of normal vocal folds is presented in Chapter 2, followed by the epidemiology, and classification of cancer of the larynx in general, and the definition of early glottic cancer in particular. An overview of treatments is presented and radiotherapy as treatment of the patients in this research project is described in more detail. In Chapter 3, background information of patients and control speakers is presented. Chapter 4 presents results from the perceptual ratings by trained and naive raters and from ratings by the speakers themselves and their partners. Relations between the ratings by various rating groups are discussed as are relations between ratings on the two types of speech material: running speech and vowels. In Chapter 5 perceptual, acoustical, and electroglottographic pitch measures are compared. In Chapter 6, the results from the perceptual analyses as described in Chapter 4 are compared with the results from acoustical analyses. Two speech processing systems are used: a commercially available system "MDVP" developed by Kay Elemetrics, and PRAAT, developed by Boersma at the Institute of Phonetic Sciences Amsterdam. Chapter 7 comprises the results on vocal function compared to voice quality; also, the influence of some factors on voice characteristics is investigated, in order to answer the main clinical question on voice quality following radiotherapy. In Chapter 8, relations between voice characteristics and daily life situations are explored. Finally, in Chapter 9 a general discussion is given, conclusions of this thesis are presented and recommendations for future research are described.

2

EARLY GLOTTIC CANCER

Abstract

Prior to the epidemiology and the classification of laryngeal cancer, a brief description of the "normal" anatomy and physiology of the larynx and in particular the vocal folds is given in this chapter. Early glottic cancer is defined and an overview of possible treatments is given. Since patients with early glottic cancer in this research project were treated with radiotherapy, this particular treatment is described in more detail. Finally, the randomised trial study on the effect of two different radiation doses on early glottic cancer is described. The present study, on voice characteristics of patients before and after radiotherapy, is part of this clinical trial.

2.1 Introduction

The main biological functions of the larynx and vocal folds are breathing (open airway), swallowing (reflective closure of the laryngeal entrance), and pressure activities, like coughing, defecating and lifting. The use of the larynx as a communication tool is a secondary function. However, since the present study is on voice characteristics, the description in section 2.2 on the anatomy and physiology of the vocal folds is focused on this function of voice production. In section 2.3 the epidemiology of laryngeal cancer is given, followed by the classification of laryngeal cancer. Since this general classification of laryngeal cancer does not seem to fit to early glottic cancer, this particular tumour is defined in detail, according to current standards. Also, prognosis and possible treatments are described in this section. Because voice characteristics following radiotherapy is the subject of the present study, this particular treatment is presented in more detail in section 2.4. Finally, the study protocol on the effect of two different radiation doses of which the present study is part, is described in section 2.5.

2.2 The vocal folds

2.2.1 Anatomy

The basic source of voice is the respiratory system pushing air out of the lungs. This air goes from the lungs up to the trachea and into the larynx, where it passes the vocal folds. The larynx consists of the thyroid, cricoid and arytenoid cartilages, intrinsic and extrinsic laryngeal muscles and vocal folds. The laryngeal muscles regulate the position, shape and tension of the vocal folds, which determine the mode of vocal fold vibration. In breathing, the vocal folds are apart and the air can pass freely. When the vocal folds are adjusted, the pressure of the airstream will cause them to vibrate and generate the voice sound. The vocal fold vibrations during phonation are not only dependent on laryngeal muscle activities, but also on the mechanical properties of the vocal folds. Detailed information on laryngeal muscles and their relation to vocal fold vibration is not given here, because this information is available in many textbooks on anatomy and physiology of the larynx, such as Hardcastle (1976) and Tortora & Grabowski (1993). An excellent overview of the molecular and cellular structure of the vocal folds, and the mechanical properties of the vocal folds is given by Gray, Hirano & Sato (1993). A part of their insights is reproduced here, because it is expected that the anatomy and thereby the mechanical properties of the vocal folds can be influenced by small tumours or by radiation treatment. The vocal folds are composed of surface tissue and the underlying muscle tissue. The surface tissue consists of the epithelium at the vocal fold edge, and the lamina propria, which is divided into a superficial (pliable tissue, also called Reinke's space), an intermediate (mainly consisting of elastic fibers), and a deep (mainly consisting of collagenous fibers) layer. The basement membrane zone, which structure is complex, anchors the surface tissue of the epithelium to the superficial layer of the lamina propria. The main body of the vocal fold is formed by the vocalis muscle (figure 2.1).

The vibration of the vocal folds is facilitated by the parallel arrangement of the elastic and collagenous fibers of the lamina propria and the muscle fibers of the vocalis muscle into the vocal edge. During phonation, the loose and pliable tissue of the superficial layer of the lamina propria is the most vibrating part. This layered structure of the vocal fold edge varies along the length of the vocal fold, which means that the tissue stiffness changes gradually from the stiff thyroid cartilage at the anterior end to the pliable membranous vocal fold to the stiff arytenoid cartilage at the posterior end.

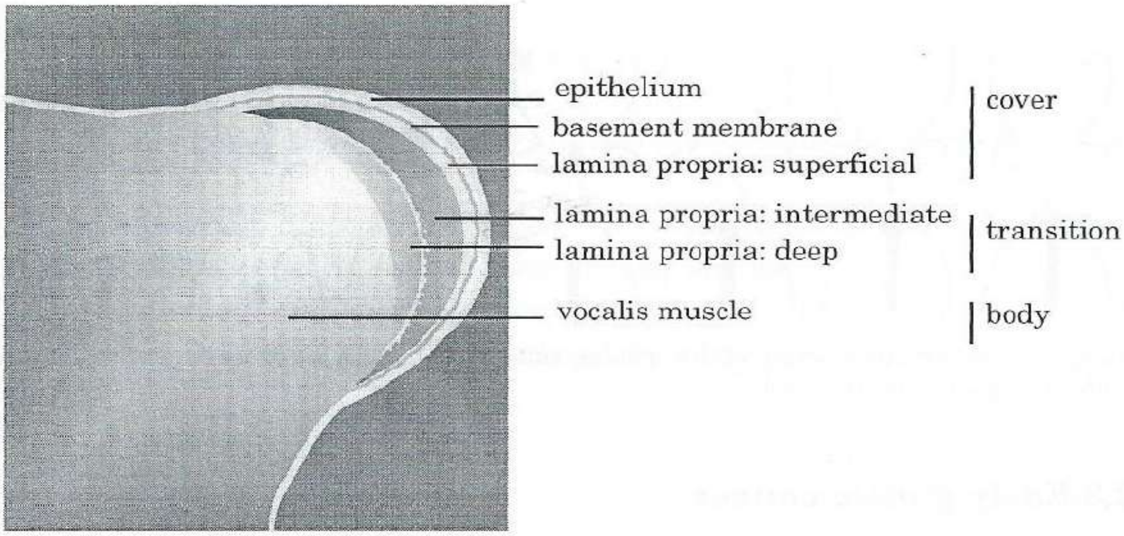

Figure 2.1. Schematic drawing of the layered structure of a vocal fold. Terminilogy after Gray, Hirano & Sato, 1993.

2.2.2 Vocal fold vibration

Vocal fold vibration is a result of interaction between aerodynamic forces and laryngeal muscle forces according to the aerodynamic-myoelastic theory of phonation (Fant, 1960). The vibration cycle begins when subglottal pressure removes the vocal fold resistance (closed phase) by blowing the glottis open. The glottis remains open until the subglottal pressure is reduced because of escaping air (open phase). When supra- and subglottal pressure are equal, vocal fold resistance causes the closure of the vocal folds, the glottis remains closed until subglottal pressure is build up again (closed phase).

A schematic and photographic representation of a vocal fold vibration cycle is given in figure 2.2 and figure 2.3, respectively. The myoelastic part of the theory concerns both muscle activity and mechanical properties of the vocal folds. Muscle activity is necessary to maintain vocal fold resistance: the resistance is dependent on the stiffness (longitudinal tension) and on the mass of the vocal folds (thickness). Furthermore, voice production can be controlled for pitch, loudness and quality. Besides the muscle activity, the mechanical properties of the vocal folds are indispensable for the vibration cycle in creating phase differences. The stiffer underlying parts of the vocal folds are blown apart first, gradually followed by the pliable superficial layer of the lamina propria. Also, in the closing phase, the vocal fold closure occurs at the underlying parts, followed by the mucosal wave of the superficial layers. If this pliable tissue becomes stiff as a result of some pathological causes (e.g. neoplasm, inflammation or scarring) vibratory movements are impaired; and even worse, if the underlying parts are affected, vibratory movement may be absent (figure 2.4).

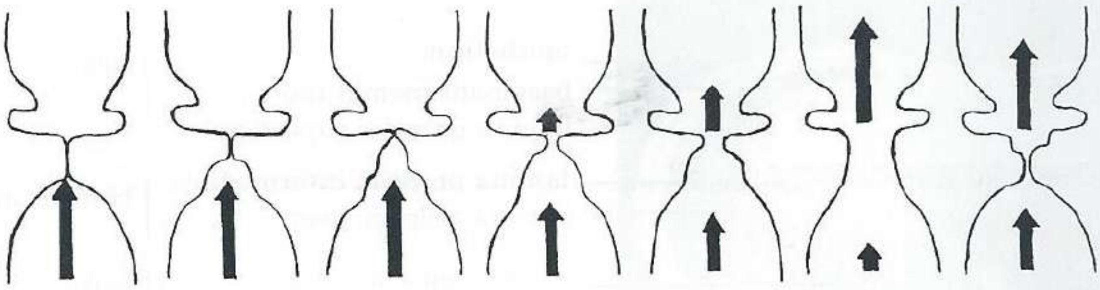

Figure 2.2. Schematic drawing of the opening and closure and mucosal wave of normal vocal folds during one vibration cycle.

2.3 Early glottic cancer

2.3.1 Epidemiology

The development of cancer is a process that occurs over many years (UICC, 1987). There is clear evidence that smoking and alcohol consumption are the major etiologic factors in the development of laryngeal cancer; for combined alcohol and tobacco consumption, the risk ratio increases more in a multiplicative than in a additive manner. Wynder et al. (1976) reported that the risk of developing laryngeal cancer was 7 times greater for persons who smoked more than 35 cigarettes per day compared to non-smokers, and that the risk was 22 times greater for persons who smoked more than 35 cigarettes per day and consumed more than 7 ounces of alcohol per day. But also other environmental factors may be involved in the development of laryngeal cancer,

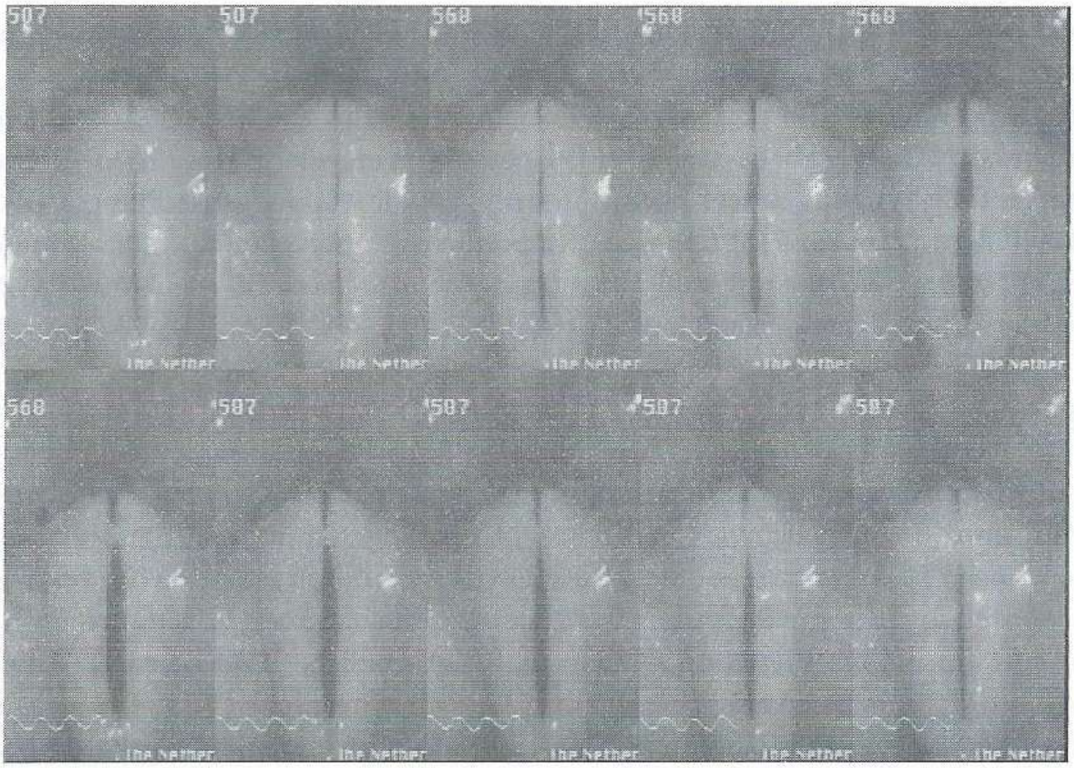

Figure 2.3. Series of photographs of one vibration cycle of normal vocal folds, as observed during video-stroboscopy.

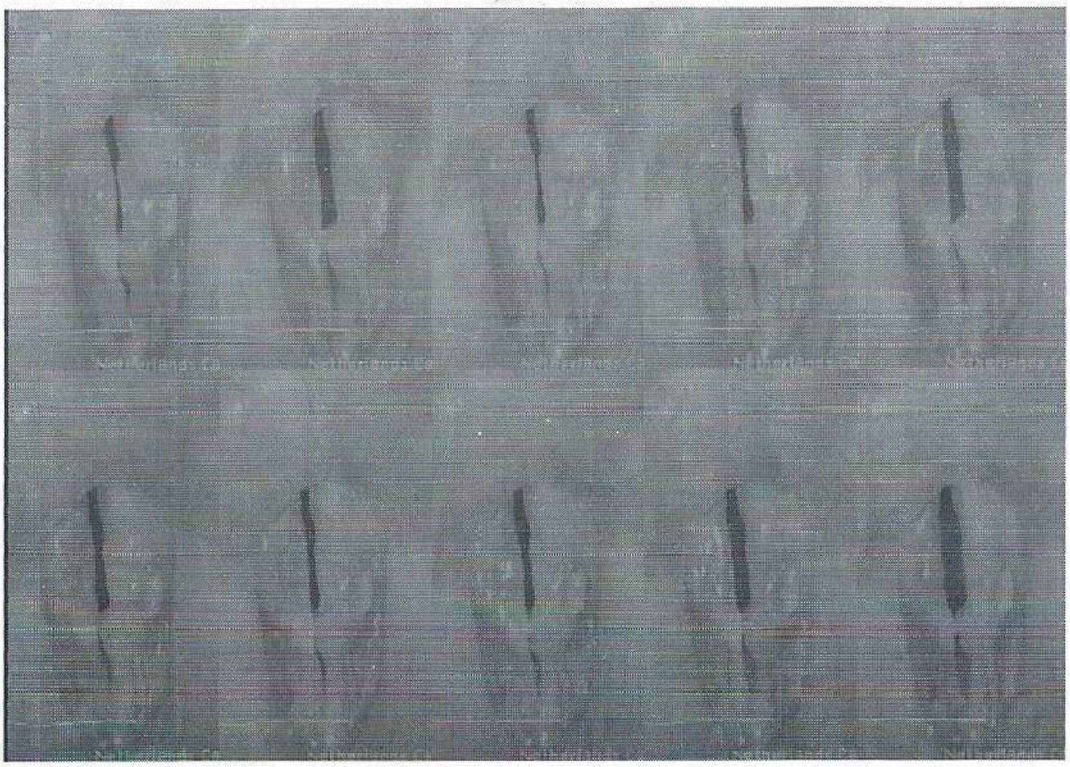

Figure 2.4. Series of photographs of one vibration cycle of vocal folds with a glottic tumour, as observed during video-stroboscopy.

like exposure to radiation, drugs, occupation related factors like asbestos and diesel fumes (Muscat & Wynder, 1991; Maier et al., 1992).

From the results of a project on cancer incidence in the South East of the Netherlands (Coebergh et al., 1995), it has become clear that also in the Netherlands laryngeal cancer occurs mainly in elder males (prevalence of 73 per 100.000), and is rare for females (prevalence of 6 per 100.000). Of all male laryngeal cancer patients, 3% is under the age of 45 years, 35% is between 45 and 60 years, 49% between 60 and 75 years, and 13% is older than 75 years.

2.3.2 TNM-classification of laryngeal cancer

Classification of carcinoma depends on the localisation and the size of the tumour, on the depth of invasion, and on the presence of regional and distant metastases. The most current classification of carcinoma is the TNM-staging as proposed by the International Union Against Cancer (UICC) and the American Joint Committee on Cancer (AJCC) (Ferlito, 1993). In the TNM-classification, the larynx is divided into three anatomical sites: the supraglottis (epiglottis, ventricles, aryepiglottic folds and arytenoids), the glottis (true vocal folds and anterior and posterior commissures), and the subglottis (region below the true vocal folds to the first tracheal ring). The T-division describes the primary tumour, the N-division the metastases to lymph nodes in the neck, and the M-division distant metastases. An overview of the TNM-staging of glottic cancer is given in table 2.1.

2.3.3 Definition of early glottic cancer

Early glottic cancer is often classified as T1aN0M0, indicating a tumour limited to one vocal cord, without regional lymph node metastasis and no distant metastasis. However, there is confusion about this classification of early glottic tumours and some authors have proposed an alternative classification (Karim et al., 1989; Kleinsasser, 1992; Ferlito, 1993). The problem is that the T-classification depends on anatomical regions and that these regions are not accurately defined; furthermore, tumour volume and invasiveness are not taken into account. According to Ferlito et al. (1996) early glottic cancer should be defined as a minimally invasive tumour that does not invade the vocal muscle or cartilage, but is still capable of metastasis. By this definition, carcinoma in situ (a premalignancy) and deeply infiltrating carcinoma are excluded from early glottic carcinoma. The site of the tumour is not of interest. Vocal fold mobility has to be normal (indicating that the vocalis muscle has not been invaded by the tumour). A schematic representation of different types of carcinoma of the vocal fold is given in figure 2.5. In the present thesis, early glottic cancer is interpreted as T1a and T1b tumours.

Table 2.1. TNM-staging of laryngeal glottic carcinoma (UICC, 1987).

stage	description
TX	Tumour that cannot be assessed by rules.
T0	No evidence of primary tumour.
T1S	Carcinoma in situ.
T1	Tumour limited to vocal fold(s) (may involve anterior or posterior commissures) with normal mobility.
T1a	Tumour limited to one vocal fold.
T1b	Tumour involves two vocal folds.
T2	Tumour extends to supraglottis and/or subglottis, and/or with impaired vocal cord mobility.
T3	Tumour limited to the larynx with vocal fold fixation.
T4	Tumour invades through thyroid cartilage and/or extends to other tissues beyond the larynx, e.g. to oropharynx, soft tissues of the neck.

stage	description
NX	Lymph node metastasis that cannot be assessed by rules.
N0	No regional lymph node metastasis.
N1	Metastasis in a single ipsilateral lymph node, 3 cm or less in greatest dimension.
N2	Metastasis in a single ipsilateral lymph node, more than 3 cm but not more than 6 cm in greatest dimension; or in multiple ipsilateral lymph nodes, none more than 6 cm in greatest dimension; or in bilateral or contralateral lymph nodes, none more than 6 cm in greatest dimension.
N3	Metastasis in a lymph node more than 6 cm in greatest dimension.

stage	description
MX	Metastasis that cannot be assessed by rules.
M0	No distant metastasis.
M1	Distant metastasis.

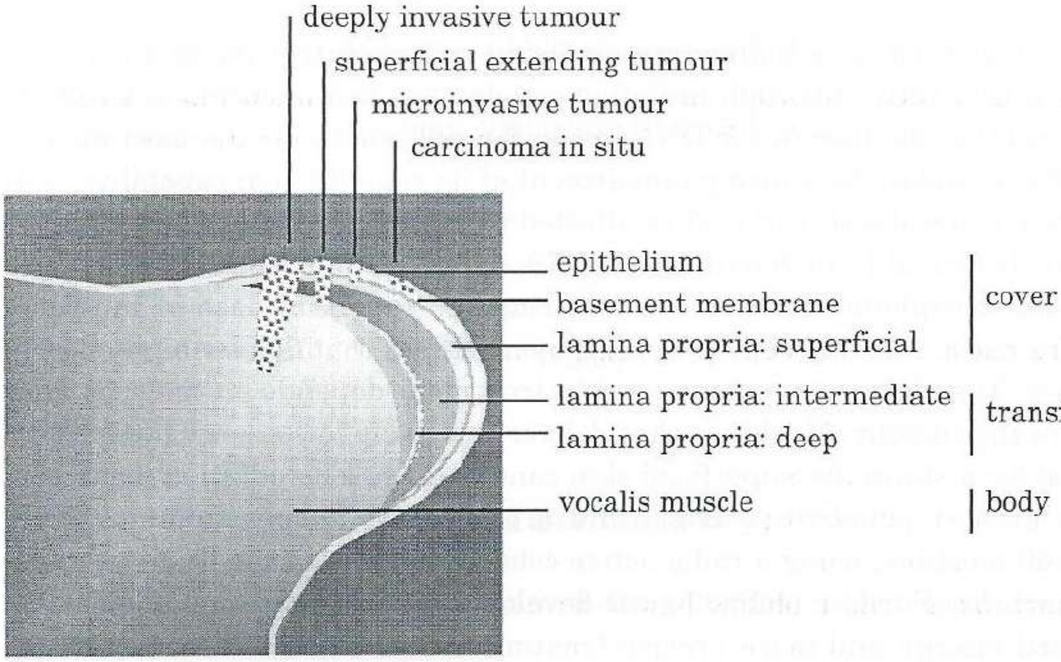

Figure 2.5. Schematic drawing of a vocal fold illustrating various types of carcinoma (terminology after Ferlito et al., 1996).

2.3.4 Prognosis and treatment

The prognosis for early glottic cancer is very good, with cure rates of 75-90%, depending on the site and degree of infiltration. As indicated above, clear risk factors for the development of glottic cancer are smoking and excessive alcohol consumption (Muscat et al., 1991; Maier et al., 1992). If a patient continues to smoke and drink after treatment, cure rates of the initial cancer are diminished, and the risk of a second tumour is increased (Browman et al., 1993). Standard treatment options for small glottic carcinoma are radiotherapy, surgery (mucosal stripping, cordectomy or hemilaryngectomy), or laser excision. Carcinoma in situ is usually managed by conservative surgery like mucosal stripping or superficial laser excision. For the treatment of early glottic cancer, radiotherapy may be selected above surgery in order to preserve the voice, reserving surgery for salvaging failures (Million et al., 1994). Voice quality of patients treated with laser can be as good as post-radiation voices, but only in the case of selected patients with carcinoma in situ or T1 lesions confined to the mobile portion of the vocal fold away from the anterior commissure (Goepfert, 1994; McGuirt et al. 1994). Since the patients in our research project are treated with radiotherapy, this treatment is described in more detail in the next section.

2.4 Radiotherapy

2.4.1 Principles

Radiation therapy uses high-energy ionising radiation (X-rays or γ-rays) to destroy cancer cells. Although radiation can damage any molecule in a cell, it is believed that damage to the DNA inside the cell nucleus is the most important effect of radiation, causing impairment of its reproduction capability and cell death. Normal cells can also be affected by ionising radiation, but usually they are better able to repair their DNA damage than tumour cells; this difference is exploited further by delivering the radiation dose in multiple small fractions. Radiotherapy is applied by machines that deliver high-energy radiation. The choice of machine used in treatment depends on the type and extent of the tumour. Machines that deliver relatively low energy (orthovolt) are used for instance for superficial skin cancer. Megavolt radiation machines have a greater penetrative effect and are used for deeper tumours. The Cobalt-60 machine, using a radio-active cobalt source, was the first megavoltage machine. Further technological development delivered machines with increased energy and more precise treatment beams. Linear accelerators, using megavolt X-rays beams, are now the most commonly used machines.

2.4.2 Simulator and treatment planning

An X-ray machine called a simulator is used to visualise and define the exact treatment area. This is necessary in order to be able to focus the radiation beam to the tumour region and to limit the radiation dose to normal tissues. In order to deliver treatments precisely, ink marks on the skin can be used as reference points. In the case of head and neck malignancies, a plastic shell is made of the patient's head and neck that is marked with the precise region to be irradiated. This shell not only has the advantage that visible marks on the patient's skin can be avoided, but provides also excellent immobilisation in the treatment position that is reproducible every next treatment. An example of a shell is given in figure 2.6. The results from the simulator are also used in treatment planning. Precise information about tumour volume and region are used to define the radiation target. By means of a computer planning system, the dose distribution can be calculated in a central transversal plane through the target volume. Various radiation beams (direction, energy) are used in order to concentrate radiation dose in the target field and to spare normal tissue as much as possible. In early glottic cancer the most common technique used is radiation by two lateral opposed beams. Wedge filters are used to compensate for the difference in diameter of the neck in anterior-posterior direction (figures 2.7 and 2.8). With this technique a homogeneous dose distribution in the larynx is obtained.

Figure 2.6. Photographic example of a plastic immobilisation shell.

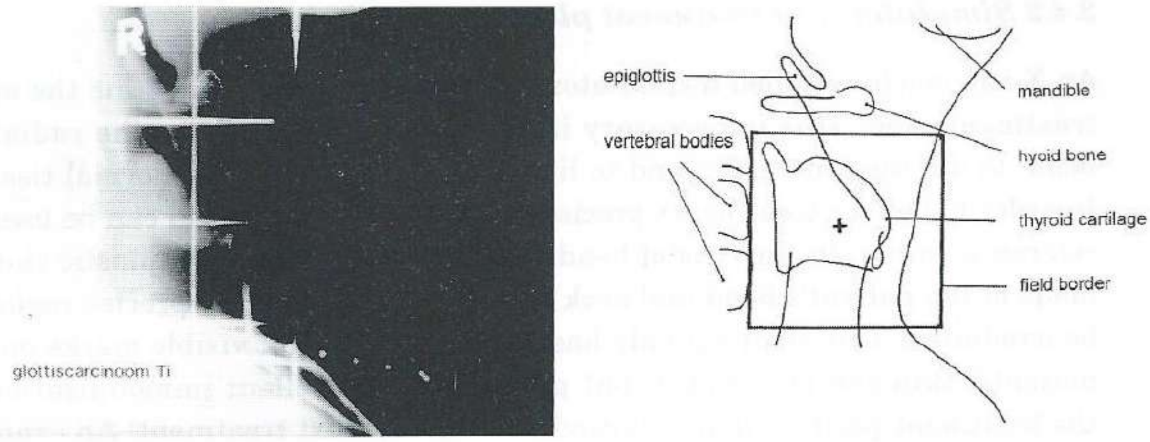

Figure 2.7. Photograph (at the left) and schematic drawing (at the right) of the right side-view of the radiation field for glottic carcinoma.

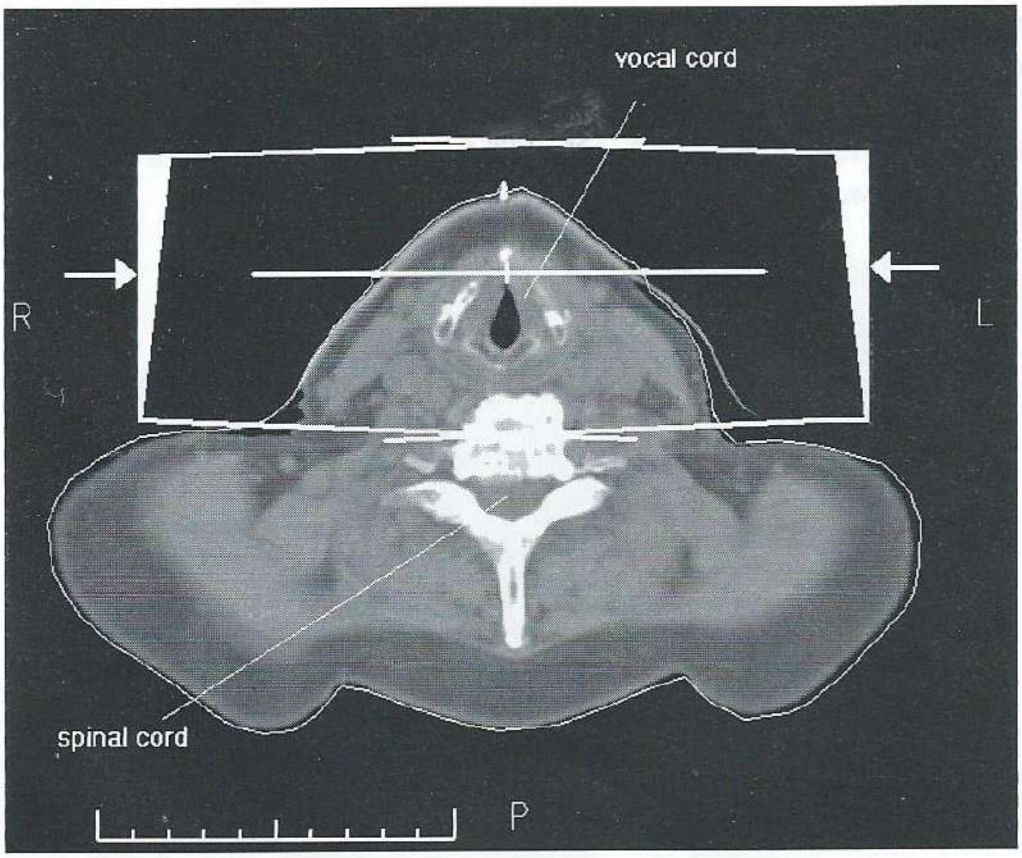

Figure 2.8. Ct-scan of the radiation field for glottic carcinoma.

2.4.3 Radiation dose and fractionation

Radiation dose is expressed in Gray (Gy) or centiGray (cGy). The total dose of radiation to be delivered depends on tumour volume, radio-sensitivity of the tumour, the dose per fraction, and the normal tissue tolerance. For laryngeal carcinoma, total radiation dose varies from 50 Gy to 80 Gy.

To allow for recovery of normal tissue damage, the total dose is delivered in smaller doses, called fractions. Conventional treatments are usually given on a daily basis, 2 Gy per day, five days per week. By the use of fractionation, a tumour-lethal dose can be delivered while minimising the damage to normal tissues. Concerning early glottic carcinoma, there is still uncertainty about the optimal total dose level to be delivered. Also, new schedules of fractionation are investigated: hyperfractionation (using multiple daily fractions in order to reach a higher total dose) and accelerated fractionation schedules (using multiple daily fractions in order to reach the same dose in a shorter overall period). The idea of accelerated fractionation is that repopulation of tumour cells during treatment might be decreased. Results of recent studies reveal that these radiotherapeutic schedules might improve local tumour control, although they can be accompanied by higher complication rates (Horiot et al., 1997). No extensive research on voice characteristics after various radiotherapeutic schedules has been conducted yet.

The purpose of the trial study described in the next section was not to investigate the effect of different fractionation schedules, but rather to investigate differences between the effect of standard total radiation dose delivered to early glottic cancer (66 Gy in 33 fractions, 2 Gy per fraction) compared to lower total radiation dose (60 Gy in 30 fractions, 2 Gy per fraction), in terms of tumour control rate and voice quality.

2.5 Clinical trial

In cancer research, a clinical trial is a study conducted with cancer patients, usually to evaluate a new treatment. Before a new treatment is applied to patients, it is studied extensively in the laboratory. In a clinical trial, a formal study plan, called a "protocol" is designed to answer research questions as well as to safeguard the medical and psychological health of patients.

The aim of the protocol N87RTG entitled "A randomised trial on dose response in radiotherapy of early glottic cancer" started at the Netherlands Cancer Institute/Antoni van Leeuwenhoekhuis in 1988, is to evaluate whether a lower radiation dose (60 Gy in 30 fractions: 5 fractions per week, 2 Gy per fraction) results in improved voice quality with the same local control, compared to the standard high dose regimen (66 Gy in 33 fractions: 5 fractions per week, 2 Gy per fraction).

The eligibility criteria for patients to enter the study are: a) small glottic tumours (T1, T2N0M0) with no bilateral or 3 regions involvement and with no impaired vocal fold mobility, b) WHO (World Health Organization) performance status scale under 2 (indicating that the patient is up for more than 50% of the time and is able to take care of himself), (c) age under 75 years, and d) infiltrative squamous cell carcinoma.

Ineligibility criteria are: a) T2N0M0 glottic cancer with bilateral or 3 regions involvement and with impaired cord mobility, b) N1, N2 or N3 disease, c) carcinoma in situ or variants e.g. adenosquamous, basosquamous or verrucous types, d) previous surgery (except biopsy or stripping), radiotherapy or chemotherapy for the primary tumour, e) hematogenous metastases or f) previous or concurrent cancer at other sites except in situ carcinoma of the uterine cervix, and adequately treated basal or squamous cell carcinoma of the skin (Baris, 1986).

In order to answer the research question of the trial, several aspects of radiotherapy are taken into account. Radiation aspects like field size and type of radiation (Cobalt or photons) are part of the study. Acute effects of radiation like response to radiotherapy of the primary tumour (complete regression or not) and acute complications, like mucosal reaction and skin reaction within 6 weeks after the end of radiation, are investigated. Late effects of radiation (more than 5 months after the start of treatment) are investigated, i.e. local control of the primary tumour and late complications such as mucosal sequelae, skin reactions, larynx oedema, fibrosis or other. Data are collected before, during, at the end of radiotherapy, and 2-3 months, 6 months, 12 months, and 2, 4, 5, 10 and 15 years after radiotherapy.

In order to investigate the effect of radiotherapy on voice quality within this trial, a study was conducted to develop a protocol to analyse voice quality of patients with early glottic cancer recorded before and after radiotherapy and of control speakers. The present thesis is a description of the development of this protocol and of the use of this protocol to investigate the effect of radiotherapy on voice quality of the patients.

3

SPEAKERS

Abstract

In this research project, voice characteristics of 60 male patients with early glottic carcinoma recorded before and after radiotherapy, were compared with 20 control speakers on perceptual, acoustical, and clinical voice measures. In the present chapter, a detailed presentation of the speakers and the research design in terms of grouping of the speakers is described.

3.1 Introduction

In this chapter, the speakers (patients and controls) are presented. Sixty patients were involved in the study, all were treated with radiotherapy for early glottic cancer T1 (UICC, 1987). A longitudinal speaker group was composed as well as separate speaker groups. Because voice characteristics are speaker-dependent, a longitudinal speaker group is preferred over separate speaker groups. However, follow-up of the patients in the longitudinal group for more than 2 years after radiotherapy fell out of the range of the present study, which was conducted as a four-year project. Therefore, five separate speaker groups were introduced as well, to analyse the late effect of treatment on voice characteristics. Finally, a control group of 20 matched speakers was involved. After a detailed presentation of the speakers (patients compared to controls) and of data on diagnosis and treatment of the patients in the next section, the research design in terms of grouping of the speakers is described.

3.2. Speakers

Questions on nine topics were asked to all speakers in order to get overall insight in case histories: consumption of spirits, smoking habits (previously and at the moment of voice measurements), pulmonary diseases, hearing problems, speech therapy, working environment concerning voice (amount of speaking time, noise, or smog), hobbies concerning voice (singing, acting), changes in voice quality during the day, and changes in voice quality due to fatigue, stress, amount of speaking time. Data of patients are given in section 3.2.1; data of controls are given in section 3.2.2.

3.2.1 Patients

All patients were male, the mean age was 67 years; the youngest patient was 47, the oldest 81. Data over all 60 patients on the nine topics as described above are given in table 3.1 and can be summarised as follows:

A minority of the patients were non-drinkers (8%); the others had moderate (between 1-5) consumption (75%) or were heavy drinkers (17%).

Only 2% of the patients never smoked; most of them used to smoke, ranging from less than 10 cigarettes per day (20%), between 10 and 20 cigarettes per day (13%), to over 20 cigarettes per day (48%); 17% of the patients quit smoking more than 15 years ago. At the moment of the voice measurements, 85% of the patients did not smoke.

Most of the patients did not suffer from pulmonary diseases (92%), nor did they have hearing problems (73%). A minority of the patients received speech therapy for their voice problems (11%).

A considerable percentage of the patients (60%) used to speak a lot in the exercise of their duties (43%), in combination with a noisy working environment (15%) or with smog (2%). Most of the patients did not have a "voice hobby" (92%), 4 patients were amateur singers and one an amateur actor.

No daily changes in voice quality were reported by 69% of the patients, while 13% noticed a worse voice in the morning and 19% experienced voice decrease during the day. Finally, 65% of the patients reported voice decrease due to speaking a lot (35%), fatigue (2%), stress or emotions (7%), speaking in a room filled with smoke (2%), or in noise (3%) or a combination (17%).

3.2.2 Controls

The control group was matched with the patients with respect to sex (all male), age (the mean age was 67 years; the youngest speaker was 51, the oldest 81), as well as smoking and drinking habits. Data over all 20 control speakers are given in table 3.1 and can be summarised as follows:

A minority of the controls were non-drinkers (15%); the others had moderate (between 1-5) consumption (75%) or were heavy drinkers (10%).

As most of the patients used to smoke but stopped smoking when the tumour was discovered, the control group was built up of 10 smoking men and 10 non-smoking men. The smoking men smoked less than 10 cigarettes per day (60%), or over 20 cigarettes per day (40%). Of the non-smoking men, 2 men never smoked and 8 men used to smoke but stopped smoking longer than 15 years ago. Of these 8 men, 25% used to smoke less than 10 cigarettes per day, another 25% used to smoke between 10 and 20 cigarettes per day and 50% used to smoke more than 20 cigarettes per day.

Most of the controls did not suffer from pulmonary diseases (90%), nor did they have hearing problems (70%). None of the controls received speech therapy.

A considerable percentage of the controls (50%) used to speak a lot in the exercise of their duties (35%), in combination with a noisy working environment (15%). Most of the controls did not have a "voice hobby" (85%), 2 speakers were amateur singers and one used to talk a lot during meetings.

No daily changes in voice quality were reported by 90% of the controls; the other 10% noticed a worse voice in the morning. Finally, only 15% of the controls reported voice decrease due to speaking a lot (10%), or fatigue (5%).

Table 3.1a. Overview of alcohol consumption in percentages of 60 patients and 20 control speakers.

consumption spirits	patients	controls
none	8	15
sporadic	12	0
1-3/day	53	60
3-5/day	10	15
5-10/day	10	5
over 10/day	7	5

Table 3.1b. Overview of smoking history, previously and at the moment of voice measurements, in percentages of 60 patients and 20 control speakers (10 who never smoked or stopped smoking more than 15 years ago and 10 who still smoke).

smoking history	patients		controls	
	previously n=60	at the moment n=60	non-smokers n=10	smokers n=10
none	2	85	0	0
1-5./day	5	6	0	20
5-10/day	19	3	0	40
10-15/day	13	2	0	0
15-20/day	8	2	0	0
over 20/day	53	2	0	40

Table 3.1c. Overview of the percentages of 60 patients and 20 control speakers concerning pulmonary diseases, hearing problems and received speech therapy.

	patients		controls	
	yes	no	yes	no
pulmonary disease	8	92	10	90
hearing problems	27	73	30	70
speech therapy	11	89	0	100

Table 3.1d. Overview of the voice load during work or hobbies in percentages of 60 patients and 20 control speakers.

voice load	patients	controls
work		
none	17	45
speaking a lot	43	35
smog	2	0
noise	8	5
speaking and noise	15	15
speaking and smog	2	0
noise and smog	13	0
hobby		
none	92	85
singing/acting	8	15

Table 3.1e. Overview of changes in voice quality during the day or under certain circumstances in percentages of 60 patients and 20 control speakers.

voice changes	patients	controls
daily		
none	69	90
morning	13	10
during the day	18	0
circumstances		
none	35	85
speaking a lot	35	10
fatigue	2	5
emotion/stress	7	0
room with smoke	2	0
noise	3	0
combination	16	0

3.2.3 Diagnosis and treatment of patients

All 60 male patients were diagnosed with early glottic T1 cancer. Results for the 60 patients concerning diagnosis and treatment can be summarised as follows:

Of the 60 patients, 65% had a tumour on or originating from the left vocal fold, and 35% on or originating from the right vocal fold. Concerning the stage of the tumour, a division was made in uni- (T1a) or bilateral (T1b) (see figure 3.1).

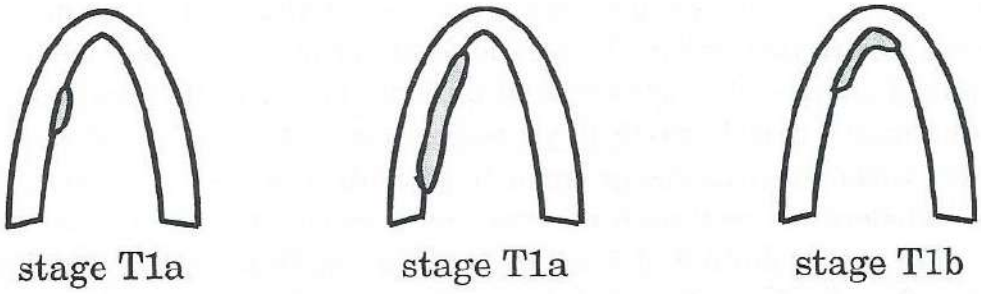

Figure 3.1 Examples of various tumour stages. Stage T1a: tumour limited to a part or stretched over one (whole) vocal fold; stage T1b: tumour extended from one vocal fold over the commissure anterior to the other vocal fold.

Concerning tumours on the left vocal fold (39 patients), 72% was unilateral (T1a) and 28% bilateral (T1b). Of the tumours on the right vocal fold (21 patients), 75% was unilateral and 25% bilateral.

A biopsy for initial diagnosis was taken for 68% of the 60 patients. For 32% of the patients, the vocal folds were stripped and/or a biopsy was taken. Because initial diagnosis took place in different hospitals and up to 10 years ago, no attempt was made to retrieve exact data on vocal fold stripping.

Radiotherapy was given to all 60 patients, varying from 66 Gy in 33 fractions (48% of the patients), 60 Gy in 30 fractions (29% of the patients) and 60 Gy in 25 fractions (23%). These radiation schedules were unknown to the experimenter until much later (see section 3.2.2).

Local recurrence occurred in 3 of the 60 patients (5%). Related to radiotherapy, 2 of these 3 patients received 66 Gy in 33 fractions, and 1 received 60 Gy in 30 fractions. Related to stage of the tumour, 1 of these 3 patients had a unilateral tumour and 2 patients a bilateral tumour.

Table 3.2. Overview of diagnosis and treatment over all 60 patients in percentages on tumour on or originating from the left or right vocal fold, tumour size uni- (T1a) or bilateral (T1b), stripping or biopsying the vocal fold for initial diagnosis, radiation schedule (66Gy in 33 fractions, 60 Gy in 30 fractions, 60 Gy in 25 fractions), and local recurrence (yes or no).

tumour on fold		tumour size		initial surgery		radiation schedule			local recurrence	
left	right	T1a	T1b	strip	biopsy	66/33	60/30	60/25	yes	no
65	35	73	27	68	32	48	29	23	5	95

3.3 Research design

3.3.1 Voices before and after radiotherapy compared to control voices

In order to develop a protocol to analyse voice characteristics, a research design was set up composed of patient groups before and after radiotherapy, and a control group. The goal was to assess which voice parameters can differentiate speaker groups best. The assumption was that a trend can be determined of voice characteristics of patients before radiotherapy and after radiation from 6 months up to 10 years, compared to control speakers.

There was a longitudinal group of 10 patients of whom voice samples were recorded before, as well as 6 months and 2 years after radiation. Furthermore, data were collected of 5 separate groups of 10 patients each, before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation. Finally, recordings were made of 20 control speakers without any known vocal defects (table 3.3).

Table 3.3. Composition of the subject sample: longitudinal group before, 6 months after, and 2 years after radiation; separate groups before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation and control group; the mean age per group is given in years.

	before	6 mths	2 yrs	3-7 yrs	7-10 yrs	control
longitudinal (n)	10 ->	10 ->	10			
mean age (yrs)	63	63	65			
separate (n)	10	10	10	10	10	20
mean age (yrs)	65	65	68	70	71	65

3.3.2 Voice characteristics following radiotherapy: influencing factors

The composition of the longitudinal and separate speaker groups as described in the previous section was made independent of initial tumour size, initial surgery, or radiation dose schedule. However, these aspects become essential in investigating voice characteristics following radiotherapy, as in Chapter 7 on voice quality and vocal function. Five factors were taken into account to investigate voice characteristics following radiotherapy:

- stage of the tumour (unilateral, T1a, or bilateral, T1b),
- initial surgery (biopsying or stripping),
- radiation dose schedule (66 Gy in 33 fractions, 60 Gy in 30 fractions, or 60 Gy in 25 fractions),
- age (younger than 65 years old, between 65 and 70 years old, between 70 and 75 years old, and older than 75 years), and
- smoking habit after treatment (yes or no).

Mean data of patients after radiotherapy of the longitudinal group and the separate groups as well as of the control group are given in table 3.4.

Table 3.4. Overview of percentages of 40 patients after radiotherapy (separate groups) and of 10 other patients after radiotherapy (longitudinal group) concerning tumour stage (T1a or T1b), initial surgery (1: biopsying or 2: stripping), radiation dose schedule (1: 66 Gy in 33 fractions, 2: 60 Gy in 30 fractions, and 3: 60 Gy in 25 fractions), smoking after treatment (yes or no), and age (1: younger than 65, 2: between 65 and 70 years old, 3: between 70 and 75 years old, and 4: older than 75 years), and of 20 control speakers concerning smoking and age.

GRP	STAGE		SURG.		SCHEDULE			SMOKING		AGE			
	T1a	T1b	1	2	1	2	3	yes	no	1	2	3	4
sep. n=40	75	25	73	27	40	25	35	23	77	25	40	20	15
long. n=10	60	40	60	40	70	30	0	40	60	60	40	0	0
contr. n=20								50	50	50	30	10	10

In the next chapter the first phase of the development of a protocol to analyse voice characteristics following radiotherapy by means of perceptual analyses of voice quality is described. The research design as described in section 3.3.1 is applied in order to investigate which perceptual voice quality parameters can differentiate speaker groups best.

4

PERCEPTUAL ANALYSES OF VOICE QUALITY*

Abstract

Voice quality of 60 male patients with early glottic carcinoma before and after radiotherapy, and of 20 control speakers was analysed by means of perceptual descriptions. Ratings were gathered from 3 trained and 20 naive raters on read-aloud text and on sustained /a/ speech material; also, the speakers themselves and their partners were asked to evaluate the voice quality of the speakers. A trend was observed for voices before radiotherapy having the most deviant voices; voices 6 months after, 2 years after and 3-7 years after radiotherapy sounded less deviant, but were significantly different from voices of the control speakers; voices 7-10 years after radiotherapy were comparable with control voices on read-aloud text. Over all patients after radiotherapy, 55% had normal voice quality, while 45% end up with pathological voice quality. Speaker group differences were more often found on read-aloud text than on sustained /a/. Correlations between the two types of speech material were significant, but low. It is concluded that, with a proposed limited set of scales, perceptual analyses by trained raters (analytical ratings) and naive raters (evaluative ratings) on read-aloud text, and evaluations by the speakers themselves and their partners are valuable in investigating voice quality before and after radiotherapy. Perceptual analyses on sustained /a/ are not recommended.

---

*This chapter is a substantially revised and extended version of de Leeuw (1991).

4.1 Introduction

The aim of this thesis is to develop a protocol to investigate voice characteristics following radiotherapy. As was already argued in Chapter 1, a distinction is made between voice quality and vocal function. Vocal function can be investigated by physiological analyses, voice quality by means of perceptual and acoustical analyses. In the present chapter, perceptual analyses of voice quality are investigated in detail. Acoustical and physiological analyses are explored in the next chapters.

Several systems to analyse voice quality perceptually are available, such as Laver's Vocal Profile Analysis Protocol (Laver, 1980), which is a phonetically based system, the GBRAS scales (Hirano, 1981), proposed by the Japanese Society of Logopedics and Phoniatrics, scales developed by means of the semantical differential technique (Osgood et al., 1957) adapted for instance for Swedish by Hammarberg (1986) and for Dutch by Fagel et al. (1983), the Buffalo Voice Profile by Wilson (1987), developed primarily for the evaluation of children's speech, and the evaluation system by Pahn & Pahn (1976), a clinical system based on voice perception and voice production. Many other procedures are in use, some of them derived from the systems mentioned above, others using visual analogue scales, direct magnitude estimation, or paired or triadic comparison tasks (Kreiman et al., 1993).

The choice of the system depends on the goal of the perceptual analysis and on the type of the raters that are asked to judge voice quality. Kreiman et al. (1992, 1993) argued that raters compare voice samples under investigation with internal standards. If the voice samples are non-pathological, all raters have the same experience and their internal standards will be similar and stable. When the voice samples to be judged are pathological, inexperienced raters have all similar internal standards and will compare pathological voice quality to their internal standards of normal voice quality. However, expert raters seem to differ in their internal standards because of different levels of experience with pathological voices. In order to handle this problem, Kreiman et al. proposed a rating protocol using fixed reference voices. Cordes (1994) and Franken (1995) have proposed that the reliability of clinical ratings can be improved by means of specific training procedures.

In the present study we asked trained and naive raters, and the speakers themselves and their partners to judge the voice quality of the speakers. The task of the trained raters was to provide an analytic and precise analysis of voice quality. The Vocal Profile Analysis Protocol by Laver (1981) was chosen as the basis for these analyses, because it includes a thorough training procedure. Furthermore, a reference tape was used to equalise internal standards of the raters. This reference tape was also used for the naive raters. Perceptual analyses by naive raters were used to find out how 'ordinary' people describe voice quality. To this end, the scaling instrument developed by Fagel et al. (1983) was chosen as a basis. This instrument was originally developed to obtain reliable ratings of normal speakers by naive Dutch raters, and was subsequently adapted to evaluate pathological speech such as cleft palate speech (van Erp, 1991), the speech of stutterers (Franken et al., 1995), and the speech of laryngectomized patients (Nieboer et al., 1988). The present author has adapted the scaling instrument for voice quality analysis before and after radiotherapy (de Leeuw, 1991). Next to the trained and naive raters, the speakers themselves and their partners were asked to describe the voice of the speaker. For this purpose, a similar scaling instrument was used as used for the naive raters. The results from these self-judgements are compared to results from self-ratings on vocal performance in Chapter 8 in order to investigate the effect of decreased voice quality on daily life.

Besides the assessment of voice quality of patients recorded before and after radiotherapy compared to control speakers by these four rater groups, direct relations between the groups are investigated in order to reveal which perceptual analyses are to be chosen in investigating voice characteristics following radiotherapy.

4.2 Methods

4.2.1 Speakers

The subject sample was described in detail in Chapter 3. Briefly, the sample consisted of 60 male patients with early glottic cancer, who were treated with radiotherapy. There was a longitudinal group of 10 patients of whom voice samples were recorded before, as well as 6 months and 2 years after radiotherapy. Furthermore, data were collected of 5 separate groups of 10 patients each, before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation. Finally, recordings were made of 20 control speakers. In summary, there were in total 100 speaker cases (longitudinal group (n=30), separate groups (n=50), and control group (n=20).

4.2.2 Speech material/recordings/preprocessing

Read-aloud text and sustained /a/ were chosen as speech material. Speakers read out the same text of neutral content for about five minutes, and produced a sustained /a/ for about three seconds at a comfortable pitch and loudness level. All the speech material was recorded in a sound treated room, using a Philips N8214 microphone and a Casio DAT-recorder. Recording level was adjusted for each speaker to optimise signal-to-noise ratio, and then kept constant for that speaker. Fragments of the texts and parts of the sustained /a/ were digitised by means of the SoundEditor of an Iris Indigo R4000 with a sample frequency of 48 kHz and a 16 bit resolution. The duration of each of these read-aloud fragments was approximately 45 seconds, which is generally assumed to be sufficiently long for obtaining reliable perceptual judgements (Laver and Hanson, 1981). The sustained /a/ material was segmented beginning at the onset plus a stable part of the vowel during two seconds. In contrast to what is usually done, the onset was added because de Krom (1994) found that adding the onset raised reliability of ratings on roughness.

Two analogue tapes (read-aloud text and sustained /a/) were composed for the perceptual analyses. For each tape, the following sequence of voice samples was composed. The first ten samples were examples of the speakers which ranged from extremely pathological to normal as judged by the present author, in order to give the raters the whole span of voice samples. After these ten reference voices, ten training samples, randomly chosen from the total of speech samples, were presented, which had a double function: to accustom the subjects to the test and to allow assessing intrarater reliability. Finally, the 100 test samples were presented in random order, including a second presentation of the ten training samples.

4.2.3 Trained raters/rating scales

The trained raters were three female phonetic researchers/speech therapists. Two had taken a training course on the Vocal Profile Analysis Protocol by Laver (1980). The third rater had been trained by one of the two others. In the present study an adapted and limited version of the Vocal Profile was used. The scales were divided into two types: 13-point and 7-point scales. The 13-point scales relate to features that are always present in speech, like pitch; they go from one extreme (very low pitch) via a neutral reference point to the opposite extreme (very high pitch). In the original Vocal Profile, these bipolar scales are two separate unipolar scales. The 7-point scales are related to voice features that can be absent in speech, like roughness.

The scales used in the present study (nine altogether) were selected on the basis of a trial study (unpublished) on five patients before radiotherapy and six months after radiotherapy. T-tests revealed that the following scales differentiated significantly between the two speaker groups both on read-aloud text and on sustained /a/: *pitch* (the impression of average pitch level), *sonority* (the extent to which the voice sounds resonant or sharp), *tension* (the impression of the muscle tension in the vocal folds), *abrupt voice onset* (the amount of hard glottal attack), *breathiness* (the amount of air escaping through the glottis), *roughness* (the amount of aperiodic vibration, resulting in a rough and rasping quality), and *creak* (the amount of discrete pulses that can be heard during phonation). These seven scales were rated for both types of speech material. During training sessions, the scale *audible breath* (the amount of audible inadequate breathing) was later added for the read-aloud text, and the scale *steadiness* (the extent to which a vowel is steady) for the sustained /a/.

Preceding the individual listening sessions in the present study, the raters judged the above mentioned ten reference voice samples together. In this way, anchor points for the scales were set. The raters judged the actual test voice samples independently from each other. First, the read-aloud samples were judged.

The sustained /a/ material was judged one month later. On average, the rating of each scale took about 30 seconds.

4.2.4 Naive raters/rating scales

The naive raters were 20 university students (6 male, 14 female), not familiar with pathological voices. They were paid for their participation. The raters listened to the tapes in a quiet room, independently from each other. The set of semantic 7-point scales was based on the scaling instrument developed by Fagel et al. (1983) and adapted for pathological speech in earlier research (De Leeuw, 1991). De Leeuw found that naive raters were reliable on various scales for read-aloud text and for sustained /a/, covering pitch, voice quality, prosodic and temporal aspects, and articulation. However, because the focus of the present study was on laryngeal voice quality, only the scales referring to pitch and voice quality aspects were taken into account. For the read-aloud text 14 scales were used and for the sustained /a/ 13 scales. The scale *not intelligible -- intelligible* was used for the read-aloud text only. The remaining scales were *unpleasant -- pleasant*, *ugly -- beautiful*, *breathy -- not breathy*, *dull -- clear*, *high -- low*, *shrill -- deep*, *unsteady -- steady*, *panting -- not panting*, *tense -- relaxed*, *rough -- not rough*, *creaky -- not creaky*, *speaking with difficulty -- speaking without difficulty*, and *deviant -- not deviant*.[1]

Preceding the actual judging, the naive raters were presented with the above mentioned ten reference voice samples, in order to give the raters the whole span of voice quality, ranging from extremely pathological to normal. First, the read-aloud samples were judged. The sustained /a/ material was judged one week later. On average, the rating of each scale took about 10 seconds.

4.2.5 Speakers themselves and their partners

At the start of the project the scaling instruments for the naive raters, the speakers themselves and their partners were identical. However, during the project, the instrument for the naive raters was modified, resulting in the final evaluation instrument described above. These modifications were not applied to the instruments for the speakers and their partners, in order to avoid confusions on the part of the speakers and the partners that would have occurred by changing forms every time. The evaluation instruments for the speakers and for the partners consisted of 8 scales *unpleasant -- pleasant*, *ugly -- beautiful*, *breathy -- not breathy*, *dull -- clear*, *high -- low*, *shrill -- deep*, *unsteady -- steady*, and *not intelligible -- intelligible*. The speakers and their partners received score forms with a written instruction. They

[1]These terms are English translations. The original Dutch terms were: slecht verstaanbaar -- goed verstaanbaar, onaangenaam -- aangenaam, lelijk -- mooi, hees -- niet hees, dof -- helder, hoog -- laag, schel -- diep, onvast -- vast, hijgerig -- niet hijgerig, gespannen -- ontspannen, schor -- niet schor, krakerig -- niet krakerig, met moeite spreken -- zonder moeite spreken, afwijkend -- niet afwijkend.

were asked to evaluate the voice of the speaker at home by filling out the form independently from each other.

4.2.6 Statistical analyses

An interrater reliability coefficient was calculated for all perceptual rating scales: Cronbach's alpha. This is a measure of the reliability of the means of the ratings given by a panel of raters (Rietveld & Van Hout, 1993). Alpha is defined as $(MS_{betw} - MS_{res}) / Ms_{betw}$ in which MS_{betw} = Mean Square between speakers and MS_{res} = Mean Square residual. A low reliability can be caused either by a high MS_{res} (the raters disagree), a low MS_{betw} (there is little variation between the speakers, i.e. the true variance is low), or by both. Intrarater reliability was established as follows: ten voice samples, selected from the available material itself and ranging from extremely deviant to normal, were presented twice: first as training samples (the first ten voice samples that had to be judged), the second time as part of the 100 test samples. Percentages of the first ratings that were within 1 scale value of the second (repeated) ratings were calculated.

Principal Component Analysis (PCA) was carried out to examine interrelations among the various perceptual voice quality rating scales per rating condition. The correlations of the ratings over the scales per rater group (4) and per speech material (2) were tabulated in 6 correlation matrices. The PCA was used to decompose these matrices into factors. For each matrix, the initial factors were rotated to a varimax criterion (Wilkinson, 1989).

Two types of one-way analysis of variance were carried out: analyses on the ratings of the three longitudinal speaker groups (with repeated measures) and analyses on the ratings on the six separate speaker groups (without repeated measures). Since the value of the F-statistic or p-value only provides information concerning the likelihood that speaker group differences are present or not, the η^2 statistic is given as an indicator of the strength of an effect. The η^2 statistic is defined as: $SS_{between} / SS_{total}$, and can be interpreted as the proportion of the total variability in the dependent variable (i.e. the voice quality parameter) that can be accounted for by the independent variable (i.e. the speaker groups) (Rietveld & Van Hout, 1993; Kirk, 1982). The F-tests on the separate speaker groups were used to analyse voice quality of patients before radiotherapy and of patients in various stages after radiotherapy and of control speakers. Post hoc tests according to Tukey (Winer, 1971) were used to analyse the differences between the separate speaker groups.

Finally, Pearson correlations were used to investigate relations between the two types of speech material (read-aloud text versus sustained /a/) and relations between the four rater groups (trained versus naive raters versus the speakers themselves versus their partners).

4.3 Results

Inspection of the data revealed that for the 13-point scale *tension* judged by the trained raters, of the 100 speaker cases, 92 cases received scores lower than 7 (indicating hyper*tension*) on read-aloud text; the other 8 cases received scores between 7 and 8. On sustained /a/, 96 of the speaker cases received scores lower than 7 and the other 4 received scores between 7 and 8. Therefore, the scale *tension* was recoded from a 13-point to a 7-point scale of (hyper)*tension*: cases with scores between 7 and 8 (slight hypo*tension*) were recoded to score 7 (no hyper*tension*).

4.3.1 Reliability

Interrater reliability was calculated: Cronbach's alpha (table 4.1). The results show that reliability of the trained raters (table 4.1, top) as well as of the naive raters (table 4.1, bottom) was high on read-aloud text as well as on the sustained /a/: all alphas, except for *abrupt voice onset*, exceeded .80, and more than half exceeded .90. Intrarater reliability was calculated by percentages of first ratings that were within 1 scale value of the second ratings of the same items (table 4.2). Reliability was high (above 85%) for the trained raters on read-aloud text as well on sustained /a/. Percentages were moderate (above 56%) for the naive raters; these percentages increased to above 75% for first ratings within 2 scale values of the second ratings. Given these moderate to high reliabilities, mean ratings per speaker of the three trained raters and mean ratings of the 20 naive raters were taken into account in further analyses per scale. Reliability could not be assessed for the judgements of the speakers themselves and their partners since they rated just one voice at the time. However, in order to compare their judgement with the judgements by the trained and naive raters, mean ratings of the speakers themselves and mean ratings of their partners were taken into account in further analyses.

4.3.2 Principal Component Analysis

Principal Component Analysis (PCA) was carried out to examine relations among the different scales per rating condition. Results were taken into account in interpreting the results of speaker group differentiation described in the next sections. Two factors were produced for all matrices except for the naive raters on read-aloud text and for the speakers themselves, where three factors were produced (table 4.2).

Table 4.1. Overview per rating scale of **interrater reliability** (Cronbach's alpha, defined as (MS_{betw} - MS_{res}) / MS_{betw}) and **intrarater reliability** (percentage of first ratings that were within 1 scale value of the second rating; percentages of ratings that were within 2 scale values (naive raters) are printed italic) for the ratings of 3 **trained raters** (top) and of 20 **naive raters** (bottom) over all 100 speaker cases for read-aloud text and sustained /a/ (* this scale was used for the read-aloud text only and ** for the sustained /a/ only).

trained raters	read-aloud text				sustained /a/			
	Inter	MS_b	MS_{res}	Intra %	Inter	MS_b	Ms_{res}	Intra %
pitch	.83	3.58	.61	90	.90	3.48	.35	100
abrupt voice onset	.73	.98	.26	93	.88	1.82	.22	100
breathiness	.93	4.72	.31	90	.92	3.55	.30	93
creak	.89	4.08	.46	97	.82	1.37	.25	97
roughness	.93	7.90	.59	87	.91	5.87	.50	86
sonority	.82	3.24	.57	93	.87	3.97	.53	87
tension	.95	9.67	.52	97	.88	3.71	.46	90
audible breath	.83	2.59	.44	100	*			*
steadiness	**			**	.88	2.10	.26	87

naive raters	read-aloud text					sustained /a/				
	Inter	MS_b	MS_{res}	Intra %		Inter	MS_b	MS_{res}	Intra%	
unpleasant -- pleasant	.93	25.65	1.77	68	*83*	.92	19.47	1.55	76	*87*
ugly -- beautiful	.92	19.45	1.48	68	*88*	.92	17.53	1.41	73	*93*
breathy -- not breathy	.94	32.48	1.98	64	*78*	.93	34.74	2.32	70	*81*
dull -- clear	.91	18.89	1.71	65	*86*	.91	17.78	1.6	70	*87*
high -- low	.91	16.77	1.54	71	*84*	.93	24.85	1.83	72	*87*
shrill -- deep	.89	12.78	1.40	75	*88*	.90	16.31	1.58	72	*87*
not intelligible -- intellig.	.92	20.60	1.63	65	*81*	*			*	
unsteady -- steady	.81	11.04	2.07	64	*79*	.87	18.88	2.42	56	*79*
panting -- not panting	.84	14.06	2.31	51	*83*	.89	18.22	1.98	68	*81*
tense -- relaxed	.86	16.53	2.33	61	*77*	.87	15.66	2.02	66	*81*
rough -- not rough	.94	36.39	2.01	56	*76*	.92	33.05	2.60	67	*81*
creaky --not creaky	.89	19.70	2.17	62	*76*	.88	18.23	2.14	60	*80*
speak. +dif. -- speak. -dif.	.90	23.19	2.32	53	*73*	.89	18.30	1.93	64	*80*
deviant -- not deviant	.92	22.48	1.73	58	*82*	.92	23.78	1.90	59	*79*

On the basis of the factor loadings (>.45) the factors were labelled as Voice Quality and Pitch for the two-factor solutions and Voice Quality, Speech Quality, and Pitch for the three-factor solutions. It can be concluded that raters to a large extent used the same dimensions: Voice Quality and Pitch; the naive raters and the speakers themselves seem to split the dimension Voice Quality into two factors, resulting in a third factor, Speech Quality. These results indicate that on read-aloud text the scales *unsteady -- steady, tense -- relaxed* and *speaking with difficulty -- speaking without difficulty* were interpreted as supralaryngeal characteristics by naive raters; the scales *unpleasant - pleasant, ugly -- beautiful, not intelligible -- intelligible* and *deviant -- not deviant* appeared to be evaluative for laryngeal as well as for supralaryngeal characteristics. The Voice Quality factor (laryngeal characteristics) always explained most of the variance.

Table 4.2. Factor loadings (>.45 is printed bold) and percentage of total variance explained for the mean ratings on read-aloud text and sustained /a/ over all 100 voices of 3 **trained raters** (top), of 20 **naive raters** (middle), and for the ratings of the 100 **speakers themselves** and **their partners** (bottom) (* this scale was used for the read-aloud text only, and ** for the sustained /a/ only). Factors were labelled as Voice Quality, Speech Quality and Pitch.

trained raters	read-aloud text		sustained /a/	
factor	Voice Q.	Pitch	Voice Q.	Pitch
% of total variance expl.	44	26	45	18
pitch	.16	**.88**	.18	**.84**
abrupt voice onset	**.61**	.04	**.58**	.01
breathiness	**.88**	.24	**.80**	.21
creak	.02	**.80**	**.55**	**.49**
roughness	**.89**	.20	**.87**	.22
sonority	**.51**	**.71**	**.76**	**.56**
tension	**.88**	.33	**.90**	.29
audible breath	**.70**	.02	*	*
steadiness	**	**	.42	.09

naive raters	read-aloud text			sustained /a/	
factor	Voice Q.	Speech Q.	Pitch	Voice Q.	Pitch
% of total variance expl.	40	31	17	65	20
unpleasant -- pleasant	.77	**.57**	.05	**.96**	.01
ugly -- beautiful	**.80**	**.53**	.04	**.97**	.01
breathy -- not breathy	**.95**	.15	.01	**.82**	.14
dull -- clear	**.61**	.28	**.68**	**.58**	**.76**
high -- low	.14	.05	**.97**	.03	**.97**
shrill -- deep	.11	.05	**.97**	.16	**.96**
not intelligible -- intelligible	**.58**	**.75**	.18	*	*
unsteady -- steady	.14	**.93**	.03	**.81**	.14
panting -- not panting	**.75**	.35	.12	**.91**	.15
tense -- relaxed	.38	**.85**	.09	**.94**	.12
rough -- not rough	**.94**	.15	.11	**.85**	.30
creaky -- not creaky	**.82**	.16	.16	**.80**	.19
speak. +dif. -- speak. -dif.	.24	**.91**	.11	**.95**	.01
deviant -- not deviant	**.70**	**.61**	.04	**.96**	.05

speakers and partners	speakers			partners	
factor	Voice Q.	Speech Q.	Pitch	Voice Q.	Pitch
% of total variance expl.	24	23	19	43	19
unpleasant -- pleasant	**.79**	.20	.05	**.79**	.05
ugly -- beautiful	**.54**	**.54**	.13	**.84**	.15
breathy -- not breathy	**.75**	.29	.03	**.78**	.18
dull --clear	**.60**	.43	.26	**.82**	.18
high -- low	.01	.07	**.85**	.04	**.76**
shrill -- deep	.09	.17	**.82**	.06	**.83**
not intelligible -- intelligible	.16	**.75**	.01	**.69**	.10
unsteady -- steady	.03	**.84**	.21	**.59**	.36

4.3.3 Differentiation among speaker groups

The next step was to investigate which scales differentiated speaker groups best. Differences between speaker group means of the longitudinal group were described in section 4.3.3.1. Scales that differentiated longitudinal speaker groups best, were used in order to investigate voice quality of patients before and after radiotherapy compared to control speakers (separate speaker groups). This is described in section 4.3.3.2.

4.3.3.1 Longitudinal speaker group

The results of the analysis of variance with repeated measures on the longitudinal speaker group before, six months after and two years after radiotherapy are given in table 4.3 as judged by trained raters (top), by naive raters (middle) and by the speakers themselves and their partners (bottom). All trends were found to be linear, indicating that voices before radiotherapy were judged to be more deviant than voices 6 months or 2 years after radiotherapy.

According to the trained raters, *breathiness* explained 54% and 38% of the variance on read-aloud text and sustained /a/, respectively. The scale *tension* explained even more of the variance according to the trained raters, namely 62% and 54% respectively. *Roughness* showed a trend on read-aloud text, explaining 53% of the variance, but not on sustained /a/.

As judged by the naive raters, the scale *breathy -- not breathy* explained 47% and 45% of the variance on read-aloud text and sustained /a/, respectively. The scale *rough -- not rough* explained 37% of the variance on read-aloud text but did not indicate a trend on sustained /a/. The scale *tense -- relaxed* used by the naive raters did not indicate a trend at all. The scale *panting -- not panting* explained 29% and 42% of the variance on read-aloud text and on sustained /a/, respectively. It is striking that, according to the naive raters on read-aloud text, the scales that loaded high both on the factors Voice Quality and Speech Quality (as described in the previous section), *unpleasant -- pleasant, ugly -- beautiful, not intelligible -- intelligible* and *deviant -- not deviant,* showed the strongest trend, explaining 71%, 66%, 70%, and 58% respectively. It seems that the scales that loaded mainly on the factor Speech Quality (*unsteady -- steady* and *tense -- relaxed*) did not differentiate the speaker groups, although this conclusion did not hold for the scale *speaking with difficulty -- speaking without difficulty* judged by the naive raters.

The significant trends that were found for the scales *ugly -- beautiful, breathy -- not breathy* and *dull -- clear* judged by the speakers themselves (table 4.3, bottom) and for the scale *unsteady -- steady* judged by their partners were less strong (explaining less than 40%) compared to the other rater groups.

Table 4.3. Results of analyses of variance (ANOVA, with repeated measures): F-statistic ($p<0.05$ is printed bold), η^2 (proportion of variance explained), SS_{betw} and SS_{total} of the various rating scales over the 3 longitudinal speaker groups for the ratings on read-aloud text and sustained /a/ of 3 **trained raters** (top) and of 20 **naive raters** (middle), and for the ratings of the **speakers themselves** and **their partners** (bottom) (* this scale was used for the read-aloud text only and ** for the sustained /a/ only; the F superscript of the speakers and the partners indicates a deviating number of speakers per speaker group).

trained raters	read-aloud text				sustained /a/			
	F	η^2	SS_{betw}	SS_{total}	F	η^2	SS_{betw}	SS_{total}
pitch	3.15				1.76			
abrupt voice onset	1.10				1.40			
breathiness	**10.35**	.54	18.91	35.35	**5.46**	.38	8.59	22.76
creak	2.22				0.91			
roughness	**10.33**	.53	22.84	42.75	1.95			
sonority	2.13				2.07			
tension	**14.64**	.62	32.71	52.83	**10.67**	.54	16.24	29.94
audible breath	2.39				*			
steadiness	**				2.77			

naive raters	read-aloud text				sustained /a/			
	F	η^2	SS_{betw}	SS_{total}	F	η^2	SS_{betw}	SS_{total}
unpleasant -- pleasant	**22.16**	.71	10.59	14.89	**3.56**	.28	4.37	15.39
ugly -- beautiful	**17.42**	.66	6.84	10.37	**3.45**	.28	3.35	12.07
breathy -- not breathy	**7.98**	.47	12.02	25.58	**7.38**	.45	10.13	22.48
dull -- clear	**4.18**	.32	4.13	13.01	1.45			
high -- low	0.06				0.64			
shrill -- deep	0.02				3.28			
not intelligible -- intell.	**20.77**	.70	9.53	13.66	*			
unsteady -- steady	2.46				**3.57**	.28	3.32	11.67
panting -- not panting	**3.61**	.29	3.09	10.78	**6.51**	.42	4.25	10.12
tense -- relaxed	3.17				3.09			
rough -- not rough	**5.28**	.37	10.86	29.38	1.93			
creaky --not creaky	1.77				0.43			
speak. +dif. -- speak. -dif.	**4.15**	.32	3.88	12.29	2.96			
deviant -- not deviant	**12.27**	.58	9.85	17.07	**4.30**	.32	7.14	22.08

speakers and partners	speakers				partners			
	F	η^2	SS_{betw}	SS_{total}	F	η^2	SS_{betw}	SS_{total}
unpleasant -- pleasant	1.75				0.97^{9}			
ugly -- beautiful	**4.06**	.31	5.60	18.00	2.19^{8}			
breathy -- not breathy	**5.65**	.39	45.27	117.33	2.61^{9}			
dull -- clear	**4.94^{9}**	.38	19.85	51.99	1.61^{9}			
high -- low	1.04				0.68^{8}			
shrill -- deep	0.03				1.44^{9}			
not intelligible -- intell.	1.44				0.02^{9}			
unsteady -- steady	0.65^{9}				**4.13^{8}**	.38	20.58	54.00

The scales that differentiated speaker groups best and were mainly part of the factor Voice Quality were chosen to investigate voice quality of patients before and after radiotherapy compared to control speakers in the next section, resulting in the following scales: *breathiness, roughness,* and *tension* for the trained raters and *breathy -- not breathy, rough -- not rough,* and *panting -- not panting* for the naive raters. The scales that showed significant differences on read-aloud text only (*roughness* for the trained raters and *rough -- not rough* for the naive raters) were taken into account for both types of speech material, in order to compare results for read-aloud text and sustained /a/ speech material. In order to compare judgements from the speakers themselves with judgements from their partners, the scales *breathy -- not breathy, ugly -- beautiful, dull - clear* (significant results for the speakers) and *unsteady -- steady* (significant results for the partners) were all included for both patients and partners in the next section.

4.3.3.2 Separate speaker groups

Analyses of variance without repeated measures were carried out on the separate speaker groups (patients before radiotherapy, 6 months after, 2 years after, 3-7 years after, 7-10 years after radiotherapy, and control speakers) for the selected scales. All scales showed significant differences between the speaker groups. Histograms for the six rating conditions are given in figure 4.1.

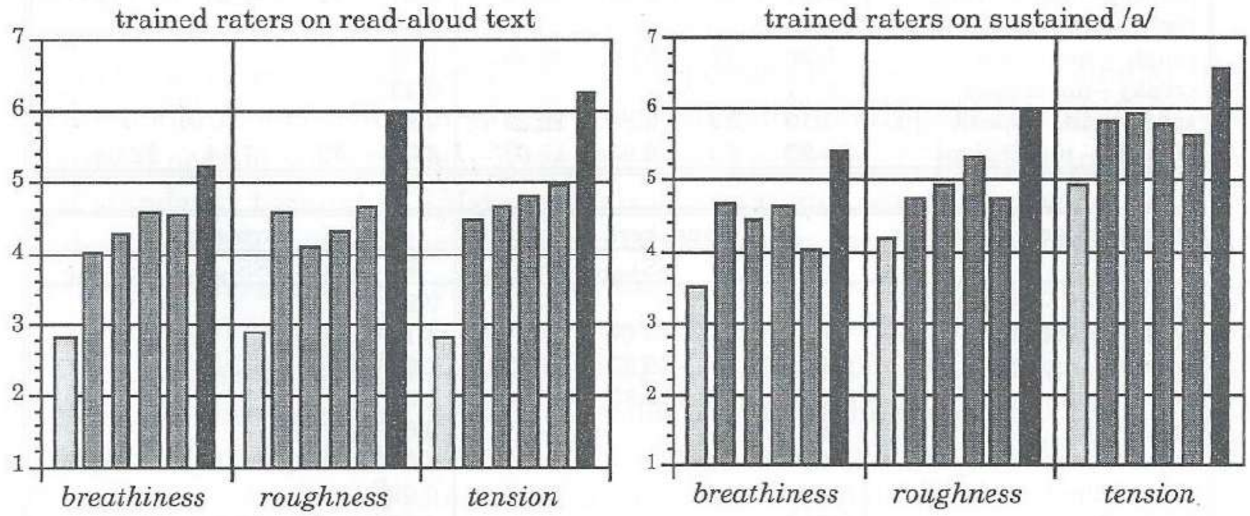

Figure 4.1a. Mean scale ratings of 3 trained raters on *breathiness*, *roughness* and *tension* on read-aloud text (at the left) and on sustained /a/ (at the right) of the separate speaker groups (from left to right): before radiotherapy (n=10), 6 months after (n=10), 2 years after (n=10), 3-7 years after (n=10), and 7-10 years after radiotherapy (n=10), and control speakers (n=20)). Ratings below 4 were considered to be pathological, ratings above 4 to be normal. An overview of significant differences between speaker groups is given in table 4.4.

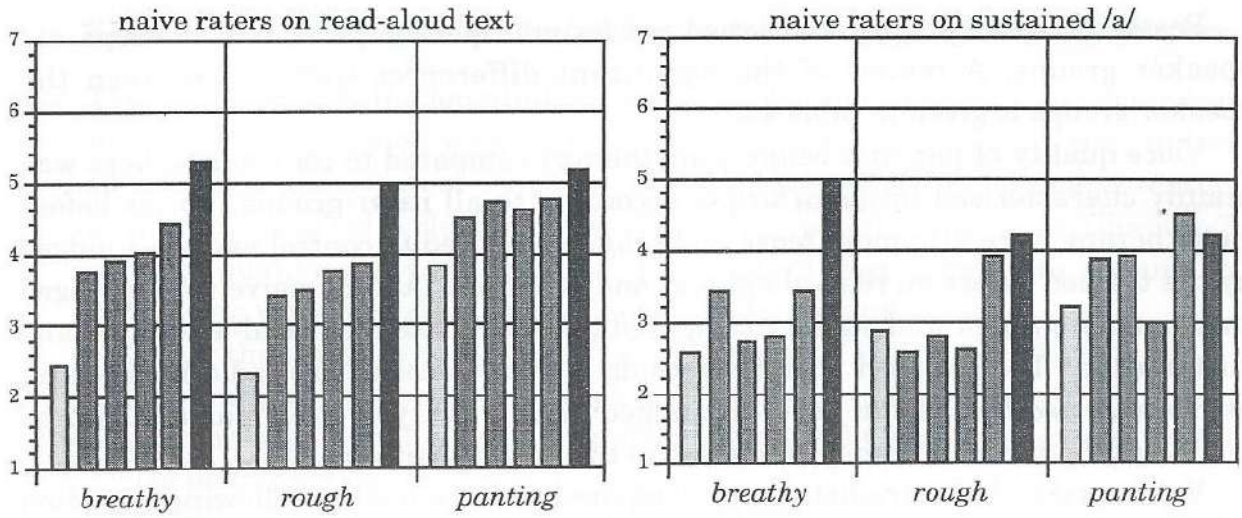

Figure 4.1b Mean scale ratings of 20 naive raters on *breathy -- not breathy*, *rough -- not rough*, and *panting -- not panting* on read-aloud text (at the left) and on sustained /a/ (at the right) of the separate speaker groups (from left to right: before radiotherapy (n=10), 6 months after (n=10), 2 years after (n=10), 3-7 years after (n=10), and 7-10 years after radiotherapy (n=10), and control speakers (n=20). An overview of significant differences between speaker groups is given in table 4.4.

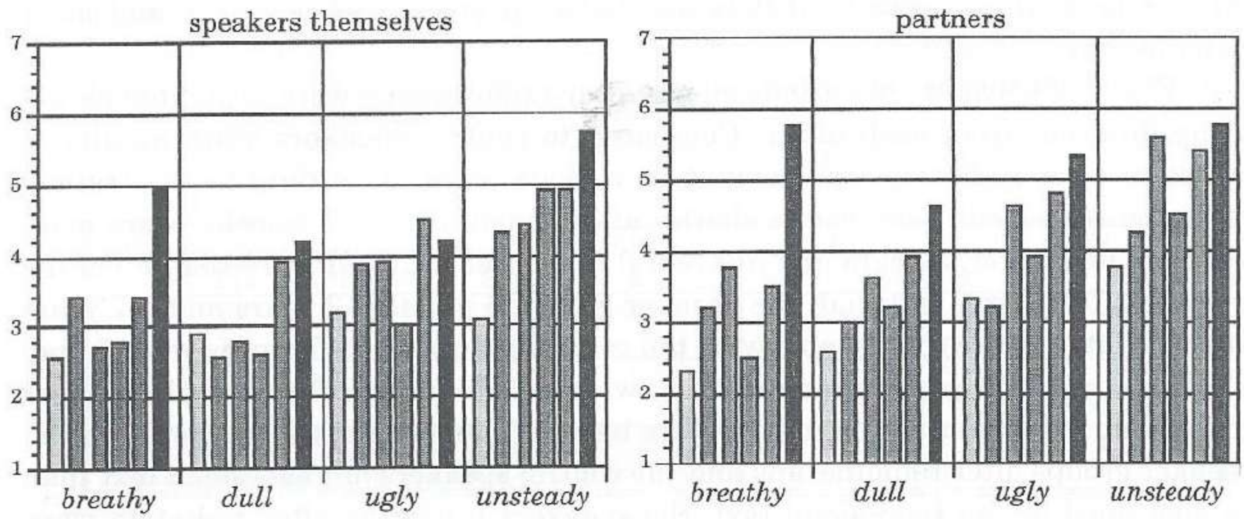

Figure 4.1c Mean scale ratings of the speakers themselves (at the right) and their partners (at the left) on *breathy -- not breathy*, *dull -- clear*, *ugly -- beautiful*, and *unsteady -- steady* of the separate speaker groups (from left to right: before radiotherapy (n=10), 6 months after (n=10), 2 years after (n=10), 3-7 years after (n=10), and 7-10 years after radiotherapy (n=10), and control speakers (n=20). An overview of significant differences between speaker groups is given in table 4.4b.

As judged by the trained and naive raters, it is obvious that voice quality of patients before radiotherapy is the most deviant; after radiotherapy voice quality improved, but was still different from voices of control speakers. A similar but less obvious trend was found for the ratings by the speakers themselves and their partners.

Posthoc tests (Tukey) were carried out for multiple comparison of the separate speaker groups. A review of the significant differences ($p<0.05$) between the speaker groups is given in table 4.4.

Voice quality of patients before radiotherapy compared to control speakers was mainly characterised by *breathiness* according to all rater groups. Voices before radiotherapy were also more *tense* and *rough* compared to control voices as judged by the trained raters on read-aloud text and sustained /a/. The naive raters judged them as more *rough* and *panting* compared to control voices on read-aloud text and sustained /a/. The speakers themselves judged their voices before radiotherapy also as more *unsteady* and the partners judged voice quality of their partner also as more *dull, ugly,* and *unsteady* compared to the control partners.

Voice quality before radiotherapy compared to voice quality following radiation was judged as more *breathy* by trained raters on read-aloud text and by naive raters on read-aloud text and sustained /a/. Furthermore, voice quality before radiotherapy compared to voices of some patient groups after radiotherapy was rated as more *tense* (trained raters on read-aloud text), more *rough* (naive raters on read-aloud text) and more *panting* (naive raters on read-aloud text and sustained /a/). According to the trained raters on sustained /a/, the speakers themselves and the partners, there were no differences between voice quality before and after radiotherapy.

Voices of speakers in various stages after radiotherapy were evaluated as not being different from each other. Compared to control speakers voice quality of speakers after radiation was deviant in various ways. According to the trained raters on read-aloud text, voices shortly after radiotherapy (6 months) were more *tense*, on long term (2 years and 3-7 years) the voices sounded more *rough*. On the sustained /a/ speech material, the speaker groups 6 months, 2 years and 3-7 years after radiation were comparable with the control group; no differences were found. The speakers 7-10 years after radiation were different from the control speakers concerning *breathiness*. The naive raters heard also more differences between the speaker groups after radiotherapy and the control speakers on read-aloud text than on sustained /a/: on read-aloud text, the speakers 6 months after radiation were more *breathy* and *rough* compared to the control speakers; on sustained /a/ they were more *rough*. On longer term these differences on read-aloud text seemed to diminish: 2 years after radiotherapy, the voices were evaluated as *breathy* and *rough* on read-aloud text and as more *breathy* on sustained /a/; 3-7 years after radiation speakers were *breathy* on read-aloud text. Voices of speakers 7-10 years after radiotherapy seem to decrease again. On sustained /a/ the voices sounded more *breathy, rough,* and *panting* compared to control voices.

The speakers themselves and their partners (figure 4.3 and table 4.4b) evaluated the voices after radiotherapy as more *dull, breathy,* and/or *unsteady* compared to control speakers. The speakers 3-7 years after radiotherapy judged their voices as more *ugly* compared to speakers 7-10 years after radiation.

Table 4.4a. Review of posthoc comparisons (Tukey) of the separate speaker groups (patients before radiotherapy (n=10), 6 months after (n=10), 2 years after (n=10), 3-7 years after (n=10) and 7-10 years (n=10) after radiotherapy and control speakers (n=20)) for the ratings by trained raters on read-aloud text and sustained /a/ and by naive raters on read-aloud text and sustained /a/ on significant scales ($p<0.05$); n.s. means not significant. All comparisons are positive, that is, in time, voices become less deviant. For example: according to the trained raters on read-aloud text, voices before radiotherapy are characterised by more *breathiness, tension* and *roughness* compared to control voices; furthermore, voices 6 months after radiation contain more *tension* than control voices.

trained raters: text	6 months	2 years	3-7 years	7-10 years	control
before	n.s.	*breathiness*	*breathiness, tension*	*breathiness, tension*	*breathiness, tension, roughness*
6 months		n.s.	n.s.	n.s.	*tension*
2 years			n.s.	n.s.	*roughness*
3-7 years				n.s.	*roughness*
7-10 years					n.s.

trained raters: /a/	6 months	2 years	3-7 years	7-10 years	control
before	n.s.	n.s.	n.s.	n.s.	*breathiness, tension, roughness*
6 months		n.s.	n.s.	n.s.	n.s.
2 years			n.s.	n.s.	n.s.
3-7 years				n.s.	n.s.
7-10 years					*breathiness*

naive raters: text	6 months	2 years	3-7 years	7-10 years	control
before	*breathy*	*breathy*	*breathy*	*breathy, rough, panting*	*breathy, rough, panting*
6 months		n.s.	n.s.	n.s.	*breathy, rough*
2 years			n.s.	n.s.	*breathy, rough*
3-7 years				n.s.	*breathy*
7-10 years					n.s.

naive raters: /a/	6 months	2 years	3-7 years	7-10 years	control
before	*breathy*	*breathy*	*breathy, panting*	*breathy*	*breathy, rough, panting*
6 months		n.s.	n.s.	n.s.	*rough*
2 years			n.s.	n.s.	*breathy*
3-7 years				n.s.	n.s.
7-10 years					*breathy, rough, panting*

Table 4.4b. Review of posthoc comparisons (Tukey) of the separate speaker groups (patients before radiotherapy (n=9), 6 months after (n=9), 2 years after (n=10), 3-7 years after (n=10) and 7-10 years after radiotherapy (n=10) and control speakers (n=20) for the ratings by the speakers themselves and by their partners on the significant scales ($p<0.05$).

speakers	6 months	2 years	3-7 years	7-10 years	control
before	n.s.	n.s.	n.s.	n.s.	*breathy, unsteady*
6 months		n.s.	n.s.	n.s.	*dull*
2 years			n.s.	n.s.	*breathy*
3-7 years				*ugly*	*breathy, dull*
7-10 years					n.s.
partners	6 months	2 years	3-7 years	7-10 years	control
before	n.s.	n.s.	n.s.	n.s.	*breathy, dull, ugly, unsteady,*
6 months		n.s.	n.s.	n.s.	*breathy, ugly*
2 years			n.s.	n.s.	*breathy*
3-7 years				n.s.	*breathy*
7-10 years					*breathy*

4.3.3.3 Voice quality following radiotherapy

In the previous sections, a trend was observed for patients before radiotherapy having the most deviant voices; voices 6 months after radiotherapy sounded less deviant, but were significantly different from voices of control speakers. In this section, the amount of deviant voice quality following radiotherapy is assessed. To that end, dichotomies were introduced based on normal or pathological *breathiness*, *roughness* and *tension* as scored by the trained raters on read-aloud text. These voice quality parameters were chosen, because the trained raters agreed to score normal (actual score equal or higher than 4) or pathological (actual score lower than 4) on these 7-point scales and because these 3 scales appeared to differentiate speaker groups best.

For the 10 longitudinal patients, it appeared that there are four patterns of normal or pathological (*breathiness*, *roughness* or *tension*, or a combination) voice quality before and after radiotherapy (table 4.5): voice quality is pathological before radiotherapy but becomes normal after radiotherapy (speaker 1 to 5), voice quality is pathological before radiotherapy and remains pathological (speaker 6, 7 and 8), voice quality is normal, becomes pathological 6 months after radiotherapy, but becomes normal again 2 years after radiotherapy (speaker 9), and voice quality is pathological, becomes normal 6 months after radiotherapy, but pathological again 2 years after radiotherapy (speaker 10).

Of the 40 patients after radiotherapy in the separate speaker groups 55% had normal voice quality (in terms of pathological scores on either *breathiness*, *roughness* or *tension*), while the other 45% showed pathological voice quality.

Table 4.5. Overview of the longitudinal speaker group in terms of speaker number, pattern of voice quality before, 6 months after, and 2 years after radiotherapy (pathological *breathiness* (B), *roughness* (R), or *tension* (T), or normal (N)).

NR	PATTERN		
	before	6 months after	2 years after
1	BT	N	N
2	BRT	N	N
3	B	N	N
4	BRT	N	N
5	R	N	N
6	BRT	R	T
7	BRT	BRT	BRT
8	BT	R	BR
9	N	BT	N
10	BRT	N	RT
% normal	10	60	60

So, over all 50 patients after radiotherapy (longitudinal and separate groups), it can be concluded that approximately 55% had normal voice quality, while 45% end up with pathological voice quality.

4.3.4 Some relations between read-aloud text and sustained /a/

The trained and naive raters judged voice quality on read-aloud text and on sustained /a/. The scale *roughness* by the trained raters and the scales *rough -- not rough, dull -- clear* and *speaking with difficulty -- speaking without difficulty* by the naive raters showed significant differences between the speaker groups on read-aloud text, but not on sustained /a/; for the scale *unsteady -- steady* by the naive raters, it was just the other way around (table 4.3). These results already indicated differences between the two types of speech material. The relation between the two types of speech material was further investigated for those scales indicating significant differences between the speaker groups both on read-aloud text and sustained /a/: *breathiness* and *tension* judged by the trained raters, and *breathy -- not breathy, unpleasant -- pleasant, ugly -- beautiful, panting -- not panting*, and *deviant -- not deviant* judged by the naive raters. Correlations between the two types of speech material were significant, but low to moderate: .36 for *breathiness* and .25 for *tension* judged by the trained raters, and .68 for *breathy -- not breathy*, .50 for *unpleasant -- pleasant*, .46 for *ugly -- beautiful*, .53 for *panting -- not panting*, and .52 for *deviant -- not deviant* judged by the naive raters (Pearson correlations). These low correlations indicated again (although ratings on read-aloud text and sustained /a/ were equally reliable and differentiated among speaker groups) clear differences between the two types of speech material.

4.3.5 Some relations between the rating groups

Four rating groups were used in our research: trained raters, naive raters, the speakers themselves and their partners. The rating scales used by the trained raters were well defined: three scales differentiated among the speaker groups: *breathiness*, the amount of air escaping through the glottis, *roughness*, the amount of aperiodic vibration resulting in rough and rasping quality and *tension*, the impression of the muscle tension of the vocal folds. In order to investigate, over all speakers, in what way evaluations by the naive raters, the speakers themselves and the partners were related to these analytic ratings, the evaluations on the scales *breathy -- not breathy* and *rough -- not rough* by the naive raters and the scale *breathy -- not breathy* by the speakers and by their partners were correlated to *breathiness* and *roughness*. On read-aloud text, strong significant Pearson correlations ($r>.70$) were found between *breathiness* and evaluations of the naive raters on the scales *rough -- not rough* ($r=.73$) and *breathy -- not breathy* ($r=.80$), and between *roughness* and evaluations by the naive raters on *rough -- not rough* ($r=.73$) and *breathy -- not breathy* ($r=.75$). On the sustained /a/, there were no strong correlations ($r>.50$) found between ratings on *breathiness* and *roughness* and evaluations by the naive raters on *rough -- not rough* or *breathy -- not breathy*. There were also no strong correlations between the ratings by the trained or naive raters and the evaluations by the speakers themselves or their partners. The correlation between the speakers and the partners was .59 for the scale *breathy -- not breathy*. These results suggest that on read-aloud text, naive raters used the scales *breathy -- not breathy* and *rough -- not rough* in a similar way as the trained raters used *breathiness* and *roughness*. On a sustained /a/, however, naive raters differed strongly from the trained raters. Furthermore, the speakers themselves and their partners evaluated the voice of the speakers differently from the trained and naive raters.

4.4 Discussion

The main aim of this chapter was the development of a protocol to investigate voice quality by means of perceptual evaluations by trained raters, naive raters, the speakers themselves, and their partners. A trend was seen for patients before radiotherapy having the most deviant voices. Voices 6 months after radiation, 2 years after, 3-7 years after and 7-10 years after radiotherapy sounded less deviant. Still, all speaker groups after radiation were significantly different from the control group. Before discussing in detail specific voice characteristics of speakers before and after radiotherapy, we will focus on the differences between read-aloud text and sustained /a/ and on the differences between the four rater groups.

4.4.1 Read-aloud text versus sustained /a/

Concerning perceptual analyses of read-aloud text and sustained /a/, reliability of listeners' voice quality ratings may be equally high for read-aloud text and sustained /a/ (de Leeuw, 1991). Also de Krom (1994) found no differences between various types of speech material (post-onset vowels, whole vowel, or running speech fragments) concerning reliability of ratings on breathiness or grade of dysphonia. He did find that adding the onset of sustained vowels raises reliability of ratings on roughness. The question remains, however, whether deviant voice quality is represented more prominently during vowels or in transient parts of running speech. Results on reliability of perceptual ratings do not give information about this matter. In Section 4.3.4, correlations between read-aloud text and sustained /a/ were found to be moderate. Concerning perceptual analysis of voice quality, no data were found in the literature on running speech versus sustained vowels. Concerning acoustical analysis, Klingholz (1990) found moderate correlations (0.63) between running speech and vowels measured by means of the Signal-to-Noise Ratio. Furthermore, in Section 4.3.3, differentiation of speaker groups appeared to be stronger for ratings on read-aloud text than on sustained /a/. This disparity between read-aloud text and sustained /a/ may be an explanation for the contradictory results found in the literature on voice quality following radiotherapy.

Results in a study by Harrison et al. (1990) on acoustical measurements on sustained /a/ speech material showed that voice quality of patients 9 months after radiotherapy was normal again; Hoyt et al. (1992) showed by means of acoustical measures on sustained /a/ that voices 6 months after radiotherapy were better than before radiation; they did not compare the results with normal control speakers. Also, Miller et al. (1990) investigated voices of patients before, during, and after radiation on a sustained /a/ and found that voices seemed to return to normal again, 9 months after radiation. These conclusions support our results, at least for the analytic descriptions by the trained raters on sustained /a/, which we think to correlate best with acoustical measurements (Verdonck-de Leeuw & Koopmans-Van Beinum, 1995).

In contrast with these positive results concerning voice quality after radiation, Lehman et al. (1986) showed on read-aloud text that post-radiation voices had abnormal voice quality that was mainly characterised by greater than normal effort compared to control speakers. Stoicheff et al. (1983), found that radiotherapy positively influenced the voice quality, but that voices one year after radiation were still worse than normal voices, as evaluated by naive raters on read-aloud text.

All these studies support our results that speaker groups are differentiated more often on read-aloud text than on sustained /a/. The question remains whether these differences between read-aloud text and sustained /a/ are caused by perception or production aspects. It could be that the raters in our study were not able to perceive clear differences between speaker groups on sustained /a/ speech material because of its short duration. On the other hand, it seems likely that the cause has to be searched in voice production. Read-aloud text is argued to contain more aspects of deviant voice quality as it is physiologically more complex (Bassich & Ludlow, 1986), although it is also argued to contain more disturbances by articulatory aspects (Gobl & Ni Chasaide, 1988; Lofqvist & McGowan, 1991); sustained /a/ speech material is said to be unnatural (Hammarberg, 1986). In the next chapter, we will present results of acoustical analyses in order to obtain more insight in the perception and production process of read-aloud text and sustained /a/. On the basis of the perceptual descriptions presented in this chapter, we recommend to use read-aloud text in investigating voice quality following radiotherapy. Sustained /a/ material appeared to be less suitable for perceptual analysis of voice quality.

4.4.2 Rater groups

Trained raters judged voice quality in a reliable and analytical way. They differentiated the speaker groups strongly on the scales *breathiness, roughness* and *tension*. Naive raters were also reliable, but seemed to use the various rating scales in a more evaluative manner, as the speakers themselves and their partners did: high correlations were found between descriptive scales like *breathy -- not breathy, rough -- not rough* and evaluative scales like *unpleasant -- pleasant, ugly -- beautiful*. Differentiation of speaker groups was done on more voice aspects by the naive raters than by the speakers themselves and their partners.

As we mentioned already in the introduction, the choice of rater groups depends on the goal of the perceptual description. The task of the trained raters was to provide an analytic and precise description of voice quality. The role of the naive raters was to find out how 'ordinary' people evaluate voice quality as an impression of the communicative aspects of voice quality in the speakers' home environment. To that end, also the speakers themselves and their partners served as raters. For this purpose, it is recommended to use naive raters instead of the speakers themselves or their partners. The fact that the speakers themselves and their partners discriminated least, can be caused by the different way these data were gathered: in contrast to the other rater groups to whom the 100 voices were presented on tape, the speakers themselves and their partners were asked to evaluate just one voice, that of the speaker, and had therefore no reference. Also, it is reasonable to assume that evaluations by some of the speakers themselves were influenced negatively (before radiotherapy) by the effect that they just heard they had laryngeal cancer, or positively (after radiotherapy) by the fact that they knew the tumour had gone. Nevertheless, if the goal of perceptual evaluation of voice quality is to determine the relation between voice quality and quality of life aspects, such as social contacts and work activities, evaluations by the speakers themselves and their partners are essential.

4.4.3 Voice quality before and after radiotherapy

Before radiotherapy, voice quality was the most deviant. The actual tumour can cause changes in the vocal folds, like stiffness change, mass change, and asymmetry, which can in turn lead to aperiodic vibrations, and incomplete glottal closure. Another explanation for the poor voice quality before therapy may be an increased tension of the vocal folds by the patient in order to compensate for his voice loss. Also, little is known about the effect of microlaryngeal surgery on voice quality. Gray, Hirano & Sato (1993) described the complex layered vocal fold structure extensively. According to Benninger et al. (1994) vocal fold stripping or excisional biopsy rather than limited biopsy for initial diagnosis increased the risk of decreased voice quality.

Shortly after radiotherapy (6 months), voices were characterised by *tension* by the trained raters. Two years and 3-7 years voices were more *rough* compared to control voices according to the trained raters; on the long term (7-10 years) no differences were indicated on read-aloud text, but on sustained /a/, voices were rated as containing more *breathiness* compared to control speakers. The same results were found for the ratings by the naive raters, the speakers themselves and their partners. This trend can hardly be explained by radiation induced complications on normal tissue alone. Acute responses of the normal tissue like mucositis occur within a few weeks of treatment and continue until 4-6 weeks after the end of therapy. Late responses like late oedema or fibrosis occur months or years after radiotherapy (Hill, 1990; Ravasz and Batterman, 1989). Other factors, like the effect of microlaryngeal surgery on the vocal folds as described above can also explain the present trend and will be investigated in Chapter 7.

4.5 Conclusion

Patients before radiotherapy have the most deviant voices. Following radiotherapy voice quality is improved although not completely up to the point of making the irradiated voices indistinguishable from normal voices: voice quality of 55% of the patients is normal, while 45% remain pathological. The four rater groups give supplementary information: perceptual description by trained raters is analytical and supposed to relate best to acoustical and physiological analyses; the evaluative description by naive raters is supposed to represent the judgement of the home environment of the patients (family, friends, colleagues, etc.). The judgements of the speakers themselves and their nearest relatives, their partners, are also of interest: they will be related to questionnaires to investigate the influence of voice quality evaluations on daily life situations. For perceptual descriptions of voice quality, it is recommended to use read-aloud text as speech material.

5

ACOUSTICAL, ELECTROGLOTTOGRAPHIC, AND PERCEPTUAL MEASURES OF PITCH*

ABSTRACT

Speech samples (read aloud text and sustained /a/) of patients with early glottic cancer recorded before and after radiotherapy, and of control speakers, were analysed for various pitch measures: acoustical average fundamental frequency (F0), electroglottographic (EGG) average fundamental frequency, and perceptual pitch evaluations by trained and naive raters, and by the speakers themselves and their partners. The results of pitch measures of trained raters, EGG and acoustical F0 correlated strongly. A principal component analysis resulted in four factors: one factor for "objective" pitch measures (acoustical and EGG F0 on read-aloud text and sustained /a/, and pitch ratings by trained raters on read-aloud text), one factor for perceptual pitch ratings on sustained /a/ speech material (by trained and by naive raters), another factor for perceptual pitch evaluations by naive raters on read-aloud text, and one factor for the pitch evaluations by the speakers themselves and their partners. Because no reliable EGG data could be obtained of 21% of the speakers and because perceptual pitch ratings by trained and naive raters seem to be dependent on voice quality, especially roughness, acoustical analysis of fundamental frequency are recommended in investigating average pitch changes following radiotherapy. Results showed that the patients before radiotherapy differed from the patients 6 months after radiotherapy according to acoustical measured pitch on read-aloud text.

---

*This chapter is a substantially revised and extended version of Verdonck-de Leeuw & Koopmans-van Beinum (1995a&b).

5.1 INTRODUCTION

Pitch is supposed to be one of the parameters that can be influenced by the presence of a tumour or by side-effects of radiotherapy on the vocal fold tissue, such as mucositis and tissue oedema. In the previous chapter, however, no significant differences were found between patients before and after radiotherapy compared to control speakers, according to perceptual pitch ratings. Because a preliminary study on perceptual, electroglottograpic, and acoustical analysis of pitch revealed that perceptual pitch ratings may be influenced by deviant voice quality (Verdonck-de Leeuw & Koopmans-van Beinum, 1995), in the present chapter pitch analysis is investigated in more detail. Acoustical analysis of pitch (average fundamental frequency) is easy and quick to perform. However, acoustic signals contain, next to the fundamental frequency, several strong harmonics due to the resonant frequencies of the vocal tract, which may make pitch extraction difficult. Electroglottographic (EGG) signals represent vocal fold activity directly, since they reflect the variation in electrical conductance, due to opening and closing of the vocal folds, between two electrodes placed on either side of the thyroid cartilage. EGG signals have a weak harmonic structure and are therefore preferable to acoustical analysis in determining fundamental frequency (Fourcin, 1993; Askenfelt et al., 1980). However, recordings of EGG signals can be difficult to obtain. Short necks with a lot of subcutaneous fat, electrode placement, head movements during recordings, and even heartbeat or vascular pulses can influence EGG recordings (Colton & Conture, 1990; Orlikoff & Baken, 1989). Furthermore, most research on EGG signals so far has been done on normal subjects in order to parametrize the source signal in terms of open and closed phase, steepness of the opening and closing phase of the vocal folds, and in order to investigate fundamental frequency perturbation (jitter). Studies on EGG analysis of pathological voices are scarce. Lablance et al. (1992) found similar acoustical and EGG jitter results over 80 dysphonic voices; however, they prefer acoustical fundamental frequency analysis over EGG, because no reliable recordings could be made of 22% of the subjects. Vieira et al. (1995) argued that EGG analysis is a very precise and robust method to estimate fundamental frequency of pathological voices. They also found similar acoustical and EGG jitter measures (and thereby pitch measures) for sustained /a/ vowels, but not for /i/ or /u/ vowels (Vieira et al., 1997).

The aim of the present chapter was to assess which pitch measure to use in determining pitch changes following radiotherapy. To that end, acoustical and EGG average fundamental frequency measures were investigated. Results of these "objective" pitch measures were compared to the perceptual pitch measures, as described in detail in the previous chapter.

5.2 METHOD

5.2.1 Speakers

The subject sample is described in detail in Chapter 3. Briefly, the sample consisted of 60 male patients with early glottic cancer, who were treated with radiotherapy. There was a longitudinal group of 10 patients of whom voice samples were recorded before, as well as 6 months and 2 years after radiotherapy. Furthermore, data were collected of 5 separate groups of 10 patients each, before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation. Finally, recordings were made of 20 control speakers. In summary, there were in total 100 speaker cases (longitudinal group (n=30), separate groups (n=50), and control group (n=20).

5.2.2 Speech material/recordings/preprocessing

The same speech material as described in Chapter 4.2.2 was used. Analyses of sustained /a/ are more suitable for acoustical analysis techniques, since most techniques require more or less stable speech signals. Running speech fragments were used as well, because especially with respect to pitch characteristics, these are more representative of conversational speech. Here again, fragments of read-aloud text were chosen, because it was important to avoid variance between speakers caused by unequal texts.

The digitised voice samples were used for acoustical analyses by means of PRAAT, a speech signal processing software package developed by Boersma & Weenink (1996), implemented on an Iris Indigo R4000. From these digitised voice samples, two digital tapes (sustained /a/ and read-aloud text) with all 100 voice samples were made for acoustical analyses by means of MDVP, the Multi Dimensional Voice Program developed by Kay Elemetrics, which was implemented on a PC (Deliyski, 1993).

Furthermore, two analogue tapes (read-aloud text and sustained /a/) were composed for perceptual analyses, as decribed in Chapter 4.4.2.

5.2.3 Acoustical pitch analysis

In earlier research we used an accurate method to analyse pitch and the harmonics-to-noise ratio of pathological and control voices, implemented in the speech signal processing program PRAAT (Boersma, 1993; Verdonck-de Leeuw & Boersma, 1996). In order to meet clinical needs, in the present study acoustical analyses were also performed using the Multi Dimensional Voice Program (MDVP), implemented in the Kay Elemetrics hardware (Model 4300).

To be sure that pitch extraction is accurate, we compared the results of average Fundamental Frequency (F0) measured by means of PRAAT and MDVP. Both methods use autocorrelation for pitch extraction (Boersma, 1993; Delyiski, 1993). The main difference between the two algorithms is that MDVP uses non-linear signal-coding (dividing the sampled signal in + or - parts) with a window length of 30 ms, while PRAAT uses real numbers of the sampled signal with a window length of 60 ms.

On sustained /a/, Pearson correlation of the average fundamental frequency over all 100 speaker samples between MDVP and PRAAT results was .89. Visual inspection of the sound signals revealed that PRAAT was accurate in 99 of the 100 cases. MDVP showed erratic results in 11 cases. More specifically, in cases where the speaker was not able to produce a stable vowel, octave jumps were made. Furthermore, in cases of irregular voice onset, the voice onset was considered as unvoiced and therefore left out of parameter calculations. An example is given in figure 5.1. Since MDVP uses fixed F0 tracking strategies and does not permit access to the extraction parameters, we decided to cut off all onset parts and to segment only that steady part of the vowel where F0 determined visually and by means of MDVP were equal. In two cases, the length of the signal became rather short (0.25 and 0.38 seconds), in all other cases the length was ca 1 second. In cases with irregular onset, the onset measured by PRAAT was also set as unvoiced and left out of calculation of F0.

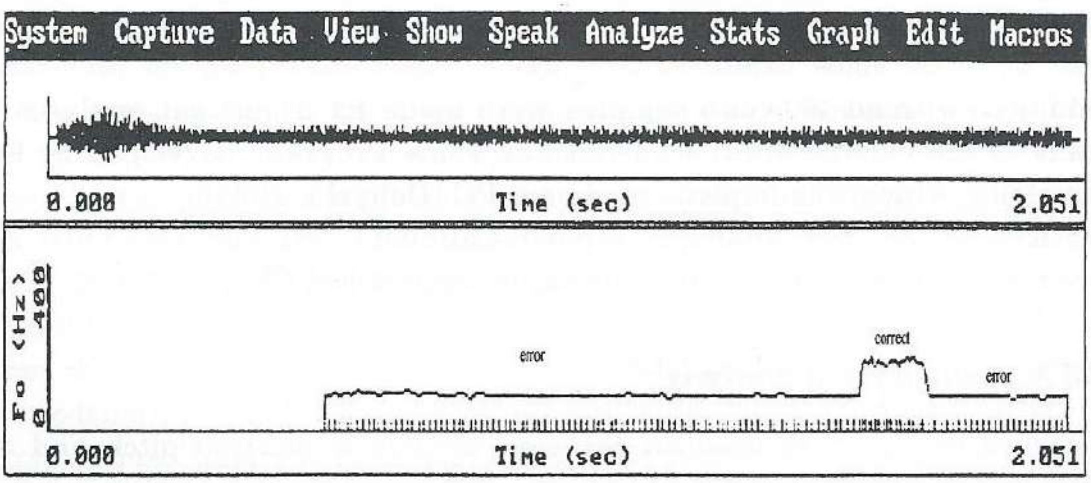

Figure 5.1. Example of a sustained /a/ from a speaker before radiotherapy, where errors (octave jumps) were made and where the onset erroneously is considered as unvoiced by the pitch algorithm used in MDVP.

On read-aloud text, Pearson correlations of the average fundamental frequency over all 100 speakers between MDVP and PRAAT results was .97. However, visual inspection revealed inaccurate pitch results. In 49 of the 100 MDVP cases, the highest F0 found was over 400 Hz (figure 5.2). Visual inspection of these cases showed clear pitch extraction mistakes in a schwa or a reduced vowel. For example, in many cases the F0-extraction in the word "gebroken" (English "broken") was erratic in the first and last schwa: /χəbrœkə/. Intensity seems to be a factor in this matter. All apparent errors (approximately at most 3% of a voice sample) were deleted from the signal by hand. Because final results of PRAAT and MDVP correlated almost perfectly, results from PRAAT were used for further investigation.

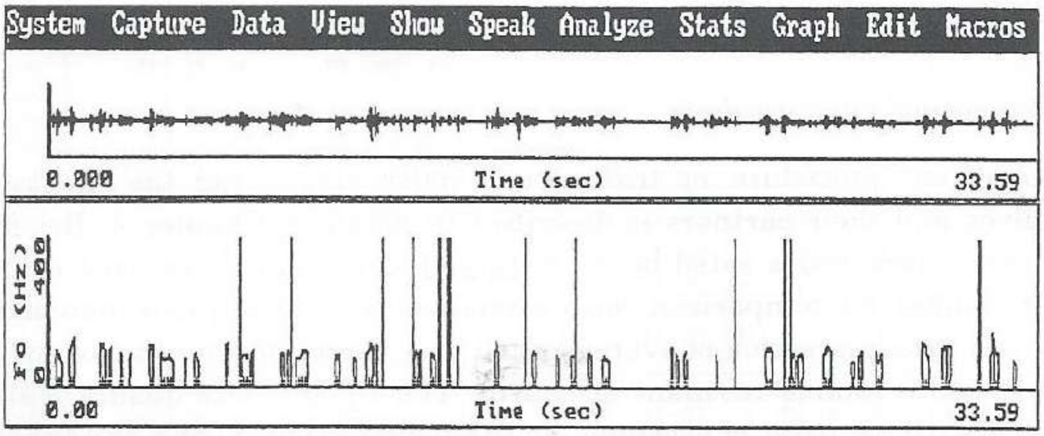

Figure 5.2. Example of read-aloud text from a speaker before radiotherapy. Pitch errors were made above 400 Hz by the pitch algorithm used in MDVP.

5.2.4 Electroglottographic pitch analysis

Two electroglottographic procedures were used, abbreviated as EGG and ELG. By means of an electroglottograph (Stöpler Teltec GFA06) average fundamental frequency was measured on read aloud text (EGG). The same text was used as was recorded for perceptual and acoustical analyses described above, but the recording itself was another realisation. The speakers read aloud the text for about 5 minutes while up to 1000 voiced samples were analysed and averaged. We experienced difficulties in obtaining EGG data for 21% of the speakers. These cases were left out of consideration, resulting in 79 speaker cases (instead of 100).

Furthermore, a portable Laryngograph (Laryngograph Ltd., London) was used to record electrolaryngeal signals (ELG). The ELG signals were recorded simultaneously with microphone sound signals in a sound treated room, using a Philips N8214 microphone and a Casio DAT-recorder. Recording level was adjusted for each speaker to optimise signal-to-noise ratio, and then kept constant for that speaker. For 15 speakers, the Laryngograph equipment was not available yet at the start of the project, leaving 85 speaker cases. Speakers were asked to produce a sustained /a/ at a comfortable pitch and loudness for about 3 seconds. For 16 speakers no reliable signals could be obtained, which is 19% of the 85 speakers, leaving 69 speaker cases. Microphone and ELG signals of these 69 speakers on sustained /a/ were digitised by means of the Sound Editor of an Iris Indigo R4000 with a sample frequency of 48 kHz and a 16 bit resolution. The sustained /a/ speech material was segmented from a stable part of the vowel for about 1 second. The microphone and ELG segmented signals were bandpass filtered between 30 Hz and 2 kHz in order to remove low as well as high frequency noise. These filtered signals were analysed by means of PRAAT (see section 5.2.3) to determine average fundamental frequency.

5.2.5 Perceptual pitch analysis

The judgement procedure by trained and naive raters and the speakers themselves and their partners is described in detail in Chapter 4. Briefly, the 13-point pitch scales rated by the 3 trained raters that were used in the present chapter for comparison with acoustical and EGG pitch measures were: *pitch* (the impression of average pitch level) and *sonority* (the extent to which the voice sounds resonant or sharp). The 7-point voice quality scales *breathiness* (the amount of escaping air through the glottis) and *roughness* (the amount of aperiodic vibration, resulting in rough and rasping quality) were used in the present chapter to investigate the effect of deviant voice quality on pitch perception. The 7-point scales rated by 20 naive raters that were used in the present chapter for comparison with acoustical and EGG pitch measures were: *high -- low* and *shrill -- deep*. These two scales *high -- low* and *shrill -- deep* rated by the speakers themselves and their partners were used as well in the present chapter. They were asked to evaluate the voice of the speaker at home by filling out the form independently from each other. Forms of 98 speakers and of 93 partners were received.

Interrater and intrarater reliability was calculated for the trained and naive raters. For the trained raters, alpha was .83 and .90 for *pitch*, .82 and .87 for *sonority*, .93 and .92 for *breathiness*, .93 and .91 for *roughness* on read-aloud text and sustained /a/ respectively. For the naive raters, alpha was .91 and .93 for *high -- low* and .89 and .90 for *shrill -- deep* on read-aloud text and sustained /a/, respectively. Intrarater reliability was moderate for the naive raters to high for the trained raters: percentages of first ratings that were within 1 scale value of the second ratings (of the same items) were above 85% for the trained raters. Percentages were above 70% for the naive raters; these percentages increased to above 80% for first ratings within 2 scale values of the second ratings. Given these reliabilities, only mean ratings of each rater group per speaker were taken into account in further analyses.

5.2.6 Statistical analyses

Univariate analysis of variance with repeated measures (the longitudinal speaker group) and univariate analysis of variance without repeated measures (six separate speaker groups) were used (Rietveld & Van Hout, 1993; Kirk, 1982). Posthoc tests (Tukey) were used to test the significance of differences between separate speaker groups. Pearson correlations were calculated to investigate relations between perceptual, acoustical, and EGG/ ELG measures. Principal Component Analysis was used to decompose the correlation matrix and to obtain more insight in distinct dimensions of pitch. Multiple regression analyses were carried out in order to investigate if perceptual pitch measures were influenced by deviant voice quality.

5.3 RESULTS

5.3.1 Relations between various pitch measures

In order to investigate relations between acoustical, EGG/ELG, and perceptual pitch measures, Pearson correlations were calculated. To obtain more insight in distinct pitch dimensions, a Principal Component Analysis (PCA) was used to decompose the correlation matrix into varimax rotated factors. With the criterion 'eigenvalue greater than one', the PCA produced 4 factors, together explaining 74% of the total variance (table 5.1). On the basis of the factor loadings (>.50) the factors were labelled as:

1. objective pitch (acoustical pitch, EGG/ELG on text and /a/, and *pitch* and *sonority* by trained raters on read-aloud text),
2. perceptual pitch on sustained /a/ (*pitch* and *sonority* by trained raters and *low-high* and *deep-shrill* by naive raters),
3. perceptual pitch on read-aloud text (*low-high* and *deep-shrill* by naive raters),
4. self-ratings (*low-high* and *deep-shrill* by the speakers and their partners).

The correlation matrix is given in table 5.2. Pitch evaluations by the speakers and their partners did not correlate significantly with any of the other pitch measures and are therefore not included in the table.

Relations between acoustical and EGG/ELG pitch analysis on read-aloud text and sustained /a/, and the perceptual pitch ratings by trained raters on

Table 5.1. Factor loadings (>.50) and percent of total variance explained per factor after varimax rotation over all 100 speakers for acoustical pitch on read-aloud text and on a sustained /a/, EGG-pitch on read-aloud text, ELG pitch and filtered acoustical pitch on sustained /a/, pitch evaluations by 3 trained and by 20 naive raters on read-aloud text and on a sustained /a/, and pitch evaluations by the speakers and by the partners.

	factors	1	2	3	4
% of total variance explained		28%	18%	15%	13%
pitch measure	material				
acoustical pitch	text	.84			
acoustical pitch	/a/	.75			
EGG pitch	text	.69			
ELG pitch	/a/	.79			
acoustical pitch	/a/	.84			
pitch trained raters	text	.71			
sonority trained raters	text	.83			
pitch trained raters	/a/		.83		
sonority trained raters	/a/		.70		
low-high naive raters	text			.90	
deep-shrill naive raters	text			.85	
low-high naive raters	/a/		.81		
deep-shrill naive raters	/a/		.80		
low-high speakers					.78
deep-shrill speakers					.78
low-high partners					.60
deep-shrill partners					.72

Table 5.2. Significant ($p<0.05$) pairwise Pearson correlations (x 100) between acoustical pitch on read-aloud text and sustained /a/ (n=100), EGG pitch on read-aloud text (n=79), filtered ELG and filtered acoustical pitch on sustained /a/ (n=69), and perceptual pitch ratings by 3 trained raters on *pitch* and *sonority* and 20 naive raters on *high -- low* and *shrill -- deep* on read-aloud text and sustained /a/ (n=100). High correlations (r>.50) are printed bold. Pitch evaluation by the speakers and their partners did not correlate significantly with any of the other pitch measures and are therefore not included.

			2	3	4	5	6	7	8	9	10	11	12	13
1	acoustical	txt	**52**	**72**	**73**	**76**	**73**	**72**	22	29	46	**57**	35	39
2	acoustical	/a/		32	**77**	**80**	40	41	47	35	27	35	**55**	**57**
3	EGG	txt			**51**	**54**	**68**	**66**		27	38	47	22	29
4	ELG filtered	/a/				**90**	**56**	**54**	28	27	29	40	45	44
5	acoustical filtered	/a/					**59**	**57**	25	25	32	40	46	47
6	trained *pitch*	txt						**70**	21	22	**51**	**57**	32	37
7	trained *sonority*	txt								29	29	48	21	35
8	trained *pitch*	/a/								**60**	20	24	49	49
9	trained *sonority*	/a/										21	29	36
10	naive *high-low*	txt										**90**	49	47
11	naive *shrill-deep*	txt											48	49
12	naive *high-low*	/a/												**93**
13	naive *shrill-deep*	/a/												

read-aloud text are clear; they loaded highly on the same factor and correlations between these pitch measures were high (r>.50), as can be seen in table 5.1 and table 5.2, respectively. Perceptual pitch ratings on sustained /a/, perceptual pitch ratings on read-aloud text by naive raters, and self-ratings seem to be separate dimensions of pitch.

5.3.2 Pitch changes following radiotherapy

In order to assess pitch changes following radiotherapy, analyses of variance were carried out on all perceptual, acoustical, and electroglottographical pitch measures. No significant differences (p<0.05) for any of the pitch measures were found between the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation, and the control speakers), nor for the longitudinal speaker group (before radiotherapy, 6 months after, and 2 years after radiotherapy).

At a lower significance level (p<0.10), the results of acoustical pitch on read-aloud text revealed differences between the speaker groups. For the separate speaker groups, posthoc tests after the analysis of variance (F=2.34, p=0.06) revealed that differences between patients before radiotherapy and patients 6 months after radiotherapy were significant (p<0.05). The same results were found for the longitudinal groups (F=3.18, p=0.07). Patients before radiotherapy tend to have higher pitched voices, while patients 6 months after radiation tend to have lower pitched voices, compared to the control speakers (figure 5.1).

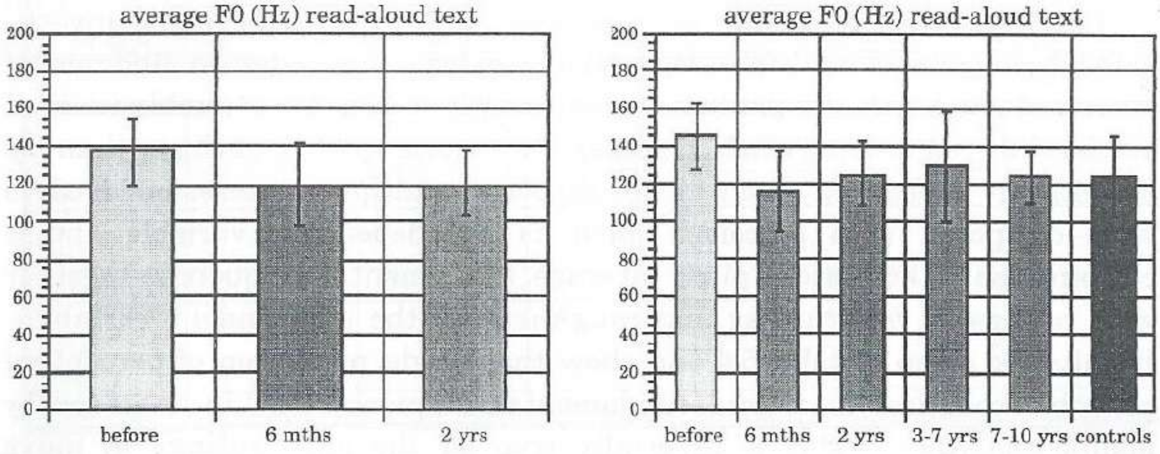

Figure 5.1. Mean and standard deviation of average fundamental frequency (acoustical pitch) in Hz on read-aloud text for the longitudinal speaker groups (before radiotherapy and 6 months and 2 years after radiotherapy) at the left and for the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after and 7-10 years after radiotherapy, and control speakers) at the right.

5.3.3 Perceptual pitch measures and voice quality

The previous two sections revealed that although highly correlating, acoustical pitch measures on read-aloud text showed differences between patients before and 6 months after radiotherapy, while perceptual pitch ratings by trained raters on read-aloud text did not. A possible explanation, namely the effect of voice quality on perceptual pitch ratings is investigated in the present section. Single correlations between acoustical and perceptual pitch measures were calculated, separately for pathological and normal voices, and multiple regression analysis were carried out. EGG and ELG measures were left out of consideration because no reliable recordings could be made of about 20% of the speakers.

In order to assess which speakers were normal or pathological, ratings by trained raters on *breathiness* and *roughness* were involved. During the rating sessions, the trained raters agreed to score lower than 4 when *breathiness* or *roughness* was considered to be pathological; they scored equal or higher than 4 in the case of normal *breathiness* or *roughness*. Results revealed that 64 speaker cases were normal, while 36 cases fell in the pathological range, both on *breathiness* and *roughness*. Variances of pathological and normal groups were comparable.

Correlations between perceptual and acoustical pitch measures were compared for normal voices and pathological voices. Results are given in table 5.3 and reveal that correlations between perceptual and acoustical measured pitch are always lower for pathological voice quality than for normal voice quality, indicating that perceptual pitch ratings are dependent on voice quality of speakers. However, this effect was fractional for the trained raters (*pitch*) on read-aloud text.

The effect of voice quality on perceived pitch was furthermore investigated by means of multiple regression analysis, in order to find out if perceived pitch can be predicted significantly better by a combination of acoustically measured pitch together with voice quality ratings than by acoustically measured pitch alone. Stepwise multiple regresssion models were composed with perceived pitch as the dependent variable and a combination of acoustical pitch (average fundamental frequency) together with ratings on *breathiness* and *roughness* as the independent variables. Results are given in table 5.4 and show that single prediction of perceptual pitch by acoustical measured fundamental frequency (top) is improved by adding *roughness*. This is especially true for the pitch ratings by naive raters. Furthermore, it is striking that *sonority* as judged by trained raters on sustained /a/ is influenced by *breathiness*.

Table 5.3. Pearson correlations ($p<0.01$) over 100 speaker cases (64 scoring normal on *breathiness* or *roughness*, 36 scoring pathological) between acoustical pitch and perceptual pitch by 3 trained and 20 naive raters on read-aloud text and on a sustained /a/.

Perceptual pitch	acoustical pitch							
	read-aloud text				sustained /a/			
	breathiness		*roughness*		*breathiness*		*roughness*	
	path.	normal	path.	normal	path.	normal	path.	normal
trained *pitch*	.71	.73	.72	.75	.46	.48	.37	.53
trained *sonority*	.56	.73	.60	.73	.26	.44	.30	.34
naive *high -- low*	.55	.66	.51	.59	.55	.60	.49	.64
naive *shrill -- deep*	.60	.66	.62	.63	.50	.63	.45	.67

Table 5.4. Results in terms of percentage of variance explained of single correlations between acoustical and perceptual pitch measures (top) and of stepwise multiple regression analysis (bottom) over 100 speaker cases of prediction of perceptual pitch ratings by 3 trained (*pitch, sonority*) and 20 naive raters (*high -- low, shrill -- deep*) by acoustical measured pitch together with voice quality ratings by trained raters (*roughness, breathiness*) and naive raters (*breathy -- not breathy, rough -- not rough*) on read-aloud text (at the left)and on a sustained /a/ (at the right).

	read-aloud text				sustained /a/			
	trained raters		naive raters		trained raters		naive raters	
% var. expl.	*pitch*	*sonority*	*high-low*	*shrill-dp*	*pitch*	*sonority*	*high-low*	*shrill-dp*
single prediction								
acoustical pitch	53	52	21	32	22	12	30	32
multiple prediction								
acoustical pitch	53	52	21	32	22	12	30	32
+ roughness			19	5			11	
+ breathiness						31		
total	53	52	40	37	22	43	41	32

5.4 Discussion

The main aim of the present chapter was to assess which pitch measure to use in order to investigate pitch changes following radiotherapy. As objective measures pitch extraction results of acoustical and EGG voice recordings were compared. Results on the same sustained /a/ showed that acoustical and EGG correlated almost perfectly; only small differences (less than 1 Hz) were found. This means that average fundamental frequency can be analysed reliably by means of either method. A disadvantage of EGG analysis however, was that no reliable EGG recordings could be made of about 20% of the speakers. Similar percentages (15-21%) were found in the literature (Colton & Conture, 1996; Lablance et al., 1992).

When these objective pitch measures were compared to subjective perceptual pitch ratings, it appeared that strong relations were found between acoustical and EGG pitch analysis and pitch evaluations by the trained raters on read-aloud text. The naive raters judged pitch differently on read-aloud text. Also, on sustained /a/ speech material there was a clear difference between acoustical and perceptual pitch measures, both for trained and naive raters. Furthermore, it appeared that the speakers themselves and their partners judged their voices differently from the trained and naive raters. This may be due to the different way they were asked to evaluate the voices: one voice at a time instead of all voices at one time from tape recordings, as the raters did.

Correlations between perceptual and acoustical measured pitch were always found to be lower for pathological voice quality than for normal voice quality, indicating that pitch ratings are influenced by deviant voice quality. This effect was fractional for the trained raters on read-aloud text, but considerable for the other perceptual ratings. Especially rough voices were perceived as low pitched while fundamental frequency was not that low. This may be explained by strong subharmonics and pitch perturbation which may interfere with fundamental frequency, which is known as a problem in investigating pathological voice quality (Hammarberg & Gauffin, 1995; Titze, 1994). Apparently, trained ears and accurate pitch extraction methods are not diverted by these aspects, while naive raters can be fooled.

Finally, it can be concluded that there was no statistically significant effect of radiotherapy on the various pitch measures. It seems that patients before radiation have higher pitched voices compared with patients six months after radiotherapy, measured by means of acoustical pitch analysis. This may be due to mechanical effects of the tumour on the vocal folds or the effect of microlaryngeal surgery that most of the patients have undergone before radiation. Another explanation may be an increased tension of the vocal folds by the patient in order to compensate for his voice loss. Shortly after radiation, lower pitched voices may be explained by tissue oedema as a side-effect of radiotherapy.

5.5 Conclusion

Since no reliable EGG data could be obtained of 21% of the speakers and perceptual pitch ratings seem to be dependent on voice quality (especially *roughness*), acoustical analysis of pitch is recommended in investigating pitch changes following radiotherapy.

6

ACOUSTICAL VERSUS PERCEPTUAL ANALYSIS OF VOICE QUALITY*

Abstract

Acoustical analyses of voice quality, such as perturbation and harmonics-to-noise ratio, on read-aloud text and on a sustained /a/ of 60 male patients with early glottic carcinoma recorded before and after radiotherapy, and of 20 control speakers were explored in the present chapter. Results were compared to perceptual scale judgements on breathiness, roughness, and tension by trained and naive raters as described in Chapter 4. Differentiation between speaker groups was stronger for perceptual analyses than for acoustical analyses of voice quality. Single correlations between perceptual and acoustical parameters appeared to be low. Results of multiple regression analyses (prediction of one perceptual parameter by a set of acoustical parameters) showed that at most 47% of the perceptual variance was explained, which is considered to be moderate. It is concluded that acoustical analyses give some insight in major differences between normal and deviant voices but do not reveal small voice quality differences. More research is needed to improve acoustical analyses on running speech and voice onset. For the moment, perceptual analyses by trained raters on running speech are recommended in investigating voice quality following radiotherapy in the case of separate speaker groups. For the longitudinal study of voice quality of a patient during therapy, acoustical analyses are easy and quick to perform and come close to the judgements by naive raters.

---

*This chapter is a substantially revised and extended version of Verdonck-de Leeuw & Boersma (1996).

6.1 Introduction

There are four major applications of automatic acoustical analysis of voice quality: screening for laryngeal pathology of populations known to be at risk, priority assessment of patients with evidence of serious pathology, diagnostic support, and monitoring a patient during therapy (Laver et al., 1986). Also Hammarberg (1997) argues that quantifiable and objective measures of voice quality are necessary in the clinic for the evaluation of therapy and surgery effects. In Chapter 5, it was already stated that acoustical analysis of pitch is essential for diagnostic support. The main aim of this Chapter 6 was to investigate the feasibility of acoustical analysis in monitoring patients during therapy. If feasible, this would mean that voice quality can be investigated in a relatively easy and quick way by means of acoustical analyses rather than the time-consuming perceptual analyses. For clinical purposes, this would be an important advantage. In earlier research (Boersma, 1993; Verdonck-de Leeuw & Boersma, 1996), the speech signal processing program PRAAT was used for acoustical analysis of voice quality. However, in order to meet clinical needs, in the present study acoustical analyses were also performed using the Multi Dimensional Voice Program (MDVP), implemented in the Kay Elemetrics hardware (Model 4300). MDVP is an extension of the Computerised Speech Lab (CSL) and is used more and more (also in the Netherlands) as a clinical tool by otolaryngologists and speech therapists. However, few publications about results are available.

The starting point of the present chapter was the investigation whether the same trend that was found by means of perceptual analyses is found by means of acoustical analyses as well. In Chapter 4, voice quality of patients before and after radiotherapy compared with normal speakers was investigated in detail by means of perceptual ratings by trained and untrained raters on read-aloud text and on sustained /a/. Although reliability of listeners was comparable for read-aloud text and sustained /a/, speaker group differences were more prominent on read-aloud text than on sustained /a/. Low correlations were found ($r<.50$) between the two types of speech material. Furthermore, ratings on *breathiness, roughness* and *tension* appeared to differentiate speaker groups best. A trend was found that patients before radiotherapy had the most deviant voices; voices following radiotherapy sounded less deviant, but were significantly different from voices of control speakers.

Furthermore, results of acoustical analyses were also compared directly with perceptual ratings by trained and naive listeners in order to obtain more insight in production and perception aspects of voice quality on read-aloud text and on sustained /a/.

6.2 Methods

6.2.1 Speakers

The study sample is described in detail in Chapter 3. Briefly, there was a longitudinal group of 10 patients of whom voice samples were recorded before, as well as 6 months and 2 years after radiation. Furthermore, data were collected of 5 separate groups of 10 patients each, before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation. Finally, recordings were made of 20 control speakers without any known vocal defects. In summary, there were in total 100 speaker cases (longitudinal group (n=30), separate groups (n=50), and control group (n=20).

6.2.3 Acoustical analyses

The same speech material as described in Section 4.2.2 was used: read-aloud text and sustained /a/. The digitised voice samples were used for acoustical analyses by means of PRAAT and MDVP, as described in Section 5.2. MDVP determines various voice parameters, mainly covering fundamental frequency (pitch), frequency and amplitude perturbation (jitter and shimmer), voice breaks and irregularities, subharmonic components, and noise and tremor aspects. The MDVP procedures are described in detail in Deliyski (1993) and in the Operations Manual of MDVP (1993). An overview of all acoustical parameters is given in table 6.1.

The major problem in investigating pathological voices acoustically, is the periodicity as such. According to Titze (1994) it appears that there are three categories of sound signals: periodic with small random perturbations, periodic with subharmonic structure and modulation, and nonperiodic; perturbation measures can be applied only to the first category. In order to be sure that calculations (which are based on short-term autocorrelation periodicity analysis) of the MDVP time-domain parameters (fundamental frequency, perturbation, voice breaks, voice irregularities, and subharmonic components) are accurate, results of average Fundamental Frequency (F0) were investigated in detail in Chapter 5. Briefly, on sustained /a/ we decided to cut off all onset parts of sustained /a/ vowel samples and to segment only that steady part of the vowel, where F0 determined visually and by means of MDVP were equal. In two cases, the length of the signal became rather short (0.25 and 0.38 seconds), in all other cases the length was ca 1 second.

Table 6.1. Overview of the manually measured DVO, of PRAAT, and of MDVP voice parameters with their descriptions in the categories Fundamental Frequency, Frequency Perturbation, Amplitude Perturbation, Voice Breaks, Subharmonic Components, Voice Irregularities, Noise, and Tremor (Noise and Tremor parameters were calculated for sustained /a/ only). Because the additional MDVP parameters JITa, ShdB, NVB, NSH, NUV, Fftr and Fatr are equivalent to JITT, Shim, DVB, DSH, DUV, FTRI and ATRI, respectively, the first mentioned parameters are left out of consideration in our actual data analyses.

DVO	Duration Voice Onset (ms)
PRAAT	
F0	Median fundamental frequency (Hz)
HNR	Median (sustained /a/) or 90% (read-aloud text) Harmonics to Noise Ratio (dB)
MDVP	
Fundamental Frequency	
F0	Average Fundamental frequency (Hz)
Fhi	Highest Fundamental frequency (Hz)
Flo	Lowest Fundamental frequency (Hz)
stdF0	Standard deviation of the Fundamental frequency (Hz)
PFR	Phonatory F0 range (semitones); distance between highest and lowest F0
Frequency Perturbation	
JITa	Absolute jitter (ms); cycle to cycle pitch perturbation
JITT	Jitter percent (%); cycle to cycle pitch perturbation
RAP	Relative average perturbation (%); short term pitch perturbation over 3 cycles
PPQ	Pitch period perturbation quotient (%); short term perturbation over 5 cycles
sPPQ	Smoothed pitch period perturbation quotient (%); long term perturbation over 55 cycles
vF0	Fundamental frequency coefficient variation (%); ratio STD-F0
Amplitude Perturbation	
ShdB	Absolute shimmer (dB); cycle to cycle amplitude perturbation
Shim	Shimmer percent (%); cycle to cycle amplitude perturbation
APQ	Amplitude perturbation quotient (%); short term perturbation over 11 cycles
sAPQ	Smoothed amplitude perturbation quotient (%); long term perturbation over 55 cycles
vAM	Peak-to-peak amplitude coefficient of variation (%); standard deviation of perturbation
Voice Breaks	
DVB	Degree of voice breaks (%); ratio of the length of voice breaks areas to the total voiced sample length
NVB	Number of voice breaks
Subharmonic Components	
DSH	Degree of subharmonics (%); ratio of the number of windows with incorrect sub-harmonic period classification to the total number of windows
NSH	Number of subharmonics
Voice Irregularities	
DUV	Degree of irregular vocalisation (%); the ratio of the number of windows classified as unvoiced to the total number of windows
NUV	Number of unvoiced segments
Noise	
NHR	Noise to Harmonics Ratio (dB); ratio of inharmonic energy in the range 1500-4500 Hz to the harmonic spectral energy in the range 70-4500 Hz
VTI	Voice Turbulence Index (dB); ratio of the inharmonic energy in the range 2800-5800 Hz to the harmonic spectral energy in the range 70-4500 Hz
SPI	Soft Phonation Index (dB); ratio of the harmonic energy in the range 70-1600 to the harmonic energy in the range 1600-4500 Hz for the first group of windows
Tremor	
FTRI	F0-tremor Intensity Index (%); value of the global maximum of the average autocorrelation curve and the corresponding position F0-tremor frequency
ATRI	Amplitude Tremor Intensity Index (%); value of the global maximum of the average autocorrelation curve and the corresponding position F0-tremor amplitude
Fftr	F0-tremor frequency (Hz)
Fatr	Amplitude-tremor Frequency (Hz)

In order to get some insight in vowel onset aspects, we measured manually the duration of the vowel onset (DVO); DVO was defined as the duration from the beginning of of the vowel to the point that the signal becomes regular concerning periodicity. Examples of short and long DVO are given in figure 6.1.

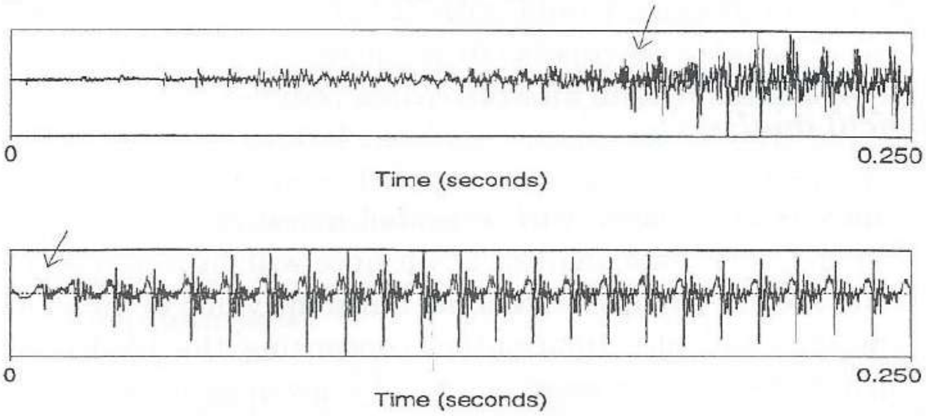

Figure 6.1. Example of a long (at the top) and a short (at the bottom) voice onset (Duration Voice Onset (DVO)) of a sustained /a/ of a patient with early glottic cancer before radiotherapy and a control speaker, respectively. Arrows indicate the end of the voice onset.

6.2.3 Perceptual analyses

The scale judgement procedure by trained and naive raters on read-aloud text and sustained /a/ speech material is described in detail in Chapter 4. The three scales, rated by the trained raters, that differentiated the various speaker groups best and that were used in the present chapter for comparison with acoustical parameters were: *tension* (the impression of the muscle tension in the vocal folds), *breathiness* (the amount of air escaping through the glottis), and *roughness* (the amount of aperiodic vibration, resulting in rough and rasping quality). The scales *breathy -- not breathy*, *rough -- not rough*, and *tense -- relaxed,* rated by the naive raters, were used in the present chapter for comparison with the judgements of the trained raters and with the acoustical parameters.

Interrater reliability was high: Cronbach's alpha was .93 and .92 for *breathiness*, .93 and .91 for *roughness*, and .95 and .88 for *tension* for the trained raters on read-aloud text and sustained /a/ respectively. For the naive raters, alpha was .94 and .93 for *breathy -- not breathy*, .94 and .92 for *rough - - not rough*, and .86 and .87 for *tense -- relaxed* on read-aloud text and sustained /a/, respectively. Intrarater reliability was moderate for the naive raters to high for the trained raters: percentages of first ratings that were within 1 scale value of the second ratings (of the same items) were above 85% for the trained raters. Percentages were above 56% for the naive raters; these percentages increased to above 75% for first ratings within 2 scale values of the second ratings. Given these reliabilities, mean ratings of the trained and mean ratings of the naive raters per speaker were taken into account in further analyses.

6.2.4 Statistical analyses

Univariate analysis of variance with repeated measures (the longitudinal speaker group) and univariate analysis of variance without repeated measures (six separate speaker groups) were used. Since the value of the F-statistic or the p-value only provides information concerning the likelihood that speaker group differences are present or not, η^2 is given as an indicator of the strength of an effect. The η^2 statistic is defined as: $SS_{between}$ / SS_{total}, and can be interpreted as the proportion of the total variability in the dependent variable (i.e. the scale) that can be accounted for by the independent variable (i.e. the speaker groups) (Rietveld & Van Hout, 1993; Kirk, 1982). Posthoc tests (Tukey) were used to describe differences between the separate speaker groups. Pearson correlations were calculated to describe direct relations between the perceptual and acoustical voice parameters. Multiple regression analyses (step-wise) were carried out in order to investigate if a set of multiple acoustical parameters can predict a particular perceptual aspect better than single correlations.

6.3 Results

6.3.1 Differentiation among speaker groups

The first step was to assess whether acoustical analyses of voice quality show the same trend for the various speaker groups as was found for the earlier derived perceptual analyses (Chapter 4). Results of analyses of variance on the longitudinal speaker group and the six separate speaker groups are given in table 6.2 for acoustical parameters (top) and perceptual parameters by trained and naive raters (bottom). Results show that perceptual ratings by trained raters were the strongest indicators of a trend to be present, followed by the ratings by naive raters. The longitudinal groups were differentiated more clearly than the separate groups. Speaker groups were differentiated more strongly on read-aloud text than on sustained /a/ speech material.

Acoustically, results for the longitudinal speaker group concerning stdF0 (deviation of fundamental frequency), and JITT, RAP, PPQ and sPPQ (frequency perturbation measures), came close to results of naive raters.

Posthoc tests (Tukey) on these acoustical measures on the separate groups revealed that the same trend was found as for perceptual measures, namely that voice quality before radiotherapy was the most deviant, and that voice quality after radiotherapy improved, but still differed significantly from control speakers. As examples, histograms of PPQ and DVO are given in figure 6.2 and 6.3, respectively.

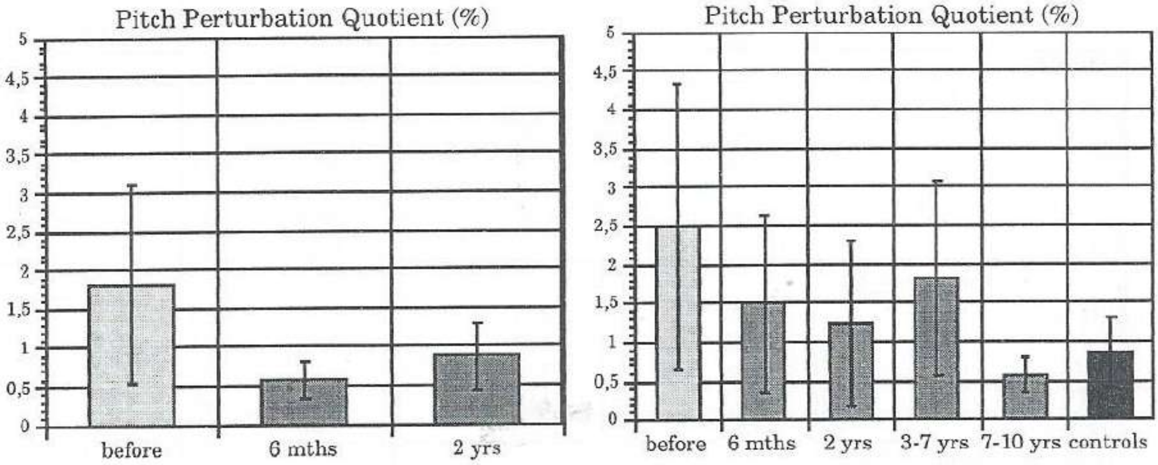

Figure 6.2. Mean and standard deviation of standard deviation of the Pitch Perturbation Quotient (PPQ) in percentage on sustained /a/ for the longitudinal speaker group (before radiotherapy, and 6 months and 2 years after radiotherapy) at the left and for the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiotherapy, and control speakers) at the right.

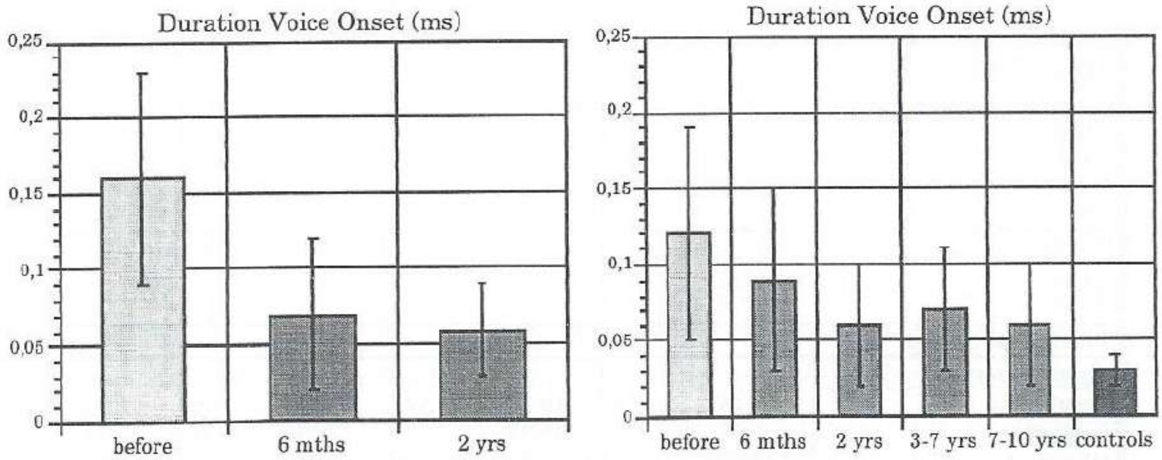

Figure 6.3. Mean and standard deviation of standard deviation of the Duration Voice Onset (DVO) in ms on sustained /a/ for the longitudinal speaker group (before radiotherapy, and 6 months and 2 years after radiotherapy) at the left and for the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiotherapy, and control speakers) at the right.

Table 6.2. Results of analyses of variance (F-statistic, $p<0.05$ is printed bold, whereas η^2 is printed for the significant F-statistics) for the longitudinal speaker group at the left hand, and for the 6 separate speaker groups at the right hand of this table, of the manually measured Duration Voice Onset, PRAAT and MDVP parameters, and perceptual parameters by trained and naive raters on read-aloud text and sustained /a/ (Duration Voice Onset, Noise and Tremor parameters were calculated for sustained /a/ only). An asterisk indicates that analysis of variance was not feasible, since one of more groups showed no variance.

Voice Parameters	longitudinal speaker groups				separate speaker groups			
	sustained /a/		read-aloud text		sustained /a/		read-aloud text	
	F	η^2	F	η^2	F	η^2	F	η^2
Manually								
Duration Voice Onset	**5.76**	.30			**2.40**	.16		
PRAAT								
F0	0.81		3.19		1.42		1.89	
HNR	2.99		0.41		**3.96**	.24	**2.42**	.16
MDVP								
Fundamental Frequency								
F0	2.01		**3.62**	.28	1.81		1.92	
Fhi	3.04		2.63		**2.56**	.17	1.10	
Flo	1.28		0.09		1.63		1.09	
stdF0	**4.28**	.32	**6.43**	.42	**4.09**	.24	1.97	
PFR	2.24		2.65		**5.91**	32	1.28	
Frequency Perturbation								
JITT	**5.95**	.40	**4.92**	.35	**4.34**	.25	1.75	
RAP	**5.85**	.39	**5.06**	.36	**3.95**	.23	2.09	
PPQ	**6.14**	.41	**4.63**	.34	**4.57**	.26	1.15	
sPPQ	**5.11**	.39	1.04		**5.23**	.29	0.94	
vF0	**3.67**	.28	2.79		**5.06**	.28	1.45	
Amplitude Perturbation								
Shim	1.97		1.96		**2.78**	.18	1.33	
APQ	1.88		0.41		**2.80**	.18	1.41	
sAPQ	1.85		0.53		2.13		1.37	
VAM	1.21		1.31		1.10		0.97	
Voice Breaks								
DVB	*		0.19		*		0.27	
Subharmonic Components								
DSH	*		3.39		*		1.17	
Voice Irregularities								
DUV	1.04		0.47		**3.94**	24	0.43	
Noise								
NHR	**3.53**	.25			**5.25**	.29		
VTI	**3.68**	.29			1.72	.22		
SPI	0.41				1.01			
Tremor								
FTRI	0.15				0.75			
ATRI	*				0.70			
Perceptual ratings								
trained raters								
breathiness	**5.46**	.38	**10.35**	.54	**10.28**	.45	**7.09**	.36
roughness	1.95		**10.33**	.53	**7.89**	.38	**7.04**	.35
tension	**10.67**	.54	**14.64**	.62	**3.85**	.22	**7.26**	.36
naive raters								
breathy -- not breathy	**7.38**	.45	**7.98**	.47	**6.72**	.39	**12.66**	.49
rough -- not rough	1.93		**5.28**	.37	**3.34**	.20	**8.37**	.39
tense -- relaxed	3.09		3.17		**4.48**	.26	**3.70**	.22

6.3.2 Acoustical correlates of perceptual ratings

In order to get more insight in perceptual and production aspects of voice quality, acoustical parameters were correlated to the perceptual parameters *breathiness, roughness* and *tension* by the trained raters, and to *breathy -- not breathy, rough -- not rough* and *tense -- relaxed* by the naive raters, on read-aloud text and on sustained /a/. Results of single correlations are given in table 6.3 in terms of percentage of variance explained (r^2).

Table 6.3. Percentage of variance explained by single acoustical predictors for *breathiness, roughness* and *tension* by trained raters, and *breathy -- not breathy, rough -- not rough* and *tense -- relaxed* by naive raters on read-aloud text and sustained /a/. Blanks indicate that correlations were not significant at the 5% level; an asterisk indicates that this particular acoustical parameter was not available on read-aloud text.

	trained raters						naive raters					
	breathiness		*roughness*		*tension*		*breathy*		*rough*		*tense*	
	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/
Manually												
DVO	*		*	10	*	8	*	16	*	26	*	18
PRAAT												
F0	26	4	24		37	6	19	8	11			7
HNR	5	12	4	12		9	4	18	5	22		8
MDVP												
Fund. Frequency												
F0	24		23		36	4	18	10	11			10
Fhi	22	5	17		31	6	22	17	12	6	4	13
Flo				5	29							4
stdF0	21	12	22		21	9	17	31	13	27		18
PFR	14	8	14	6	21	7	16	23	11	27	4	11
Freq. Perturbation												
JITT	31	12	19	5	19	8	24	27	24	28		10
RAP	35	12	22	4	23	7	28	26	26	27		10
PPQ	27	12	17	5	16	8	20	27	21	28		9
sPPQ		12		7		7		24		25		12
vF0	9	12	10	7	10	9	7	30	7	32		17
Ampl. Perturbation												
Shim		12		8		7		20		28		12
APQ		10		10	4	6		15		27		14
sAPQ		6		8		8		6		25		14
VAM						4				14		7
Voice Breaks												
DVB												
Subh. Components												
DSH	18		18		19	11	17		14	5		
Voice Irregularities												
DUV		18		12				14		18		7
Noise												
NHR	*	8	*	8	*	5	*	13	*	22	*	
VTI	*		*		*		*	4	*	4	*	
SPI	*		*		*		*		*		*	
Tremor												
FTRI	*		*		*		*		*		*	
ATRI	*		*		*		*		*		*	9

Inspection of the results shows that single correlations were low (variance explained < 20%) or moderate (between 20% and 50%). Moderate predictors were fundamental frequency and frequency perturbation parameters. The best predictor was F0 measured by means of PRAAT (explaining 37% of the variance of *tension* as judged by the trained raters on read-aloud text). For the trained raters more of the variance was explained on read-aloud text than on sustained /a/ while for the naive raters results were equal or just the other way around. Furthermore, results on amplitude perturbation parameters appeared to be significant on sustained /a/ but not on read-aloud text.

In order to investigate if a set of multiple acoustical parameters can predict perceptual scores better, stepwise multiple regression models were composed for the acoustical analyses from PRAAT and from MDVP. In order to avoid multicollinearity (which can inflate the computational accuracy seriously), we decided to enter only those acoustical parameters in the regression analysis models that explained more than 20% of the variance of at least one perceptual parameter (see table 6.2) and that showed intercorrelations less than .70 (Pearson correlation; an overview of all correlations between acoustical parameters is given in Appendix 6.1). This happened to be the case for both parameters from PRAAT: F0 and HNR. Results are given in table 6.4 (middle). From MDVP, the parameters F0, PPQ, APQ, and NHR were eligible; however, in order to compare read-aloud text and sustained /a/ material, NHR (noise parameter) was not entered, because it was not available on read-aloud text. Results are given in table 6.4 (top). Furthermore, analyses were performed with F0 and NHR only from MDVP as acoustical predictors in order to be comparable with PRAAT (table 6.4, bottom). Finally, in order to investigate the influence of vowel onset aspects, multiple regression analyses were carried out with DVO (Duration of Voice Onset) added to the acoustical parameters mentioned.

Results in table 6.4 show that prediction was best for the trained raters on read-aloud text. Again a clear difference was found between read-aloud text and sustained /a/ for the trained raters, both for results from MDVP (top) and from PRAAT (middle). On read-aloud text, F0 and PPQ together predicted all rating scales, except *tense -- relaxed* judged by the trained raters. The same holds for F0 and HNR measured by PRAAT. On sustained /a/, APQ contributed to the prediction of *roughness* and *tension*. Measured by means of PRAAT, *roughness* was predicted by HNR alone. Adding DVO (Duration Voice Onset) to MDVP and PRAAT parameters was not effective for correlations with *breathiness*, but was for *roughness* and *tension*. Especially the judgement of *roughness* and *tension* by naive raters was predicted much better.

Table 6.4. Results of stepwise multiple regression analyses in terms of percentage of variance explained by acoustical predictors from MDVP at the top (F0 (Fundamental Frequency), PPQ (Pitch Period Perturbation Quotient) and APQ (Amplitude Period Perturbation Quotient), and F0, PPQ, APQ combined with DVO (Duration Voice Onset) measured manually), from PRAAT at the middle (FO and HNR (Harmonics to Noise Ratio), and F0 and HNR combined with DVO), and from MDVP at the bottom (F0 and NHR (Noise to Harmonics Ratio), and F0 and NHR combined with DVO) for *breathiness, roughness* and *tension* by trained raters, and *breathy -- not breathy, rough -- not rough* and *tense -- relaxed* by naive raters on read-aloud text and sustained /a/. Plus signs indicate the acoustical parameters that actually contributed to the stepwise regression analyses; blanks indicate that correlations were not significant at the 5% level (two-tailed). An asterisk means that no multiple regression analyses was carried out because one of the acoustical parameters was not available on read-aloud text.

MDVP	trained raters						naive raters					
	breathiness		*roughness*		*tension*		*breathy*		*rough*		*tense*	
	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/
F0	+		+	+	+	+	+	+	+			+
PPQ	+	+	+		+		+	+	+	+		
APQ				+		+				+		+
% var. expl.	44	12	35	13	46	13	34	32	28	32		12
F0				+		+		+				+
PPQ		+						+		+		
APQ				+		+				+		+
DVO				+		+		+		+		+
% var. expl.	*	12	*	17	*	17	*	38	*	45	*	33

PRAAT	trained raters						naive raters					
	breathiness		*roughness*		*tension*		*breathy-not b*		*rough- not r*		*tense-relax*	
	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/
F0	+	+	+		+	+	+	+	+			+
HNR	+	+	+	+	+	+	+	+	+	+		+
% var. expl.	41	17	38	12	47	16	31	27	22	22		17
F0		+		+		+		+		+		+
HNR		+		+		+		+		+		+
DVO				+		+		+		+		+
% var.expl.	*	17	*	18	*	21	*	36	*	40	*	30

	trained raters						naive raters					
	breathiness		*roughness*		*tension*		*breathy-not b*		*rough- not r*		*tense-relax*	
MDVP	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/	text	/a/
F0		+		+		+		+				+
NHR		+		+		+		+		+		+
% var. expl.	*	11	*	11	*	9	*	22	*	23	*	19
F0		+		+		+		+		+		+
NHR		+		+		+		+		+		+
DVO				+		+		+		+		+
% var. expl.	*	11	*	18	*	15	*	31	*	39	*	31

The best results on acoustical correlates of perceptual ratings showed that single correlations (table 6.2) can be improved by multiple predictions (table 6.3) on read-aloud text from 37% (F0 measured by means of PRAAT as predictor of *tension* as judged by trained raters) to 47% (F0 with HNR measured by means of PRAAT as predictors of *tension* as judged by trained raters). Results from PRAAT appeared to predict perceptual ratings better than comparable results from MDVP, measured by single correlations as well as by multiple correlations. The best results for trained raters were found on read-aloud text, for the naive raters results on both types of speech material were similar. The duration of the voice onset seems to be an important aspect, especially for naive raters. However, the best acoustical multiple prediction of perceptual aspects of voice quality was no more than 47%, which is considered to be moderate.

6.4 Discussion

6.4.1 Differentiation among speaker groups

The first question of this chapter was if the same trend that was found by perceptual analyses of voice quality of speakers with early glottic carcinoma recorded before and after radiotherapy, was found by means of acoustical analyses as well. On read-aloud text, no similar trends were found on the separate speaker groups; on the longitudinal speaker group, acoustical trends concerning fundamental frequency and frequency perturbation (jitter) aspects were found to be similar but less strong than for the perceptual trends. Although on sustained /a/ similar trends were found for acoustical and perceptual voice aspects, the perceptual trends were stronger and indicated more significant differences between speaker groups than the present acoustical trends. Results were better for the longitudinal groups than for the separate speaker groups. In the case of a longitudinal speaker group research design, results from acoustical frequency perturbation analyses come close to the results from naive raters on read-aloud text as well as on sustained /a/. In the case of a separate speaker group design, the best choice for voice quality analysis should be the use of perceptual analysis by trained raters on read-aloud text.

6.4.2 Acoustical correlates of perceptual ratings

The results on acoustical correlates of perceptual ratings showed that single correlations can be improved by multiple predictions; prediction still remains moderate (47%, at best). The best prediction of judgements by trained raters was found on read-aloud text; for the naive raters results on both types of speech material were similar. Results from PRAAT appeared to predict perceptual ratings better than comparable results from MDVP, measured by single correlations as well as by multiple correlations. The duration of the voice onset seems to be an important aspect, especially for naive raters. It should be taken into account that perceptual ratings on sustained /a/ were performed on the whole vowel (onset plus 2 seconds), as were the analyses of HNR measured by PRAAT, whereas MDVP analyses were carried out on a (hand-selected) stable steady part of the vowel (without the onset).

Various results ranging from low to high are found in studies concerning acoustical correlates of perceptual ratings on sustained vowels. Overviews of these studies are given in Martin et al. (1995) and De Krom (1995). Although comparison of results is difficult, because there is neither consensus concerning speech material (running speech, sustained vowel (often /a/), with or without onset), nor perceptual ratings (trained or naive raters, terminology of the rating scales), nor acoustical analyses (different algorithms), it seems that results concerning noise parameters (signal-to-noise ratio) are better predictors of several pathological voice qualities than perturbation parameters.

Moderate correlations (variance explained between 20% and 50%) were found of perturbation analyses for example by Takahaski and Koike (1975) between breathiness and shimmer (31%), between roughness and jitter (30%), and between roughness and shimmer (52%), by Deal & Emanuel (1978) between roughness and jitter (48%), by Yumoto et al. (1984) between hoarseness and jitter (50%), by Wolfe and Steinfatt (1987) between roughness and jitter (26%), by Wolfe et al. (1995) between shimmer and dysphonic severity (29%) and by Martin et al. (1995) between severity of roughness and shimmer (43%).

Concerning noise parameters, moderate to high correlations were found by Kojima et al. (1980) between hoarseness and HNR (74%), by Yumoto et al. (1984) between HNR and severity of hoarseness (64%), by De Krom (1995) between a combination of several spectral noise parameters and breathiness (68% for vowel onset and 69% for post-onset) and roughness (61% for vowel onset and 35% for post-onset) and by Martin et al. (1995) between HNR and severity of rough voices (73%) and between a combination of jitter, shimmer and HNR, and severity of breathy voices (74%).

The reason that in the present study only moderate results were found, may be due to the fact that acoustical analyses were carried out on the whole frequency span of the speech material. In the above mentioned studies, analyses were carried out in different frequency ranges. As Hammarberg & Gauffin (1995) summarise, breathiness seems to correlate with high frequent noise (4-5 kHz), while roughness seems to correlate with noise in the formant region.

Still, all these results on sustained /a/ speech material indicate that perceptual ratings cannot be predicted very well by acoustical analyses. One explanation may be the influence of fundamental frequency and intensity on perturbation and noise analyses. Among others, Gelfer (1995) found that both aspects are of interest: the highest perturbation values were found in low frequency - low intensity condition. It is no surprise then, that some authors recommend that for stable voices at least six tokens are needed and for instable voices at least 15 tokens to obtain representative averages of perturbation measures (Scherer et al., 1995). Another explanation may be that most of the research is done on static parts of sustained vowels, while one can imagine that the more dynamic parts of the speech signal are of importance on perceived voice quality.

In contrast to the multitude of studies on sustained vowels, there are hardly any data available about prediction of perceptual aspects by acoustical parameters on running speech of pathological speakers. Klingholz (1990) used SNR analyses on sustained vowels as well as running speech and found a moderate correlation (r=.63) between two types of speech material; furthermore he found that running speech is better correlated to pathological conditions than sustained vowels. Hillenbrand & Houde (1996) found for read sentences that signal periodicity and spectral tilt were accurate predictors of breathy voices (explaining 70-80% of the variance of breathiness). However, they question if this accuracy can be achieved for voices that are rough or hoarse. It is clear that more research is needed on acoustical analyses on running speech.

6.5 Conclusion

Perceptual analyses by trained raters on read-aloud text appeared to differentiate speaker groups best. Acoustical analyses on running speech are not yet as sophisticated as the judgements by trained raters. More research is needed on acoustical analyses on running speech; especially the onset of vowels (duration and voice quality aspects) and transitions between voiceless and voiced parts seem to be of interest. However, in the case of a longitudinal study on voice quality of a patient during therapy, acoustical analyses are objective, and easy and quick to perform. Results come close to the subjective, time-consuming perceptual analyses by naive raters.

Appendix 6.1 Correlations between acoustical parameters

In this appendix, significant correlations (Pearson) are given between the acoustical parameters from PRAAT, MDVP, and the manually derived DVO (Duration Voice Onset) for sustained /a/ (top) and read-aloud text(bottom). For abbreviations, see table 6.1. Negative correlations are printed italic.

	F0	hnr	F0	fhi	flo	stdF0	pfr	jitt	rap	ppq	sppq	vF0	shim	apq	sapq	vam	nhr	vti	spi	dvb	dsh	duv	ftri	atri
hnr		1.																						
F0	.93		1																					
Fhi	.88		.95	1																				
Flo	.90	.27	.95	.83	1																			
stdF0	.36	*.50*	.46	.68	.20	1																		
PFR		*.61*		.44		.85	1																	
jitt		*.55*	.25	.48		.90	.80	1																
RAP		*.53*	.26	.48		.89	.78		1															
PPQ		*.55*	.24	.48		.91	.81	.99	.99	1														
sPPQ		*.52*	.20	.42		.87	.72	.89	.87	.91	1													
vF0		*.60*	.20	.46		.95	.90	.91	.90	.92	.89	1												
shimm		*.80*		.28		.66	.66	.70	.69	.71	.67	.71	1											
APQ		*.76*		.20		.60	.63	.64	.62	.65	.63	.67	.95	1										
sAPQ		*.59*				.44	.44	.44	.43	.44	.50	.52	.72	.80	1									
VAm		*.37*		.21		.36	.34	.27	.27	.27	.29	.37	.48	.57	.67	1								
NHR		*.80*		.23		.62	.67	.66	.65	.67	.62	.68	.80	.74	.58	.29	1							
VTI		*.69*				.31	.38	.36	.35	.35	.28	.36	.54	.45	.30		.67	1						
SPI		.35															*.30*	*.47*	1					
DVB																				1				
DSH		*.28*	.32	.42	.20	.48	.35	.42	.42	.42	.38	.39	.37	.27	.12	.02	.34	.33			1			
DUV		*.65*		.21		.59	.60	.66	.64	.67	.64	.65	.75	.73	.64	.48	.77	.34				1		
FTRI						.39	.36	.21		.23	.34	.43	.23	.25	.29								1	
ATRI				.24		.28	.28					.27	.35	.47	.45	.76							.33	1
DVO		.26				.26	.29	.32	.32	.30	.22	.29	.27	.27	.17		.26	.29						

	F0	hnr	F0	fhi	flo	stdF0	pfr	jitt	rap	ppq	sppq	vF0	shim	apq	sapq	vam	dvb	dsh	duv
hnr	.29	1.																	
F0	.97	.32	1																
fhi	.73		.75	1															
flo	.67	.41	.65	.41	1														
stdF0	.74		.80	.75	.27	1													
pfr	.33		.37	.78	.22	.61	1												
jitt	.20	*.44*	.20	.32		.48	.45	1											
rap	.25	*.41*	.26	.35		.49	.45	.98	1										
ppq		*.48*		.31	*.22*	.46	.46	.99	.96	1									
sppq		*.28*			*.27*	.56	.40	.54	.48	.58	1								
vF0	.32		.40	.52		.86	.63	.55	.52	.57	.83	1							
shim	*.39*	*.64*	*.45*	*.21*	*.50*			.51	.47	.55	.28		1						
apq	*.57*	*.56*	*.63*	*.37*	*.56*	*.37*		.25		.31	.20		.90	1					
sapq		*.18*			*.28*								.47	.56	1				
vam													.36	.38	.76	1			
dvb	*.28*	*.47*	*.31*		*.25*			.28	.26	.33	.24		.51	.43			1		
dsh	.45		.46	.46		.58	.49	.54	.58	.50	.23	.46						1	
duv	.37	*.60*	*.41*	*.23*	*.42*			.42	.38	.47	.32		.72	.64	.23		.91		1

7

VOCAL FUNCTION VERSUS VOICE QUALITY*

Abstract

Vocal function of patients with early glottic carcinoma, before and after radiotherapy, was investigated by means of video-laryngo-stroboscopy, phonetography, maximum phonation time, and phonation quotient. Results were compared with voice quality measured by means of perceptual and acoustical analyses. Correlations between parameters of vocal function and voice quality were moderate to high. Although voice quality was mainly affected by the age of the speaker, also stripping rather than biopsying the vocal fold for initial diagnosis had some effect. Analyses of vocal function by means of stroboscopy revealed that in addition to increasing age and stripping the vocal fold, continuing smoking after treatment deteriorated vocal function following radiotherapy.

---

*This chapter is a substantially revised and extended version of Verdonck-de Leeuw, Koopmans-van Beinum, Hilgers & Keus (1997).

7.1 Introduction

Although there are no standards in investigating voice characteristics, there is some consensus nowadays that voice is a multidimensional phenomenon and should be investigated by means of perceptual, acoustical, and physiological analyses. Even though clearly related, a distinction has to be made between voice quality and vocal function (see Chapter 1). Voice quality reflects the outcome of the sound signal and is therefore analysed perceptually and/or acoustically. Vocal function describes habitual laryngeal behaviour and capabilities. A battery of tests is needed to investigate the complexity of vocal function. This battery should include measurements of vocal fold motion, aerodynamic aspects, and maximum performance tasks (Hirano & Bless, 1993).

In Chapter 4, 5, and 6, detailed investigations were described concerning voice quality following radiotherapy for early glottic cancer by means of perceptual and acoustical analyses. We found that voices before radiation were the most deviant, and that 55% of the patients after radiation had normal voices, compared to matched control speakers. These results confirmed findings in the literature. Despite the variety of voice parameters used in the literature (for an overview, see table 1.1 in Chapter 1), it can be concluded that some studies report abnormal postradiation voices (Heeneman et al., 1994; Hirano et al., 1994; McGuirt et al., 1994; Lehman et al., 1986; Stoicheff et al., 1983), whereas others report voice improvement to a normal or near-normal level for at least 70% of the patients (Benninger et al., 1994; Hoyt et al., 1992; Harrison et al., 1990; Miller et al., 1990; Karim et al., 1983; Colton et al., 1978; Stoicheff, 1975; Murry, 1974; Werner-Kukuk et al., 1968).

The question arises then, why some patients have normal voice quality after radiation, while others end up with deviant voice quality. The answer cannot be found in radiation-induced complications on normal tissue alone (Hill, 1990; Keane, 1994). The findings of Benninger et al. (1994) revealed that voice changes may be present after radiotherapy in patients (33%) with associated risk factors, like smoking after treatment, stripping or excision rather than biopsying for initial diagnosis, and complications of radiotherapy (like oedema), but that voices are normal in those patients without these risk factors. These findings were based upon the physician's impression of the voice quality (good or raspy/poor).

The present chapter describes the investigation of vocal function compared to voice quality, measured against stage of the tumour, initial surgery for diagnosis, radiation dose, age of the speaker and smoking habit, in order to get more insight in the source of deviant voice quality following radiotherapy for early glottic carcinoma.

7.2 Methods

7.2.1 Speakers

The subject sample is described in detail in Chapter 3. Briefly, there were in total 100 speaker cases (longitudinal group (n=30), separate groups (n=50), and control group (n=20)). Five factors were taken into account in order to investigate voice characteristics following radiotherapy:
- stage of the tumour,
- initial surgery,
- radiation dose schedule,
- age, and
- smoking habit.

Mean data of the longitudinal and separate patient groups, and control group are given in table 3.4 (Chapter 3).

7.2.2 Vocal function

In the present study, four vocal function tests were applied to analyse vocal function:
- evaluation of stroboscopic video-recordings,
- the Voice Range Profile (phonetogram), and
- maximum phonation time and
- phonation quotient.

7.2.2.1 Video-laryngo-stroboscopy

Video-laryngo-stroboscopy (stroboscopy) was chosen to investigate vocal fold aspects and movement directly. Stroboscopy is the application of intermittent flashes of light that are sent out at 1 - 2 Hz faster than the fundamental frequency of the vocal folds. In this way, the flashes of light illuminate the vocal folds at different phases of the glottal cycle, resulting in a pseudomotion of vocal fold movement at slow motion. Stroboscopic video-recordings were performed using a Wolf rigid endoscope (Model 4452) and a Wolf stroboscope (Model 5510). The endoscope was connected to a Wolf camera (Model 5355) and a Sony video-tape-recorder. All stroboscopic examinations were performed by or under supervision of an experienced phoniatrician. Recordings of the vocal folds were made with continuous light during breathing, and with stroboscopic flash light during phonation. Each speaker produced a sustained /i/ or /a/ at habitual, high and low pitch at comfortable loudness. Recordings were made of 88 speakers. No recordings were made of 12 speakers, because they refused to participate. Recordings of 33 speakers were unsuitable for evaluation, because of unfocused images, no clear vision of the vocal folds or no stroboscopic images, leaving 55 recordings for further analyses.

Two rating tapes (images without sound) were composed for evaluation. The first tape consisted of samples of 18 of the 55 speakers, of whom good recordings were available. This tape was used to accustom the raters to the test. The second tape consisted of the actual test samples of the 55 speakers in random speaker order. Of each speaker, a stationary view of the vocal folds during approximately 5 seconds was followed by stroboscopic images of the vocal folds. The best stroboscopic images were chosen, regardless of frequency, loudness or vowel. A rating form was composed based on the rating form suggested by Hirano & Bless (1993). The rating form consisted of scales relating to overall laryngeal anatomy: the presence of glottic oedema, supraglottic oedema or vascular injection (4-point scales, ranging from none to severe), supraglottic involvement (6-point scale, ranging from none to ventricular dysphonia), and regularity of the vocal fold edge, left and right (6-point scales, ranging from smooth/straight to rough/irregular). Furthermore, scales related to vocal fold movement were included: mobility of the vocal fold, left and right (6-point scales ranging from normal to no visible mobility), mucosal wave, left and right (6-point scales, ranging from normal to absent), and non-vibrating portion, left and right (6-point scales, ranging from none to 100%). Overall glottic closure was scored as complete, or incomplete: anterior chink, irregular, bowing, posterior chink, hourglass, or unilateral mass. Finally, a 4-point scale was added to judge overall quality of the video-recordings.

Three raters (an ENT-specialist, a radiotherapist, and a phonetician/speech therapist), blinded for the clinical data, participated in the experiment. The first tape was judged by each rater, independently of each other. In a training session, these individual ratings were discussed, the final rating form was composed and anchor points for each rating scale were set. The actual test samples on the second tape, were judged in two rating sessions of about 1 hour each by the three raters agreeing a consensus.

Overall recording quality of 3 cases was judged as being too bad for further rating; they were left out, leaving 52 speaker cases for further analyses. Since mobility of the vocal folds of all 52 speaker cases was judged as normal, this scale was left out for further analyses. Of the 52 speaker cases, 7 speakers were controls, 31 patients were diagnosed with a unilateral tumour, and 14 with a bilateral tumour. Of the 31 patients with a unilateral tumour, 18 patients had a tumour on the left vocal fold and scored normal (score=0) on the right vocal fold and normal or deviant on the left vocal fold (score>0); for the 11 patients diagnosed with a unilateral tumour on the right vocal fold, scores were normal on the left vocal fold and normal or deviant on the right vocal fold for 10 patients. For one patient with a unilateral tumour on the right vocal fold, scores (on vocal fold edge and mucosal wave) were deviant on both vocal folds, indicating that the opposite vocal fold was influenced as well. For the 14 patients diagnosed with bilateral tumours, scores were normal for 7 patients, scores were deviant on both vocal folds for 3 patients, and scores were deviant on the left vocal fold for 3 patients (diagnosed with a bilateral tumour originating from the left vocal fold) or deviant on the right vocal fold for 1 patient (diagnosed with bilateral tumour originating from the right vocal fold). Based on these observations, the judgements on left and right vocal fold were combined to one judgement (left and/or right) for the scales amplitude, vocal fold edge, mucosal wave, and nonvibrating portion.

Finally, results showed that the majority of the 52 speaker cases scored normal on all scales (table 7.1). Therefore, next to the original scale ratings, a dichotomy was introduced based on normal (score=0) or deviant (score>0) ratings.

Table 7.1. Results in percentages normal (score=0) or deviant (score>0) ratings on stroboscopic measures on all 52 speaker cases: the presence of glottic oedema, supraglottic oedema, or vascular injection (ranging from none (0) to severe (3)), supraglottic involvement (ranging from none (0) to ventricular dysphonia (5)), regularity of the vocal fold edge (ranging from smooth/straight (0) to rough/irregular (5)), amplitude of the vocal fold (ranging from normal (0) to no visible movement (5)), mucosal wave (ranging from normal (0) to absent (5)), non-vibrating portion (ranging from none (0) to 100% (5)), and vocal fold closure (ranging from complete (0) or incomplete (1:anterior chink, 2: irregular, 3: bowing, 4: posterior chink, 5: hourglass, 6: unilateral mass).

	0	1	2	3	4	5	6
	normal	deviant					
glottic oedema	83	13	4				
supraglottic oedema	69	25	4	2			
vascular injection	60	30	10				
supraglottic involv.	86	4	2	4	4		
amplitude	100						
vocal fold edge	65	13	10	2	2	8	
mucosal wave	64	12	6		6	12	
nonvibrating portion	78	6			6	10	
vocal fold closure	81	4		2	2	4	7

7.2.2.2 Voice Range Profile

The Voice Range Profile (VRP), also called phonetogram, gives insight in pitch and intensity range of a speaker's voice. The basic instrumentation consisted of a tone generator and a SPL measuring device. An experienced investigator performed the VRP. The speaker was asked to produce a sustained /a/

as loud and as soft as possible at selected frequencies; the minimum and maximum SPL level was noted at the selected frequencies in the VRP. A problem occurs in acquiring a VRP as described, when a speaker is not able to produce a sustained /a/ at the given pitch. In cases this happened, the speaker was asked to phonate at an alternative pitch as loud and soft as possible while the investigator matched the actual realised pitch.

In the present study, two parameters from the VRP were taken into account. The pitch range (highest minus lowest frequency) in Hz was calculated for each speaker. Furthermore, the intensity at speaking fundamental frequency was calculated (speaking fundamental frequency was analysed for each speaker by means of acoustical analysis of the average fundamental frequency of read-aloud text by means of PRAAT (see also Chapter 5). The intensity range (loudest minus softest level) in dB at this particular frequency was calculated for each speaker.

7.2.2.3 Maximum phonation time/Phonation Quotient

The phonation quotient (PQ) is defined as the vital capacity divided by the maximum phonation time. The vital capacity was measured by means of a spirometer (Pneumomescreen 11/1,84) and is defined as the amount of air that can be forcefully expelled from fully inflated lungs. The speaker was asked to take a deep breath and to exhale into the mouthpiece of the spirometer with pinches on the nose to avoid nasal escape. No measurements of vital capacity were made of 9 patients, because of logistic reasons.

The maximum phonation time (MPT) is defined as the time in seconds phonated at comfortabel pitch and loudness as long as possible after a maximum inhalation; the best attempt out of three was taken as the maximum phonation time to calculate the phonation quotient.

7.2.3 Voice quality

Methods concerning voice quality are described in detail in Chapter 4 (perceptual analysis) and Chapter 6 (acoustical analysis). A brief description is given here.

7.2.3.1 Perceptual analysis

In the present chapter we limited ourselves to the 7-point scales *breathiness* (the amount of air escaping through the glottis), *roughness* (the amount of aperiodic vibrations, resulting in rough and rasping quality), and *tension* (the impression of muscle tension in the vocal folds), as judged by the trained raters, because these scales appeared to differentiate speaker groups best (Chapter 4). Interrater reliability was high: Cronbach's alpha was above .90 both on read-aloud text and sustained /a/ for *breathiness*, *roughness*, and *tension*, except for *tension* on sustained /a/ (.88). Intrarater reliability was determined by calculating the percentage agreement (within 1 scale value) between the first and the second (repeated) ratings. Intrarater reliability was equally high: above 90% for *breathiness* and *tension* and above 85% for *roughness* both on read-aloud text and sustained /a/.

Next to the actual ratings, dichotomies were introduced based on normal (actual score equal or higher than 4) or pathological (actual score lower than 4) *breathiness*, *roughness* and *tension*.

7.2.3.2 Acoustical analysis

Acoustical analyses were performed using MDVP and PRAAT. In the present chapter, mainly parameters from MDVP were used to compare voice quality and vocal function, because MDVP is frequently used in clinical settings. Mean fundamental frequency (F0), standard deviation of F0 (stdF0), the jitter coefficient Relative Average Perturbation (RAP) were taken into account, both on read-aloud text and sustained /a/, because these parameters appeared to differentiate speaker groups best (Chapter 6). Since MDVP does not allow measuring noise aspects on running speech, we used our in-house method to analyse the harmonics-to-noise ratio (HNR), implemented in the speech signal processing program PRAAT.

7.2.4 Statistical analyses

Pearson product-moment correlation coefficients were used for calculation of relations between vocal function and voice quality measures. A univariate analysis of variance was performed on the ratings on the six separate speaker groups to describe voice characteristics of patients before radiotherapy and of patients in various stages after radiotherapy and of control speakers. Post hoc tests (Tukey) were used to describe the differences between the various separate speaker groups. Multivariate variance analysis (general linear model) on patients after radiotherapy of the separate speaker groups was carried out to investigate five aspects that might influence voice quality: tumour stage, initial surgery, radiation dose, age of the speaker, and smoking habit after treatment. Chi-square tests were performed on the introduced dichotomies (normal or deviant rating scores). Because of the small sample size (cell counts less than 5), in some tests Fisher's Exact Test was used to test whether differences were significant.

7.3 Results

7.3.1 Correlations between vocal function and voice quality measures

Pearson correlations were calculated to investigate relations between vocal function (phonation quotient, maximum phonation time, Pitch Range and Intensity Range, and stroboscopic parameters) and voice quality (perceptual and acoustical) measures. Relations between vocal function measures are given in table 7.2. Relations between vocal function and voice quality measures are given in table 7.3.

Table 7.2. Pearson correlations (paired) between vocal function measures: phonation quotient (PQ) (n=91) and maximum phonation time (MPT) (n=100), VRP measures pitch range (PR) and intensity range (IR) (n=100), stroboscopy measures (n=52) glottic oedema, supraglottic oedema, supraglottic involvement, vascular injection, vocal fold edge, mucosal wave, and non-vibrating portion. Results are given only in the case of significant correlations (p<0.01)).

	PQ	MPT	PR	IR	gl.o	sgl.o	sgl.i	v.in	edge	wave
MPT	-.78	1								
Pitch Range			1							
Intensity Range				1						
glottic oedema					1					
supragl. oedema			-.30		.58	1				
supragl. involv.					.61	.60	1			
vascular inject.				.35				1		
vocal fold edge				-.30					1	
mucosal wave									.66	1
nonvib. portion									.72	.91

Table 7.3. Pearson correlations (paired) between vocal function (variable n, see also table 7.3) and voice quality measures (n=100): perceptual measures *breathiness*, *roughness* and *tension*, and acoustical measures fundamental frequency (F0), standard deviation of F0 (stdF0), and relative average perturbation (RAP). Results are given only in the case of significant (p<0.01) correlations. Harmonics-to-noise ratio (HNR) did not correlate significantly with any of the vocal function measures, and is therefore left out of this table.

	breathiness		*roughness*		*tension*		F0		stdF0		RAP	
	txt	/a/	txt	/a/	txt	/a/	txt	/a/	txt	/a/	txt	/a/
MPT	.27										.30	
PQ	.24											
Pitch Range												
Intensity R.	.33	.26		.32		.31						
glottic oed.												
supragl. oed.												
supragl. inv.												
vascular inj.												
voc. fold edge	.57		.50		.52		.43		.40		.39	
mucos. wave	.54		.42		.46		.48		.42		.35	
nonvib. port.	.60		.51		.53		.54		.52		.42	

Results from table 7.2 and 7.3 show that there are no significant correlations between aerodynamic measures (Phonation Quotient and Maximum Phonation Time) and the other vocal function measures. Correlations between Voice Range Profile (VRP) and stroboscopic measures were moderate (<.50), as were correlations between aerodynamic and VRP measures on the one hand and voice quality measures on the other. Based upon strong intercorrelations between stroboscopic measures and upon strong correlations (>.50) between some of the stroboscopic measures and voice quality measures, stroboscopic measures can be divided in the presence of oedema (glottic or supraglottic) or supraglottic involvement on the one hand, and regularity of the vocal fold edge, mucosal wave and nonvibrating portion on the other.

It is striking that voice quality measures on sustained /a/ did not correlate significantly with stroboscopic measures, maximum phonation time, phonation quotient, or F0-range, although these vocal function measures were collected on sustained vowels. Intensity range was the only vocal function measure that correlated significantly with (perceptual) voice quality measures on sustained /a/.

It is obvious that there are moderate to strong relations between vocal function and voice quality measures on read-aloud text. In the next section vocal function of patients before and after radiotherapy will be investigated to assess whether the same trend will be found as was found for voice quality measures in chapter 4 and 6.

7.3.2 Vocal function before and after radiotherapy

To investigate vocal function of patients before and after radiotherapy compared to control speakers, analyses of variance were carried out on the longitudinal speaker group (patients before, 6 months and 2 years after radiotherapy) and on the separate speaker groups (patients before, 6 months, 2 years, 3-7 years and 7-10 years after radiotherapy, and control speakers). Results of aerodynamic and VRP measures are given in table 7.4; results on voice quality measures are given as well.

Results show that Intensity Range differentiated longitudinal as well as separate speaker groups, whereas Pitch Range differentiated the separate groups only. Posthoc tests on the separate speaker groups on Intensity Range showed the same trend as was found for perceptual and acoustical measures of voice quality. However, the trend on Intensity Range appeared to be less strong: only the difference between patients before radiotherapy and controls was statistically significant (figure 7.1).

Table 7.4. Results of analyses of variance (F-statistic, $p < 0.05$ is printed bold, whereas η^2 is only printed for the significant F-statistic) for the longitudinal speaker group (patients before, 6 months and 2 years after radiation) at the left hand, and for the 6 separate speaker groups (patients before, 6 months, 2 years, 3-7 years, 7-10 years after radiation and controls) at the right hand, of vocal function measures (maximum phonation time (MPT), phonation quotient (PQ), Pitch Range, and Intensity Range) and of voice quality measures (*breathiness, roughness, tension*, mean fundamental frequency (F0), standard deviation of F0 (stdF0), relative average perturbation (RAP), and harmonics-to-noise ratio (HNR)), on sustained /a/ and on read-aloud text.

Voice Parameters		longitudinal speaker groups				separate speaker groups			
		F	η^2	F	η^2	F	η^2	F	η^2
Vocal function	n								
Pitch Range	100	0.46				**6.30**	.33		
Intensity Range	100	**6.48**	.42			**3.35**	.21		
MPT	100	0.25				1.28			
PQ	91	0.28				0.16			
Voice quality		sustained /a/		read-aloud text		sustained /a/		read-aloud text	
breathiness	100	**5.46**	.38	**10.35**	.54	**10.28**	.45	**7.09**	.36
roughness	100	1.95		**10.33**	.53	**7.89**	.38	**7.04**	.35
tension	100	**10.67**	.54	**14.64**	.62	**3.85**	.22	**7.26**	.36
F0	100	2.01		**3.62**	.28	1.81		1.92	
stdF0	100	**4.28**	.32	**6.43**	.42	**4.09**	.24	1.97	
RAP	100	**5.85**	.39	**5.06**	.36	**3.95**	.23	2.09	
HNR	100	2.99		0.41		**3.96**	.24	**2.42**	.16

Results of posthoc tests of the separate speaker groups on Pitch Range appeared to be significant as well. However, a different trend curve was found: patients 6 months after radiotherapy differed significantly from control speakers (figure 7.2). This acute effect of radiation 6 months after treatment on Pitch Range confirmed results on average pitch measures in chapter 5.

Results of stroboscopic measures showed little or no variance in one or more speaker groups; therefore, no analyses of variance were carried out on these measures. Instead, percentages normal (score 0) or abnormal (score > 0) were calculated per speaker group. Results are given in table 7.5.and show that patients before radiotherapy have normal laryngeal anatomy in terms of absence of glottic oedema, vascular injection and supraglottic involvement. Supraglottic oedema was seen in one patient. It is striking that for the control speakers supraglottic oedema (two speakers), vascular injection and supraglottic involvement (one speaker) was seen. However, deviancy in these cases was slight (score 1). Over all patients after radiotherapy, laryngeal anatomy was deviant more frequently and more severely (scores ranged, up to maximum scores) compared to patients before radiation and control speakers.

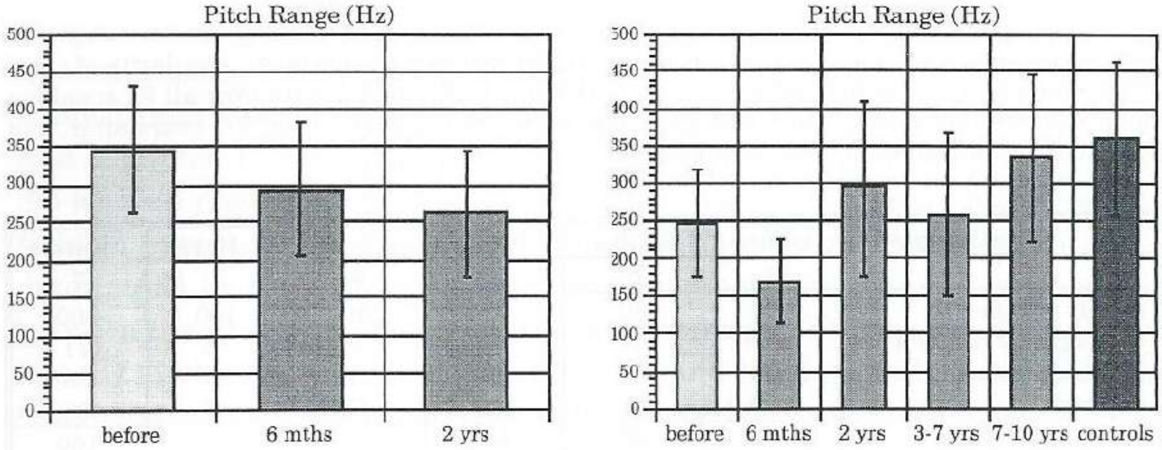

Figure 7.1. Mean and standard deviation of the Pitch Range (PR) in Hz from the Voice Range Profile for the longitudinal speaker group (before radiotherapy, and 6 months and 2 years after radiotherapy) at the left and for the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiotherapy, and control speakers) at the right.

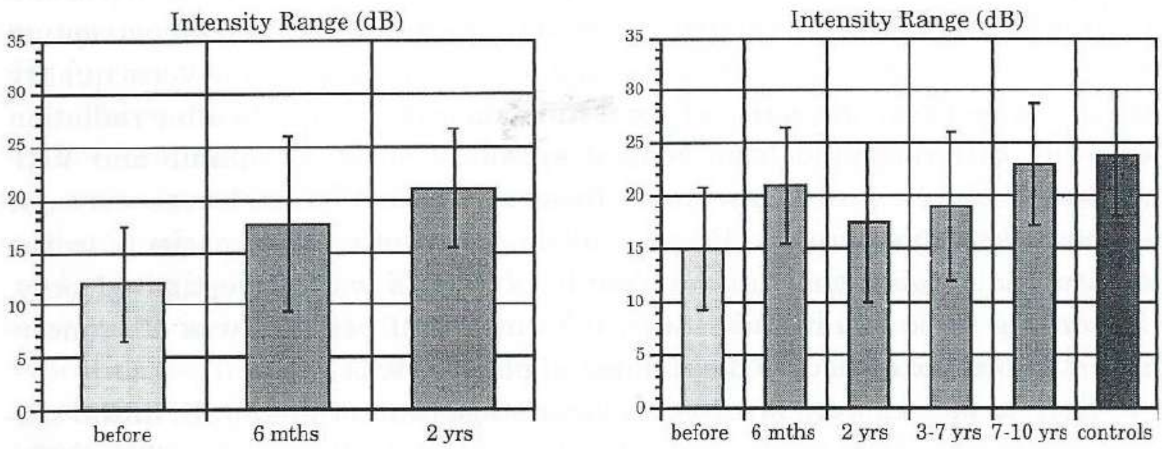

Figure 7.2. Mean and standard deviation of the Intensity Range in dB from the Voice Range Profile for the longitudinal speaker group (before radiotherapy, and 6 months and 2 years after radiotherapy) at the left and for the separate speaker groups (before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiotherapy, and control speakers) at the right.

It is obvious that vocal fold edge, mucosal wave, and nonvibrating portion of the vocal folds was deviant for most of the patients before radiation, and thereby vocal fold closure. Control speakers scored all normal. It seems that vocal fold edge, mucosal wave, nonvibrating portion and vocal fold closure improved for the majority of patients after radiation.

Table 7.5. Results in percentages normal ratings (score=0) on stroboscopy measures glottic oedema, supraglottic oedema, vascular injection, supraglottic involvement, regularity of vocal edge, mucosal wave, nonvibrating portion of the vocal fold, and closure over all 52 speakers divided in speaker group before radiation, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation, and controls.

speaker group	before	6 mths	2 yrs	3-7 yrs	7-10 yrs	controls
n	9	10	11	8	7	7
glottic oedema	100	80	73	50	100	100
supraglottic oedema	89	60	55	63	86	71
vascular injection	100	80	45	13	29	86
supraglottic involv.	100	70	100	75	86	86
vocal fold edge	11	80	64	50	100	100
mucosal wave	11	60	91	75	57	100
nonvibrating portion	22	100	100	88	57	100
vocal fold closure	56	80	100	63	86	100

7.3.3 Voice characteristics following radiotherapy

To investigate voice characteristics of patients after treatment, multivariate analysis of variance were carried out on patients after radiation on perceptual (*breathiness*, *roughness*, and *tension*) and acoustical measures of voice quality and on stroboscopic measures of vocal function. Since patients after radiation were not differentiated from control speakers on aerodynamic and VRP measures (see the previous section), these vocal function measures were not further taken into account. Because of the small number of patients in the longitudinal group, the separate patients groups were investigated only. Concerning voice quality measures, the number of patients was 40, concerning stroboscopic measures the number of patients was 23.

Several factors were involved: tumour stage (uni- or bilateral), initial surgery (stripping or biopsying), radiation dose schedule (66Gy/33fr., 60Gy/30fr., 60Gy/25fr.), age (younger than 65 years, between 65 and 70, between 70 and 75, or older than 75 years), and smoking habit after treatment (yes or no). Because of the limited number of speakers, major effects were investigated only; interaction effects were left out. Results are given in table 7.6.

On 40 patients after radiotherapy, increasing age increased *breathiness* and RAP (jitter). Also, increasing age seemed to increase average fundamental frequency (this effect of age was also found for the control speakers on *breathiness* ($F=2.94$, $p=0.08$, $df1=2$, $df2=16$)) and average fundamental frequency ($F=2.98$, $p=0.08$, $df1=2$, $df2=16$)). Radiation dose schedule (60Gy/25fr. instead of 60GY/30fr. or 66Gy/33fr.) seemed to increase RAP.

Table 7.6. Results of multivariate analysis of variance (F-value with correspondgin p-value) of voice quality parameters (perceptual ratings of breathiness and roughness and acoustical measures of fundamental frequency (F0), harmonics-to-noise ratio (HNR), and frequency perturbation (RAP)) and stroboscopic vocal function parameters (glottic oedema, supraglottic oedema, vascular injection, supraglottic involvement, regularity of vocal edge, mucosal wave, and nonvibrating portion of the vocal fold) over 23 patients after radiation of the separate speaker groups with the independent variables stage of the tumour, initial surgery, radiation dose schedule, smoking, and age. Results are given only in the case of a (nearly) significant effect of the independent variable (second p-value).

	F	p	effect	p
Voice quality (n=40)				
breathiness	1.89	.12	age	.01
roughness				
F0	1.26	.30	age	.07
HNR				
RAP	2.29	.08	age	.00
			dose schedule	.09
Vocal function (n=23)				
glottic oedema				
supraglottic oedema				
vascular injection				
supraglottic involv.				
vocal fold edge	2.68	.06	dose schedule	.02
			smoking	.08
mucosal wave	2.06	.12	initial surgery	.04
nonvibrating portion	2.43	.08	initial surgery	.03
			age	.08

Furthermore, on 23 patients after radiotherapy, initial surgery (stripping the vocal fold) affected negatively the mucosal wave and the vibrating portion of the vocal fold. Radiation dose schedules 66Gy in 33 fractions or 60Gy in 30 fractions instead of 60Gy in 25 fractions decreased vocal fold regularity. Continuing smoking after treatment seemed to increase vocal fold edge irregularity. Increasing age of the speaker seemed to increase the nonvibrating portion and average fundamental frequency.

These results were confirmed by results of chi-square tests on stroboscopic dichotomies: no significant effect was found for the presence of oedema (glottic or supraglottic), vascular injection or supraglottic involvement, nor for vocal fold closure; a significant effect ($p < 0.05$) was found for vocal fold edge (radiation dose schedule:66Gy/33fr. or 60Gy/30fr. instead of 60Gy/25fr.)), for mucosal wave (initial surgery: stripping) and nonvibrating portion (initial surgery: stripping). Finally, chi-square tests on perceptual dichotomies revealed that pathological *roughness* is associated with stripping the vocal folds ($p=0.08$), instead of taking a small biopsy for initial diagnosis.

7.4 Discussion

As was stated in the introduction of this chapter, a battery of tests is needed to investigate vocal function. In our study, we used four tests: stroboscopic recordings to examine the vocal folds directly, the phonetogram to investigate laryngeal capabilities concerning pitch and intensity range, and maximum phonation time and phonation quotient to investigate aerodynamic aspects. Although no significant correlations between aerodynamic tests and other vocal function tests were found, maximum phonation time and phonation quotient correlated moderately with *breathiness*. However, no speaker group differences were found by means of the maximum phonation time or phonation quotient. Therefore, it is concluded that these particular tests on aerodynamic aspects are not sensitive enough in investigating voice characteristics following radiotherapy.

Concerning the Voice Range Profile, some moderate correlations were found between Pitch Range and Intensity Range and stroboscopic measures. Intensity Range correlated moderately with *breathiness*, *roughness* and *tension*, and differentiated between speakers before radiation and control speakers. Pitch range correlated moderately with supraglottic oedema; no correlations were found with voice quality measures. Pitch Range differentiated between patients 6 months after radiotherapy and control speakers only. Therefore, we decided not to investigate Voice Range Profiles in more detail, as suggested by other studies (Airaner & Klingholz, 1993; Sulter et al. (1994). Some general comments can be made about this decision. Phonetography contains two main factors that might introduce unreliable Voice Range Profiles: the effort by the speaker and the registration procedure. The effort of the speaker is an important restriction in obtaining reliable and valid results. We experienced that part of the speakers in our study felt uncomfortable in producing extreme high pitch and intensity levels, even though they were encouraged to do so; furthermore, part of the speakers had trouble producing the target pitch. Besides and because of these problems, it was not unusual that time-investment was about half an hour to obtain a satisfactory VRP, which is very long in clinical practice. It is often argued that automatic VRP registration can decrease this time-investment and the occurrence of pitch errors compared to a "hand-made" VRP (Pabon & Plomp, 1988). Titze et al. (1995) however, concluded that there is no obvious preference for the use of clinician-assisted or fully automated procedures for normal voices; nevertheless, intervention by the clinician can be needed in case of underestimated high pitch ranges (due to unstable efforts) and overestimated lower pitch ranges (due to subharmonics) of the automatic VRP procedure. For pathological voices, results in chapters 5 and 6 also revealed that pitch errors can occur by means of automatic pitch extraction as well as perceptual ratings, especially in the case of severe breathy or rough voices. All in all, it was concluded that the present results of phonetography were not valuable enough in investigating vocal function following radiotherapy.

The main disadvantage of video-laryngo-stroboscopy is that examination is uncomfortable for the speaker. In our study, twelve percent of the speakers refused to participate because of this reason. Furthermore, hypersensitivity in 33 percent of the speakers lead to unsuitable recordings, which could not be evaluated. However, for the remaining half of the speakers, stroboscopic measures appeared to be highly valuable in investigating vocal function. Correlations with voice quality measures were high. Moreover, where voice quality measures were associated mainly with age of the speaker and stripping rather than biopsying the vocal folds, analyses of vocal function by means of stroboscopic measures revealed that in addition to increasing age and stripping the vocal fold, continuing smoking after treatment decreased vocal function following radiotherapy. These findings confirm the findings of Benninger et al. (1994), who found that voice changes may be present after radiation in patients with associated risk factors, such as smoking after treatment, stripping rather than biopsying the vocal folds; however, their conclusion that complications of radiotherapy, like oedema, might increase voice changes after radiation, was not confirmed by our study. The fact that vocal function was better for patients treated with 60Gy in 25 fractions instead of 66Gy in 33 fractions or 60Gy in 30 fractions was striking. In our view, an explanation might be that this group consists of long-term survivors, since most of the patients in this groups were treated 7-10 years ago; all patients with local recurrence or other complications were therefore not included. Furthermore, 92% of the patients in this particular group stopped smoking after treatment, against 77% and 50% of the patients treated with the other radiation schedules. It is clear that future prospective studies should give more insight into vocal function and voice quality following radiotherapy on the long term.

7.5 Conclusion

Voice Range Profile measures (Intensity Range and Pitch Range), maximum phonation time and phonation quotient seem not relevant in investigating vocal function following radiotherapy. Video-laryngo-stroboscopy revealed that next to increasing age and stripping the vocal fold (which also decreased voice quality), continuing smoking after treatment and two of the three radiation dose schedules decreased vocal function of patients after radiotherapy.

8

SELF-RATINGS OF VOCAL PERFORMANCE RELATED TO VOICE CHARACTERISTICS

Abstract

Self-ratings of vocal performance related to voice characteristics such as perceptual and acoustical measures of voice quality and stroboscopic measures of vocal function were investigated of patients diagnosed with early glottic cancer before and after radiotherapy compared to control speakers. The trend was that patients before radiotherapy experienced decreased vocal performance, which improved after treatment but remained worse than vocal performance as reported by control speakers. It appeared that surgery for initial diagnosis (stripping rather than biopsying) and continuing smoking after treatment decreased vocal performance following radiotherapy. High correlations were found between self-ratings of vocal performance and voice characteristics, which results are promising for future research on predictability of vocal performance by objective voice quality and vocal function measures.

8.1 Introduction

Traditionally, the main objective of cancer treatment has been prolongation of disease-free length of life. Nowadays, it is generally accepted that quality of life can be as important as quantity of life in medical treatment. Moreover, both dimensions are integrated into one measure of medical outcome: Quality Adjusted Life Years. Multiplying the number of life years spent in a certain health state by a factor that represents quality of life, provides an outcome measure in which both dimensions are represented. However, before applying such a measure in treatment choice, the concept of quality of life has to become clear. Both the theoretical framework and the methods available to measure quality of life are far from adequate (Kiebert, 1995). Concerning quality of life of patients diagnosed with laryngeal cancer, most research is done on (hemi)laryngectomized patients (Pruyn et al. 1986; Jones et al., 1992; Ackerstaff et al., 1994; Olsen et al., 1995; List et al., 1996). However, studies on quality of life following radiotherapy for early glottic cancer are scarce. Llewellyn-Thomas et al. (1984) developed a self-assessment scale of quality of life in laryngeal cancer. Results on 30 patients during radiotherapy and 29 patients 18 months after radiation showed that most of the scales were reliable for patients during treatment and posttreatment. Furthermore, most of the scales were able to demonstrate differences between the start and the end of radiotherapy. No attempt was made to assess quality of life 18 months after treatment. In a study by Bjordal et al. (1994) on 204 patients treated for head and neck cancer, patients reported a high level of symptoms 7 to 11 years after radiation, like dryness in the mouth and mucus production; patients treated with a hypofractionated radiation schedule reported a better overall quality of life than patients treated with a conventional radiation schedule. However, these results were found over all patients, including various tumour sites (oral cavity, pharynx, larynx, nose/sinus), various tumour stages (T1 to T4), and previous treatment (surgical treatment or not). De Boer et al. (1995) investigated rehabiliation outcomes of long-term survivors and found that patients treated with radiotherapy for early glottic cancer, 2 to 6 years previously, experienced a considerable number of physical complaints, such as sore muscles and fatigue, and complaints specific to head and neck tumours (phlegm, frequent colds), speech problems and problems in swallowing; only 10% reported psychosocial problems. List et al. (1996), on the contrary, found that patients treated with radiotherapy showed little overall dysfunction 6 months after treatment. These few studies reveal that research on the assessment of some aspects of quality of life of patients treated with radiotherapy for early glottic cancer is far from complete.

The first aim of the present chapter was to assess one aspect of quality of life, namely vocal performance in daily life situations, of patients diagnosed with early glottic cancer, 6 months to 10 years after radiotherapy compared to control speakers. The second aim was to investigate if these self-ratings of vocal performance can be predicted from voice characteristics.

8.2 Methods

8.2.1 Speakers

The subject sample is described in detail in Chapter 3. Briefly, there was a longitudinal group of 10 patients of whom voice samples were recorded before, as well as 6 months and 2 years after radiation. Furthermore, data were collected of 5 separate groups of 10 patients each, before, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiation. Finally, recordings were made of 20 control speakers without any known vocal defects. In summary, there were in total 100 speaker cases (longitudinal group, (n=30), separate groups (n=50), and control group (n=20)).

Five factors were taken into account in order to investigate vocal performance following radiotherapy: stage of the tumour, initial surgery, radiation dose schedule, age, and smoking habit. Mean data of the longitudinal and separate patient groups, and control groups are given in table 3.4 (Chapter 3).

8.2.2 Self-ratings of vocal performance

Together with the rating scales on voice quality (see next section), the speakers and their partners received rating scales concerning vocal performance. Thirteen questions were asked (in the form of 7-point scales analogue to the voice quality scales), covering vocal abilities and social situations. An overview is given in table 8.1.

The speakers and their partners received score forms with a written instruction. They were asked to evaluate the voice of the speaker at home by filling out the form independently from each other. Of the speakers, 97% responded; 95% of the forms filled in by the partners were returned.

From written or oral feedback, it appeared that many of the speakers or their partners had trouble rating 3 questions. The question "Can you sing?" was often interpreted in a musical way, (and the majority of the speakers and of their partners scored lower than 4, indicating that they were not impressed by the singing capacities of the speakers). The questions "Have your social contacts changed because of your voice" and "Has your voice changed?" was difficult to answer for the control speakers and their partners and for the patients and their partners after treatment, since they wondered what to refer to. Obviously, these questions were badly formulated and were left out of further investigation. Furthermore, the question "Do you smoke?" was left out, because these data were already known from anamnesis, leaving 9 questions for further investigation.

Table 8.1. Overview of vocal performance (7-point) rating scales. Three scales appeared to be ambiguous and were left out, as was the scale on smoking, because these data were already known from anamnesis, leaving 9 questions (printed regular) for further investigation.

Vocal performance questions	scale extremities	
Does your voice change from day to day ?	not at all	very much
Can you shout ?	not at all	very good
Can you have normal conversation regarding your voice ?	not at all	very good
Can you sing ?	*not at all*	*very good*
Do you cough ?	never	always
Do you smoke ?	*never*	*always*
Have your social contacts changed because of your voice ?	*strongly*	*normal*
Has your voice changed ?	*not at all*	*very much*
Can you use your voice in normal work routine?	not at all	normal
Can you make a telephone call regarding your voice ?	not at all	very good
Do you get tired from speaking for a long time ?	not at all	quickly
Do you avoid a smoky room because of your voice ?	not at all	very much
Do you avoid a large party because of your voice ?	not at all	very much

8.2.3 Voice quality

Methods concerning voice quality are described in detail in Chapter 4 (perceptual analysis) A brief description is given here. Next to the vocal performance scales described above, the speakers and their partners received a voice quality rating form. This form consisted of 8 voice quality rating scales: *unpleasant -- pleasant*, *ugly -- beautiful*, *breathy -- not breathy*, *dull -- clear*, *high -- low*, *shrill -- deep*, *unsteady -- steady*, and *not intelligible -- intelligible*. The speakers and their partners received score forms with a written instruction. They were asked to evaluate the voice of the speaker at home by filling out the form independently from each other. Of the speakers, 97% responded; 95% of the forms filled in by the partners were returned. Results described in Chapter 4 revealed that the *scales breathy -- not breathy, dull -- clear, ugly -- beautiful*, and *unsteady -- steady* differentiated speaker groups best. Therefore, these scales were taken into account in the present chapter.

Furthermore, assessment of voice quality was done by trained and naive raters. In the present chapter we limited ourselves to the 7-point scales *breathiness* (the amount of air escaping through the glottis), *roughness* (the amount of aperiodic vibrations, resulting in rough and rasping quality), and *tension* (the impression of muscle tension in the vocal folds), as judged by the trained raters, because these scales appeared to differentiate speaker groups best (Chapter 4). Interrater reliability was high: Cronbach's alpha was above .90 both on read-aloud text and sustained /a/ for *breathiness*, *roughness*, and *tension*, except for *tension* on sustained /a/ (.88). Intrarater reliability was determined by calculating the percentage agreement (within 1 scale value) between the first and the second (repeated) ratings. Intrarater reliability was equally high: above 90% for *breathiness* and *tension* and above 85% for *roughness* both on read-aloud text and sustained /a/.

8.2.4 Vocal function

Methods of vocal function are described in detail in Chapter 7. In the present chapter, results of stroboscopic video-recordings were used only, because they appeared to differentiate speaker groups best. A detailed description of video-laryngo-stroboscopy is given in Section 7.2.2.1. Briefly, recordings were made of 88 speakers. No recordings were made of 12 speakers, because they refused to participate. Because of several reasons, only 52 recordings were suitable for further investigation.

Three raters participated in the experiment, an ENT-specialist, a radiotherapist and a speech therapist/phonetician. The rating form consisted of scales relating to overall laryngeal anatomy: the presence of glottic oedema, supraglottic oedema or vascular injection, supraglottic involvement, and regularity of the vocal fold edge. Furthermore, scales related to vocal fold movement were included: mobility of the vocal fold, mucosal wave, and non-vibrating portion. Overall glottic closure was scored as complete, or incomplete: anterior chink, irregular, bowing, posterior chink, hourglass, or unilateral mass.

8.2.5 Statistical analyses

Two types of analysis of variance were carried out: univariate analysis of variance with repeated measures on the ratings of the longitudinal speaker groups and univariate analysis of variance on the ratings on the six separate speaker groups. The F-tests on the separate speaker groups were used to describe vocal performance of patients before radiotherapy and of patients in various stages after radiotherapy and of control speakers. Posthoc tests (Tukey) were used to describe the differences between the various separate speaker groups. Multivariate variance analysis (general linear model) on patients after radiotherapy of the separate speaker groups was carried out to investigate five aspects that might influence vocal performance: tumour stage, initial surgery, radiation dose, age of the speaker, and smoking habit after treatment. Pearson product-moment correlation coefficients were used to investigate single relations between vocal performance and voice characteristics.

8.3 Results

8.3.1 Vocal performance before and after radiotherapy

To investigate vocal performance before and after radiotherapy compared to control speakers, analysis of variance were carried out on the longitudinal group and on the separate speaker groups (table 8.2). Results show that the question "Do you cough?" did not differentiate the speaker groups: no significant F-values were found. Results of posthoc tests (Tukey) on the scores of the remaining 8 scales of the separate speaker groups reveal that, both for the speakers themselves and their partners, vocal performance before radiotherapy is significantly deteriorated compared to vocal performance after radiation (abilities to "shout", "have normal conversation" and "making a telephone call" were decreased) and compared to control speakers (ratings on all remaining 8 scales were decreased). Patients after radiation experienced deteriorated vocal performance compared to control speakers according to the ratings on all 8 scales but one (they did not have trouble making phone calls). Differences between patient groups after radiotherapy (ranging from 6 months to 10 years) were significant for the question "Can you shout?" and "Can you have a normal conversation?" (figure 8.1): patients 6 months, 2 years and 3-7 years after radiation had significantly more problems with shouting compared to patients 7-10 years after radiation; patients 2 years after radiation reported significantly more problems in having normal conversation than patients 7-10 years after radiation.

These results on the separate speaker groups were also observed for the longitudinal group; because no control speakers were involved in the analysis of variance, not all tests were found to be significant. However, inspection of the longitudinal speaker groups revealed that the trend was that patients before radiotherapy experienced deteriorated vocal performance, which was improved 6 months after radiotherapy and remained that way 2 years after radiation.

Table 8.2. Results of analyses of variance (F-statistic, p< 0.05 is printed bold, and degree of freedom (df)) for the longitudinal speaker group at the left hand, and for the 6 separate speaker groups at the right hand, of 9 vocal performance questions (abbreviated, see also table 8.1) (an asterisk means that there was no variance in one or more groups).

	longitudinal speaker group				separate speaker groups			
	patients		partners		patients		partners	
Vocal performance questions	F	df	F	df	F	df	F	df
Does your voice change daily?	0.56	2,27	2.09	2,25	**5.52**	5,62	**5.69**	5,58
Can you shout ?	**5.08**	2,27	1.55	2,25	**6.52**	5,62	**4.43**	5,59
Can you have normal conversation?	**3.88**	2,26	**5.91**	2,25	**7.08**	5,62	**4.89**	5,59
Do you cough ?	0.84	2,27	0.25	2,25	0.23	5,62	0.34	5,59
Can you exercise your profession?	2.79	2,23	0.87	2,22	**3.72**	5,57	**4.73**	5,54
Can you make a telephone call?	*		*		**6.42**	5,62	**8.86**	5,58
Do you get tired from speaking?	**8.64**	2,26	**3.29**	2,25	**5.62**	5,62	**4.57**	5,57
Do you avoid a smoky room?	0.49	2,25	0.18	2,24	**2.84**	5,62	**5.06**	5,58
Do you avoid a large party?	**10.85**	2,26	1.12	2,25	**2.70**	5,62	**4.12**	5,58

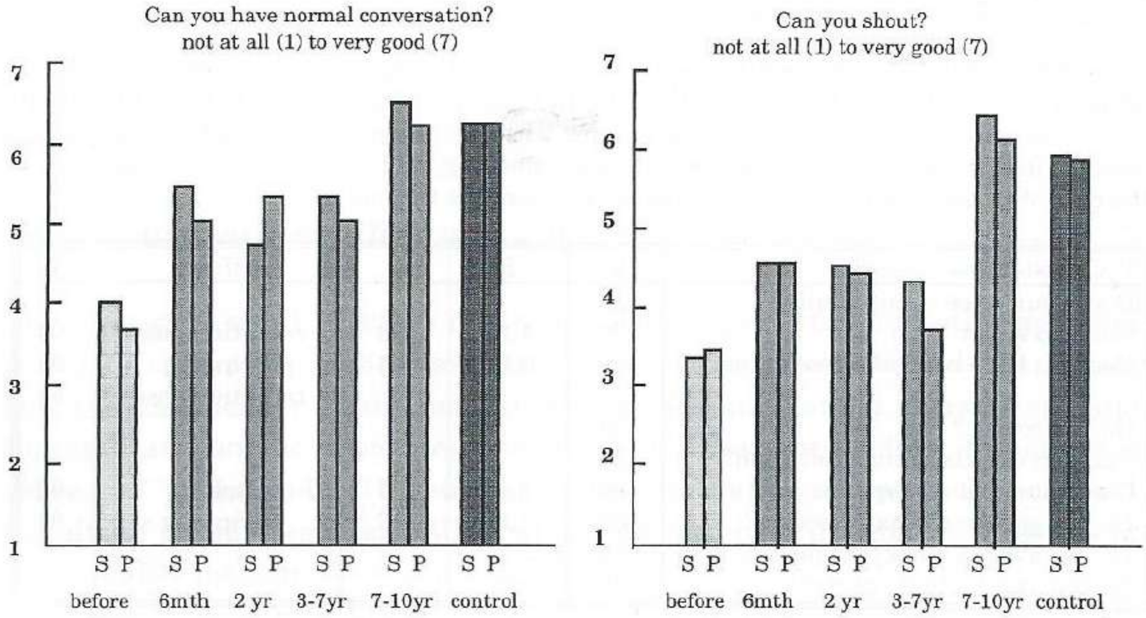

Figure 8.1. Mean scale ratings of speakers themselves (S) in dark grey and their partners (P) in light grey on the scales "Can you shout?" (at the right) and "Can you have normal conversation?" (at the left), from left to right: before radiotherapy, 6 months after, 2 years after, 3-7 years after, and 7-10 years after radiotherapy, and control speakers.

8.3.2 Vocal performance following radiotherapy

To investigate vocal performance of patients after radiotherapy, multivariate analysis of variance were carried out on the vocal performance ratings. Several factors were involved: tumour stage (uni- or bilateral), initial surgery (stripping or biopsying), radiation dose schedule (66Gy/33fr., 60Gy/30fr., 60Gy/25fr.), age (younger than 65 years, between 65 and 70, between 70 and 75, or older than 75 years), and smoking habit after treatment (yes or no). Because of the limited number of speakers, major effects are investigated only; interaction effects are left out of the analysis of variance. Results are given in table 8.3 and show that smoking after treatment decreased the ability of making phone calls, while stripping the vocal fold increased vocal fatigue (getting tired from speaking). Radiation dose schedule (66Gy/33fr. or 60Gy/30fr. instead of 60Gy/25fr.) together with stripping the vocal folds decreased the ability of having normal conversation. Furthermore, radiation dose schedule influenced the ability to shout: patients treated with 66Gy/33fr. or 60Gy/30fr. experienced more difficulties in shouting than patients treated with 60Gy/25fr..

Table 8.3. Results of multivariate analysis of variance (F-value with corresponding p-value) on scores of vocal performance questions by the patients themselves, over all 40 patients after radiation of the separate speaker groups, with the independent variables stage of the tumour, initial surgery, radiation dose schedule, smoking, and age. Results are given only in the case of a significant effect of the independent variable (second p-value).

Vocal performance questions	n	F	p	effect	p
Does your voice change daily?	39				
Can you shout ?	39	1,35	.23	radiation dose	.04
Can you have normal conversation?	39	3,65	.01	stripping	.00
				radiation dose	.01
Do you cough ?	39				
Can you exercise your profession?	35				
Can you make a telephone call?	39	3,43	.01	smoking	.00
Do you get tired from speaking?	39	1,88	.12	stripping	.01
Do you avoid a smoky room?	39				
Do you avoid a large party?	39				

8.3.3 Vocal performance and voice characteristics

Pearson correlations were calculated to investigate relations between vocal performance and voice characteristics. Relations between vocal performance ratings and voice quality ratings by the speakers themselves and their partners are given in table 8.4; relations between vocal performance ratings by the speakers themselves and their partners on the one hand and voice quality ratings by trained raters on read-aloud text on the other are given in table 8.5; relations between vocal performance ratings by the speakers themselves and their partners on the one hand and vocal function measures on the other are given in table 8.6. Results show that high correlations (r >.50) were found between scores on "Can you have normal conversation?" and scores on voice quality rated by the speakers themselves (*breathy -- not breathy*), their partners (*ugly -- beautiful, dull -- clear*), and the trained raters (*breathiness, roughness*) and scores on vocal function measure (regularity of the vocal fold edge).

Table 8.4. Pearson correlations (paired) for the speakers themselves (spk) and for the partners (prt) between vocal performance (VP) ratings and voice quality ratings on *breathy -- not breathy, ugly -- beautiful, dull -- clear,* and *unsteady -- steady*. Results are given only in the case of significant (p<.01) correlations. High correlations (r>.50) are printed bold.

	breathy		*ugly*		*dull*		*unsteady*	
VP questions answered by	spk	prt	spk	prt	spk	prt	spk	prt
Does your voice change daily?		-.37	-.29	-.36	-.39	-.44	**-.54**	**-.56**
Can you shout ?	.42	.32	.42	.42	.42	.47	.37	.58
Can you have normal conversation?	**.50**	.46	.44	**.60**	.32	**.55**	.37	.46
Do you cough ?								
Can you exercise your profession?			.30	.33				.36
Can you make a telephone call?	.31	.29	.33	.44	.32	.45	**.56**	.39
Do you get tired from speaking?	-.45				-.29		-.31	-.43
Do you avoid a smoky room?	-.28	-.32						
Do you avoid a large party?	-.39	-.35					-.29	-.39

Table 8.5. Pearson correlations (paired) between vocal performance (VP) questions rated by the speakers themselves (spk) or their partners (prt) and voice quality ratings by trained raters on read-aloud text on *breathiness*, *roughness* and *tension*. Results are given only in the case of significant (p<.01) correlations. High correlations (r>.50) are printed bold.

	breathiness		*roughness*		*tension*	
VP questions answered by	spk	prt	spk	prt	spk	prt
Does your voice change daily?			-.34		-.29	
Can you shout ?	.45	.29	.44	.33	.41	
Can you have normal conversation?	**.52**	**.50**	**.52**	.47	.49	.43
Do you cough ?						
Can you exercise your profession?	.43	.36	.34	.32	.35	.30
Can you make a telephone call?	.39	**.52**	.33	.48	.34	.47
Do you get tired from speaking?	-.33	-.32	-.40	-.44	-.38	-.32
Do you avoid a smoky room?			-.28	-.37		-.29
Do you avoid a large party?	-.31	-.30	-.29	-.34	-.33	-.31

Table 8.6. Pearson correlations (paired) between vocal performance (VP) questions rated by the speakers themselves (spk) or their partners (prt) and vocal function (stroboscope) measures (n=52) glottic oedema, vocal fold edge, mucosal wave, and non-vibrating portion. Results are given only in the case of significant correlations ($p < 0.01$)). No significant results were found for supraglottic oedema, supraglottic involvement or vascular injection; they were left out of the table. High correlations ($r>.50$) are printed bold.

	glott. oed.		edge		wave		nonvibr.	
VP questions answered by	spk	prt	spk	prt	spk	prt	spk	prt
Does your voice change daily?								
Can you shout ?								
Can you have normal conversation?			**-.51**	-.39				
Do you cough ?								
Can you exercise your profession?								
Can you make a telephone call?			-.43		**-.53**		-.49	
Do you get tired from speaking?	.38							
Do you avoid a smoky room?								
Do you avoid a large party?								

Furthermore, scores on "Does your voice change from day to day?" correlated highly with an *unsteady* voice as rated by the speakers and their partners. The voice quality rating scale *unsteady -- steady* correlated also with the ability of making a telephone call rated by the partners as did *breathiness* and the vocal function measures mucosal wave and nonvibrating portion of the vocal fold.

8.4 Discussion

In the present chapter, self-ratings of vocal performance of patients who were treated with radiotherapy for early glottic cancer were investigated before radiotherapy and 6 months to 10 years after radiotherapy compared to control speakers. A trend was found that the deteriorated vocal performance of patients before radiotherapy improved after treatment but remained worse than vocal performance of control speakers. Patients experienced more day to day voice changes and vocal fatigue, their ability to shout, have normal conversation, use their voices in normal work routine was worsened and they avoided a smoky room and a large party more, compared to control speakers. This trend was found to be similar to the trend that was found for voice characteristics of patients before and after radiotherapy (Chapter 7).

Vocal performance following radiotherapy was influenced by stripping rather than biopsying for initial diagnosis, continuing smoking after treatment, and radiation dose schedule (66Gy/33fr. or 60Gy/30fr. instead of 60Gy/25fr.). The fact that the ability of making a phone call was diminished for patients who continued smoking after treatment confirms the findings by

Benninger et al. (1984) that smoking after treatment increased the chance on voice changes. Similar results were described in Chapter 7, where regularity of the vocal fold edge was found to decrease in patients who continued smoking after treatment. The effect of stripping the vocal fold on the ability of having normal conversation and on increased vocal fatigue also confirmed findings described in Chapter 7 and reported by Benninger et al. (1984) on voice characteristics.

Furthermore, the fact that patients treated with 60Gy in 25 fractions reported better vocal performance than the others was also found for regularity of the vocal fold edge (Chapter 7). As was already argued in Chapter 7, these results remain unclear. From a radiobiological point of view one would expect to find either no differences or increased late effects in this particular group. An explanation might be that, since most of the patients in this group were treated 7-10 years ago, this group consists of survivors; all patients with local recurrence or other complications were therefore not included, while these problems might still occur in patients in the other groups. Furthermore, 92% of the patients in this group stopped smoking, against 77% and 50% of the patients treated with the other schedules. It is clear that in future research prospective studies should give more insight in vocal performance and voice characteristics following radiotherapy on the long term, although including control speakers remains of great importance to assess what normal vocal performance comprehends.

Furthermore, in the present study, no attempt was made to determine psychometric aspects, like validity, reliability and feasibility of the vocal performance scales. Also, vocal performance is only a part of quality of life aspects and a broad range of items in social, emotional, psychological, and physical functioning should be included in quality of life assessment of patients following radiotherapy for early glottic cancer (Browman et al., 1993; Morton, 1996). Further development of vocal performance and quality of life instruments is therefore required.

Another approach which we started in the present study, is the investigation of predictability of vocal performance aspects by voice quality or vocal function measures. High correlations between vocal performance ratings and voice quality and vocal function measures were found, which is promising for future research. Clinically, it would be an important advantage if vocal performance aspects could be predicted objectively by means of voice characteristics.

8.5 Conclusion

Vocal performance aspects related to voice characteristics was decreased before radiotherapy, improved after treatment but remained worse than vocal performance experienced by control speakers. Surgery for initial diagnosis and smoking after treatment deteriorate vocal performance following radiotherapy. High correlations were found between vocal performance aspects and voice characteristics, which results are promising for future research on predictability of vocal performance by objective voice quality and vocal function measures.

9

GENERAL DISCUSSION

Abstract

In this final chapter the multidimensional character of voice characteristics is discussed and the impact on clinical practice is described. Furthermore, voice characteristics following radiotherapy are placed in a broader perspective. In the conclusions, suggestions for future research are given. Finally, a concept protocol is presented.

9.1 Introduction

A patient's outcome of rehabiliation following radiotherapy for early glottic cancer can be assessed along a number of dimensions. The present thesis is a description of one of these dimensions, namely voice. As Hirano & Bless (1993) stated, voice is a complex phenomenon that requires multiple measures to describe its characteristics. Results of this thesis confirmed the multidimensional character of voice. Issues addressed herein are not to be interpreted as all-inclusive, but rather as a first attempt to assess voice characteristics following radiotherapy for early glottic cancer. While discussing the implications of our findings, we should be well aware of the limitations of the present study. First of all, a selection was made of voice parameters. Because of the emphasis on voice quality, perceptual and acoustical analyses were investigated in detail, leaving less room for analysis of vocal function and vocal performance. Secondly, the number of patients and control speakers was still rather small in the present study, although better balanced than in most other studies. One of the consequences is that no interaction effects of variables could be taken into account in investigating voice characteristics following radiotherapy. Thirdly, since the study was set up as part of a prospective clinical trial more than 6 years ago, no benefit could be gained from recent improvements of clinical tests.

In the present chapter the multidimensional voice characteristics are discussed and the impact of our findings on clinical practice is given. Furthermore, voice characteristics following radiotherapy are described. Final conclusions complete this chapter.

9.2 Multidimensional voice characteristics

In the present thesis descriptions are given of voice quality, vocal function and vocal performance. Each of these three dimensions of voice characteristics can once more be seen as multidimensional.

Voice quality can be seen as a compilation of perceptual ratings on various scales, accompanied by acoustical measures. Perceptual analyses of voice quality by trained raters on read-aloud text, were found to describe voice quality following radiotherapy best. The fact that only moderate correlations between perceptual and acoustical measures of voice quality were found in the literature and in the present study, is assumed to be a shortcoming of acoustical analysis. More research is needed on acoustical analysis of running speech of pathological voices, because the dysphonia following radiotherapy is probably originating from the dynamic part of speech. On the other hand, when trained raters are not available, objective acoustical analyses of voice quality are quick to perform with results coming close to ratings by naive raters on sustained vowels. Moreover, when average speaking fundamental frequency is investigated, acoustical analysis is preferable over perceptual analysis. It should be emphasized though that acoustical analysis can be a tricky business and should be carried out by an experienced investigator.

In the present thesis, four tests of *vocal function* were used. Evaluation of video-laryngo-stroboscopy was found to be highly valuable, although generally it is an invasive investigation for the patient and one can question the naturalness of phonation during these circumstances. As stated earlier, a battery of tests is needed to investigate vocal function. However, maximum phonation time, phonation quotient, and Intensity Range and Frequency Range as derived from the Voice Range Profile did not reveal adequate information. Future research may involve other vocal function tests, like digital high speech glottography, videokymography, oral airflow or laryngeal resistance. Relations between vocal function and voice quality appeared to be moderate to high.

Self-ratings of *vocal performance* were investigated in the present thesis and revealed striking results of vocal performance of patients following radiotherapy. Because reliability and validity of these vocal performance scales was not investigated, it is almost needless to mention that more research is needed to improve instruments to measure these vocal performance aspects properly. However, since ratings on vocal performance scales correlated strongly with voice quality and vocal function measures, another approach was suggested, namely to predict vocal performance aspects from voice quality and vocal function measures.

9.3 Voice characteristics following radiotherapy

Cancer is a chronic, life-threatening disease that requires treatment that can cause considerable side effects, whereas cure is seldom guaranteed. The typical interval from the time of diagnosis to that of treatment is short, which means that the patient must cope with various issues concerning the acknowledgment of cancer as well as aspects of treatment over a very brief period. Denial is not uncommon at the time of diagnosis; furthermore, it is common experience that once the word cancer (or another serious disease) is expressed, the ability to comprehend additional information is greatly reduced if not totally blocked. Concerning patients following radiotherapy for early glottic cancer, de Boer et al. (1995) found that coping with adequate information from the clinician, together with open discussion of the illness in the family and social support, can improve rehabiliation outcome. These aspects were all experienced during recording sessions with the patients who participated in the present study. Although the response varied from person to person, all individuals seemed to be confused, asked questions, and above all, wanted to talk about their disease, when they came for the first recording session (before treatment). After treatment (from 6 months to 10 years), most patients seemed to cope with their situation, although they still feared a laryngectomy or worse. Although I was not qualified to answer most of the questions, because they were about medical topics, my listening ear was of great importance (at least that is my interpretation in coping with being employed as a researcher on a project that did not have any direct advantages for these individual patients).

Next to these observations during recording sessions, results as described in the present thesis reveal that voice quality, vocal function as well as related quality of life aspects may be deteriorated for patients 6 months to 10 years following radiotherapy. Also, de Boer et al. (1995) found that even though several years had elapsed since treatment, patients diagnosed with early glottic cancer reported many physical complaints following radiotherapy. Furthermore, de Boer et al. found that the greater the time elapse since treatment, the fewer the psychosocial problems. These findings were confirmed by the present study: patients treated 7 to 10 years ago reported less problems than patients 6 months to 7 years after treatment. However, most of the patients treated 7 to 10 years ago, were treated with a different radiation schedule (60 Gy in 25 fractions, instead of 66 Gy in 33 fractions or 60 Gy in 30 fractions), which also deteriorated voice related problems.

Without any doubt, stripping the vocal fold rather, than biopsying the vocal fold for initial diagnosis, deteriorated voice characteristics: voice quality was more rough, the mucosal wave on the vocal fold was more frequently disturbed, and patients reported more often vocal fatigue and an inability to have normal conversation. Smoking after treatment negatively influenced the vocal fold edge and the ability to make a phone call. Increasing age of the speaker deteriorated voice quality. Unfortunately, the patient sample was too small to investigate interactions between these variables. Future research involving a larger number of patients should give more insight in the effect of these variables on voice characteristics.

9.4 Conclusions

The present study reveals a deterioration of voice characteristics of patients six months to seven years after radiotherapy. Voice quality, vocal function and vocal performance become normal in 55% of the patients, while 45% of the patients remain pathological. Therefore, more attention for patients following radiotherapy is relevant in clinical practice. Besides careful balancing the advantage and disadvantage of stripping the vocal fold for initial diagnosis and emphasising the negative effect of continuing smoking on voice characteristics (and, of even more importance, on the increased possibility of a recidive), also long-term counseling by a speech-langage therapist or a social officer might help some of the patients in coping with their disease. Future prospective research is needed to investigate voice characteristics following various radiation dose schedules on the long term.

The results in the present thesis underscore the need of a multidimensional approach in investigating voice characteristics following radiotherapy. Because of this multidimensionality, an *analysis protocol* should involve several voice dimensions. Based on the research presented in this thesis, the protocol should at least contain perceptual analyses on running speech by trained raters following a standard protocol, acoustical analyses of sustained vowels by means of standardized equipment carried out by an experienced investigator, recordings and evaluations of video-laryngo-stroboscopy carried out by an experienced physician, and self-ratings of vocal performance by the patients following a standard protocol.

Although more research is needed on reliability, validity and feasibility of some of these (and other) voice tests, this concept protocol is useful in clinical studies on the evaluation of treatment for patients diagnosed with early glottic cancer.

Summary

Prognosis concerning survival is good for patients who are treated with radiotherapy for early glottic cancer, with cure rates of 70-90%. Despite these good results, there is still uncertainty about the optimal radiation dose. The optimal dose should be based on tumour control and possible complications. Voice worsening can be a complication of radiotherapy. This thesis aims at some of the theoretical, practical, and methodological problems of voice analyses in order to assess possible outcomes of radiotherapy on voice characteristics in terms of *voice quality*, *vocal function*, and *vocal performance*.

A literature survey (Chapter 1) reveals that few studies are carried out on voice characteristics of patients following radiotherapy for early glottic cancer. In addition, results of the 19 studies reviewed are hard to compare because of methodological differences. Most striking is the variety of speakers: men and women ranging in age, with small to large tumours, treated with different radiation schedules, before, during, and right after radiation up to ten years after radiotherapy. Therefore, it is striking too that only in six studies control speakers were involved. In the other studies, patient groups were compared with themselves at various moments before and after treatment or with mean data from the literature. Furthermore, several voice analyses are applied: perceptual voice ratings, acoustical voice measurements, or clinical methods such as phonetography and stroboscopy. Although it is hard to compare results of these studies, it can be concluded that an acute effect of radiotherapy on voice characteristics has been shown, but that late effects are still obscure.

Before examining this, a description is given in Chapter 2 of the "normal" anatomy and physiology of the larynx, of early glottic cancer, and of the treatment this thesis focuses on: radiotherapy. Also, the trial study is described, that is carried out at the Netherlands Cancer Institute/Antoni van Leeuwenhoekhuis and that deals with the effect of two different radiation schedules for early glottic carcinoma; this thesis is part of that trial study.

Chapter 3 comprises a detailed description of the 60 patients and 20 control speakers who have participated in this research project. Because voice characteristics are speaker dependent, a group of ten patients is followed from before radiation, six months after up to two years after radiotherapy (n=30). Further follow-up of these patients fell out of the range of the project, but because possible late effects should become visible or audible as well, five separate groups of patients were composed: before radiation, six months after, two years after, three to seven years after, and seven to ten years after radiotherapy (n=50). Moreover, 20 control speakers were willing to partici-

pate in the project; these speakers were matched with the patients concerning sex (all male), age (between 51 and 81 years old), and smoking and drinking habits. The group arrangement is applied to develop a protocol of voice analyses, in the course of which it is investigated which analyses can differentiate these speaker groups best. Subsequently, voice characteristics following radiotherapy are examined even more precisely, dependent on five aspects: stage of the tumour (unilateral or bilateral), initial surgery (biopsying or stripping the vocal fold), radiation schedule (66 Gy in 33 fractions, 60 Gy in 30 fractions, or 60 Gy in 25 fractions), age of the speaker (younger than 65 years, between 65 and 70 years, between 70 and 75 years, or older than 75 years), and whether or not smoking was continued after treatment. But before these aspects are discussed, first a description is given of the development of the protocol concerning perceptual analyses of voice quality (Chapter 4), different pitch analyses (Chapter 5), and acoustical analyses of voice quality (Chapter 6).

Chapter 4 deals with perceptual analyses of *voice quality*. Ratings from three trained and 20 naive raters and from the speakers themselves and their partners are gathered. The trained raters are trained in the use of the 'Vocal Profile Analysis Protocol' by John Laver; the naive raters and the speakers themselves and their partners judge voice quality on seven-points scales that are especially developed for naive Dutch raters. The trained and naive raters judge voice quality on read-aloud text and on sustained /a/ vowels. Trained raters are found to be more reliable than naive raters, but reliability is satisfactory for both rater groups; reliability could neither be assessed for the ratings of the speakers themselves nor for their partners, since they rated just one voice at the time. Furthermore it appears that patients before radiotherapy have the most deviant voice quality; voice quality of patients six months, two years, and three to seven years after radiation is less deviant, but still significantly worse than voice quality of the control speakers; patients seven to ten years after radiotherapy are comparable with control speakers. This trend is found most obviously for the trained raters on read-aloud text on the scales *breathiness*, *roughness*, and *tension*. The conclusion is that perceptual analysis of voice quality by trained raters is preferred.

It would seem that voice quality can be analysed by means of perceptual judgements. However, there are still certain shortcomings attached to this method. Even though reliability of the raters has been shown, their ratings remain subjective. Furthermore, perceptual analyses are very time-consuming, which is a considerable drawback, especially in clinical practice. Sufficient reason to draw the attention to *acoustical analyses* of voice quality, which are objective and quick to perform. In Chapter 5, a closer look is taken at pitch analysis. Perceptual, acoustical, and electroglottographic analyses are compared. Earlier research revealed that perceptual pitch ratings may be influenced by deviant voice quality. Acoustical analyses of fundamental frequency (pitch is the audible feature we attach to differences in fundamental frequency) are probably less disturbed by deviant voice quality. However, acoustic signals do contain strong harmonics due to the resonant frequencies of the vocal tract (oral/pharynx cavity) which may hamper 'pitch extraction'. Electroglottographic (EGG) signals represent vocal fold activity (and thereby fundamental frequency) more directly and are therefore taken into account to determine which method can best be used to analyse pitch of pathological voices. Results show that perceptual analyses are indeed influenced by deviant voice quality. Raters have problems particularly with rough voices: these are often judged as lower, while they are not that low. Results from the objective acoustic and electroglottographic analyses are comparable, provided that the analyses are well performed. Nevertheless, preference is given to acoustical pitch analysis, because no reliable EGG-signals could be obtained from more than 20% of the speakers.

In Chapter 6, acoustical analyses of voice quality are further examined. By means of the speech processing system PRAAT developed by Boersma (Institute of Phonetic Sciences) the mean fundamental frequency and the harmonics-to-noise ratio are analysed. Besides that, the commercially available package Multidimensional Voice Program (MDVP) provides a series of parameters that are grouped under fundamental frequency, frequency and amplitude perturbation (jitter and shimmer), voice breaks, voice irregularities, noise, and tremor. Finally, a new parameter is used: duration of voice onset of the sustained /a/; this is measured manually. Again, results are compared with perceptual ratings (breathiness, roughness, and tension) by trained and naive raters on read-aloud text and the sustained /a/, to determine which analyses can best be used. It appears that acoustical analyses (especially standard deviation of the fundamental frequency, jitter, noise, and duration of the voice onset) show the same trend as was found for the perceptual ratings, albeit less strong. Direct single correlations between acoustical and perceptual voice parameters are low; results of multiple regression analyses show that a perceptual parameter can be predicted better by a set of acoustical measures. The conclusion is that, in the case of separate speaker groups, voice quality can best be analysed by means of scale judgements by trained raters. For a longitudinal research design, acoustical measures are objective and quick to perform and come close to judgements by naive raters.

Besides analyses of voice quality, measures of *vocal function* are also of interest in investigating the effect of radiotherapy on voice characteristics. In Chapter 7 the phonetogram, maximum phonation time, phonation quotient, and evaluations of video-laryngo-stroboscopy are used to investigate vocal function. It appears that frequency and amplitude range, measured by means of phonetography, maximum phonation time, and phonation quotient give insufficient insight into vocal function following radiotherapy. These measures are left aside. Stroboscopy, on the other hand, although unpleasant for the speaker and therefore not available for all speakers, does give a lot of information. It appears that patients after radiotherapy have more glottic oedema and more vascular injection on the vocal fold and that the vocal fold edge is often irregular, that the mucosal wave is often diminished, that a nonvibrating portion of the vocal fold is often present, and that vocal fold closure is often incomplete. Furthermore, it appears that in addition to increasing age of the speaker and stripping instead of biopsying the vocal fold (which was also found to have an adverse effect for perceptual analyses of voice quality), also continuing smoking after radiotherapy decrease vocal function.

In Chapter 8 the effect of a voice disorder on daily life is investigated. The speakers are asked to indicate their *vocal performance* by means of self-ratings on several scales, such as the ability to shout, have a normal (telephone) conversation, the amount of getting tired from speaking, and the avoidance of a large party. Their answers were compared with the earlier derived measures for voice quality and vocal function. Once again it appears that patients before radiotherapy experienced decreased vocal performance, which improved for patients six months to seven years after radiation but remained worse than vocal performance as reported by control speakers. Also, it appears again that diagnostic stripping instead of biopsying the vocal folds and continuing smoking after treatment have an adverse effect on vocal performance following radiotherapy.

The conclusion of this thesis (Chapter 9) is that voice characteristics remain worse for almost half of the patients six months to seven years after radiotherapy compared to control speakers. Carefully balancing the advantage and disadvantage of stripping the vocal fold for initial diagnosis and emphasising the negative effect of continuing smoking is thereby of interest. Furthermore, it appears that because of the multidimensional character of voice, an *analysis protocol* should comprise multiple voice measures. Based on the findings in this thesis, this protocol should comprise at least perceptual ratings of voice quality by trained raters on running speech, preferably complemented with acoustical measures, evaluations of stroboscopic video-recordings of vocal function, and self-ratings of vocal performance. Although more research is needed on reliability, validity, and feasibility of (other) voice analysis methods, this concept protocol is useful in clinical studies on the evaluation of treatment for patients diagnosed with early glottic cancer.

Samenvatting

De prognose voor wat betreft de levensduur van patiënten die met radiotherapie behandeld zijn voor een klein larynxcarcinoom, is goed: 70-90% van hen geneest. Ondanks deze goede resultaten bestaat er nog steeds onzekerheid over de optimale bestralingsdosis. De optimale dosis moet gebaseerd zijn op tumorcontrole en op eventuele complicaties. Eén van de complicaties van radiotherapie kan verslechtering van de stem zijn. Dit proefschrift richt zich op enkele van de theoretische, praktische en methodologische problemen van stemanalyses zodat mogelijke uitkomsten van radiotherapie op stemkarakteristieken vastgesteld kunnen worden in termen van *stemkwaliteit*, *stemfunctie* en *stemgebruik*.

Uit een literatuuronderzoek (hoofdstuk 1) blijkt dat er nog maar weinig onderzoek gedaan is naar stemkarakteristieken na bestraling van patiënten met een klein larynxcarcinoom. Bovendien zijn de resultaten van de 19 gevonden studies moeilijk te vergelijken door methodologische verschillen. Het eerste wat opvalt bij vergelijking van de studies, is de verscheidenheid aan sprekers: mannen en vrouwen variërend in leeftijd, met kleine tot grote tumoren, behandeld met verschillende bestralingsschema's, vóór, tijdens en vlak na de bestraling tot tien jaar na de bestraling. Opvallend is ook, dat maar in zes studies controlesprekers zijn opgenomen. In de andere studies worden de patiëntengroepen met zichzelf op verschillende momenten voor en na de behandeling of met gemiddelde data uit de literatuur vergeleken. Verder worden verschillende stemanalyses uitgevoerd: perceptieve stembeoordelingen, akoestische stemmetingen of klinische methoden, zoals fonetografie en stroboscopie. Alhoewel het dus moeilijk is om resultaten van deze studies te vergelijken, kan toch geconcludeerd worden, dat een acuut negatief effect van radiotherapie op stemkarakteristieken is aangetoond, maar dat effecten op langere termijn nog onduidelijk zijn.

Voordat daarop wordt ingegaan, volgt eerst in hoofdstuk 2 een beschrijving van de 'normale' anatomie en fysiologie van de larynx, van het kleine larynxcarcinoom, en van de behandeling waar het in dit proefschrift om gaat: radiotherapie. Ook wordt de trialstudie beschreven, die uitgevoerd wordt in het Nederlands Kanker Instituut/Antoni van Leeuwenhoekhuis en die gaat over het effect van twee verschillende bestralingsschema's voor kleine larynxcarinoma; dit proefschrift is een onderdeel van deze trialstudie.

Hoofdstuk 3 omvat een gedetailleerde beschrijving van de 60 patiënten en 20 controlesprekers die hebben meegewerkt aan dit onderzoek. Omdat stemkarakteristieken sprekerafhankelijk zijn, wordt een groep van tien patiënten longitudinaal gevolgd: vlak vóór, zes maanden na en twee jaar na radiotherapie (n=30). Het langer volgen van deze patiënten valt buiten het project, maar omdat ook eventuele effecten op langere termijn zichtbaar dan wel hoorbaar moeten worden, worden ook vijf verschillende groepen van tien patiënten samengesteld: vóór bestraling, zes maanden na, twee jaar na, drie tot zeven jaar na en zeven tot tien jaar na bestraling (n=50). Bovendien worden 20 controlesprekers bereid gevonden aan het onderzoek mee te werken; deze sprekers zijn vergelijkbaar met de patiënten voor wat betreft sexe (allen mannen), leeftijd (tussen de 51 en 81 jaar) en rook- en drinkgewoontes. De groepsindeling wordt gebruikt om een protocol voor stemanalyses te ontwikkelen, waarbij nagegaan wordt welke analyses deze groepen het best differentiëren. Vervolgens worden vijf aspecten beschreven waarmee in een later stadium stemkarakteristieken van patiënten na radiotherapie nog nauwkeuriger geanalyseerd worden: de grootte van de tumor (unilateraal of bilateraal), diagnostische operatie (het nemen van een biopt of strippen), het bestralingsschema (66 Gy in 33 fracties, 60 Gy in 30 fracties of 60 Gy in 25 fracties), de leeftijd van de spreker (jonger dan 65 jaar, tussen 65 en 70 jaar, tussen 70 en 75 jaar en ouder dan 75 jaar) en het al dan niet blijven roken na de behandeling. Maar voordat deze aspecten aan de orde komen, volgt eerst een beschrijving van de ontwikkeling van het protocol voor wat betreft perceptieve analyses van stemkwaliteit (hoofdstuk 4), verschillende toonhoogte-analyses (hoofdstuk 5) en akoestische analyses van stemkwaliteit (hoofdstuk 6).

In hoofdstuk 4 gaat het dus om perceptieve analyses van *stemkwaliteit*. Beoordelingen worden verkregen van drie getrainde en twintig ongetrainde luisteraars, en van de sprekers zelf en hun partners. De getrainde luisteraars zijn getraind in het gebruik van het 'Vocal Profile Analysis Protocol' van Laver; de ongetrainde luisteraars en de sprekers zelf en hun partners beoordelen de stemmen op 7-puntsschalen die speciaal zijn ontwikkeld voor ongetrainde Nederlandse luisteraars. De getrainde en ongetrainde luisteraars beoordelen de stemkwaliteit op voorgelezen tekst en op een aangehouden /a/. De getrainde luisteraars blijken betrouwbaarder dan de ongetrainde luisteraars, maar de betrouwbaarheid is voor beide groepen voldoende; van de sprekers zelf en hun partners kan de betrouwbaarheid niet gemeten worden, omdat ze slechts één stem op één moment beoordeelden. Verder blijkt dat patiënten vóór bestraling de meest afwijkende stemkwaliteit hebben; stemkwaliteit van patiënten zes maanden na, twee jaar na en drie tot zeven jaar na bestraling is minder slecht, maar nog steeds significant slechter dan die van de controlesprekers; patiënten zeven tot tien jaar na bestraling blijken vergelijkbaar met controlesprekers. Deze trend wordt het duidelijkst aangegeven door de getrainde luisteraars op voorgelezen tekst op de schalen *heesheid*, *schorheid* en *gespannenheid*. De conclusie is dat perceptieve analyses door getrainde luisteraars op voorgelezen tekst de voorkeur verdienen.

Stemkwaliteit kan dus geanalyseerd worden door middel van perceptieve beoordelingen. Er kleven echter wel nadelen aan deze analysemethode. Ook al is de betrouwbaarheid van de luisteraars aangetoond, hun oordelen blijven subjectief. Verder is het zeer arbeidsintensief, hetgeen vooral in de kliniek een groot bezwaar is. Reden dus om in dit proefschrift aandacht te schenken aan *akoestische stemanalyses,* die objectief zijn, en snel uit te voeren. In hoofdstuk 5 worden toonhoogte-analyses onder de loep genomen. Perceptieve, akoestische en electroglottografische analyses worden vergeleken. Uit eerder onderzoek bleek dat perceptieve toonhoogtebeoordelingen mogelijk worden beïnvloed door afwijkende stemkwaliteit. Akoestische grondfrequentiemetingen (toonhoogte is de hoorbare eigenschap die we toekennen aan verschillen in grondfrequentie) worden wellicht minder verstoord door afwijkende stemkwaliteit. Maar akoestische signalen bevatten wel sterke harmonischen afkomstig van resonanties in het aanzetstuk (mond-keelholte) die de 'pitchextractie' kunnen bemoeilijken. Electroglottografische (EGG) signalen representeren de stembandactiviteit (en daarmee de grondfrequentie) directer en worden daarom meegenomen om te bepalen welke toonhoogteanalyses het best gebruikt kunnen worden voor pathologische stemmen. De resultaten laten zien dat perceptieve analyses inderdaad worden beïnvloed door afwijkende stemkwaliteit. Luisteraars hebben met name moeite met schorre stemmen: deze worden vaak als lager beoordeeld dan ze in werkelijkheid zijn. De objectieve akoestische en electroglottografische analyses zijn, mits ze goed uitgevoerd worden, vergelijkbaar. Toch wordt de voorkeur gegeven aan akoestische analyses, omdat van ruim 20% van de sprekers geen betrouwbare EGG-signalen verkregen konden worden.

In hoofdstuk 6 wordt verder ingegaan op akoestische analyses van stemkwaliteit. Door middel van het spraakverwerkingsprogramma PRAAT (ontwikkeld door Boersma (Instituut voor Fonetische Wetenschappen)) worden de gemiddelde grondfrequentie en harmonischen-ruisverhouding geanalyseerd. Daarnaast levert het commercieel verkrijgbare pakket Multidimensional Voice Program (MDVP) een reeks parameters die worden gegroepeerd onder grondfrequentie, frequentie- en amplitudeperturbatie (jitter en shimmer), stembreuken, stemonregelmatigheden, ruis en tremor. Tenslotte wordt een nieuwe maat gebruikt: duur van de steminzet bij de aangehouden /a/; deze wordt handmatig gemeten. Weer worden resultaten vergeleken met perceptieve beoordelingen (heesheid, schorheid en gespannenheid) gegeven door getrainde en ongetrainde luisteraars op de voorgelezen tekst en op de aangehouden /a/. Doel is om te bepalen welke analyses het best gebruikt kunnen worden. Het blijkt dat akoestische analyses (met name de standaarddeviatie van de grondfrequentie, jitter, ruis en de duur van de steminzet) eenzelfde trend laten zien als de perceptieve beoordelingen, zij het minder sterk. Directe enkelvoudige correlaties tussen akoestische en perceptieve stemparameters zijn laag; resultaten van multipele regressie-analyses laten zien dat een perceptieve parameter wel beter wordt voorspeld door een combinatie van akoestische maten. De conclusie luidt dat, in het geval van verschillende sprekergroepen, stemkwaliteit het best geanalyseerd kan worden door middel van schaaloordelen van getrainde luisteraars. Mochten getrainde luisteraars niet beschikbaar zijn en is er sprake van een longitudinaal opgezet onderzoek, dan zijn akoestische metingen objectief en snel uit te voeren, en komen de resultaten dicht bij de beoordelingen van ongetrainde luisteraars.

Naast analyses van stemkwaliteit zijn ook *stemfunctie*metingen van belang om het effect van radiotherapie op stemkarakterstieken vast te stellen. In hoofdstuk 7 worden het fonetogram, de maximale fonatietijd, het fonatiequotiënt en evaluaties van video-laryngo-stroboscopie gebruikt om de stemfunctie te onderzoeken. Het blijkt dat het frequentie- en amplitudebereik, gemeten door middel van fonetografie, de maximale fonatietijd en het fonatiequotiënt onvoldoende inzicht geven in de stemfunctie na radiotherapie. Deze worden verder buiten beschouwing gelaten. Stroboscopie daarentegen, alhoewel onprettig voor de spreker en niet te verkrijgen van alle sprekers, geeft wel veel informatie. Het blijkt dat patiënten na radiotherapie meer glottisch oedeem en meer vaatinjectie op de stembanden hebben, en dat de stembandrand vaak onregelmatig blijft, dat de mucosagolf vaak verminderd blijft, dat vaak een deel van de stemband niet meetrilt, en dat de stembandsluiting vaak onvoldoende is. Verder blijkt dat naast het ouder worden van de spreker en het strippen in plaats van biopteren van de stemband (dat ook als nadelig effect uit de perceptieve analyses van stemkwaliteit naar voren komt), ook het blijven roken na radiotherapie de stemfunctie nadelig beïnvloedt.

In hoofdstuk 8 wordt nagegaan of patiënten in hun dagelijks leven veel last hebben van hun stemaandoening. De sprekers wordt gevraagd om aspecten van hun *stemgebruik* aan te geven door middel van zelfbeoordelingen op een aantal schalen, zoals kunnen roepen, normaal een (telefoon)gesprek kunnen voeren, moe worden van spreken, en het vermijden van een groot gezelschap. Hun antwoorden worden vergeleken met de al eerder verkregen stemkwaliteit- en stemfunctiemetingen. Opnieuw blijkt dat patiënten vóór radiotherapie het meest last hebben van hun stem in dagelijks gebruik en dat patiënten zes maanden tot zeven jaar na radiotherapie minder last hebben maar nog altijd meer dan controlesprekers. Patiënten zeven tot tien jaar na bestraling zijn weer vergelijkbaar met de controlesprekers. Ook blijkt weer dat het diagnostisch strippen van de stemband in plaats van biopteren en het blijven roken na behandeling een negatieve invloed hebben op stemgebruik na bestraling.

De conclusie van dit proefschrift (hoofdstuk 9) is dat stemkarakteristieken van bijna de helft van de patiënten zes maanden tot zeven jaar na radiotherapie vergeleken met controlesprekers slechter blijven. Hierbij blijkt dat het van belang is om de voor- en nadelen van diagnostisch strippen van de stemband zorgvuldig af te wegen en om het nadelige effect van het blijven roken na de behandeling te benadrukken. Verder blijkt dat, omdat stem multidimensionaal is, een *onderzoekprotocol* dus meerdere stemanalyses moet omvatten. Gebaseerd op de bevindingen van dit proefschrift, dient dit protocol tenminste te bestaan uit perceptieve beoordelingen van stemkwaliteit door getrainde luisteraars op lopende spraak, aangevuld met akoestische metingen, evaluaties van stroboscopische video-opnames van stemfunctie en zelfbeoordelingen van stemgebruik. Alhoewel meer onderzoek nodig is naar de betrouwbaarheid, validiteit en toepasbaarheid van (andere) stemanalysemethoden, is dit conceptprotocol bruikbaar in klinische studies betreffende de evaluatie van behandeling van patienten met een klein larynxcarcinoom.

References

Ackerstaff, A.H., Hilgers, F.J.M., Aaronson, N.K. & Balm, A.J.M. (1994). Communication, functional disorders and life style changes after total laryngectomy. *Clinical Otolaryngology* 19, 295-300.

Airaner, R. & Klingholz, F. (1993). Quantitative evaluation of phonetograms in the case of functional dysphonia. *Journal of Voice* 7, 2, 136-141.

Askenfelt, A., Gauffin, J., Sundberg, J. & Kitzing, P. (1980). A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency. *Journal of Speech and Hearing Research* 23, 268-273.

Baris, G., Bartelink, H., Daal, W.A.J. van, Eschwege, F., Karim, A.B.M.F., Bogaert, W. van den (1986). *A randomised trial on dose response in radiotherapy for early glottic cancer*. Proposal for a research project. Netherlands Cancer Institute.

Bassich, C.J. & Ludlow, C.L. (1986). The use of perceptual methods by new clinicians for assessing voice quality. *Journal of Speech and Hearing Disorders* 51, 125-133.

Benninger, M.S., Gillen J., Thieme, P., Jacobson, B. & Dragovich, J. (1994). Factors associated with recurrence and voice quality following radiation therapy for T1 and T2 glottic carcinomas. *Laryngoscope* 104, 294-294.

Bezooijen, R. van. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38, 253-266

Boer, M.F. de, Pruyn, J.F.A., Borne, B. van den, Knegt, P.P., Ryckman, R.M. & Verwoerd, C.D.A. (1995). Rehabilitation outcomes of long-term survivors treated for head and neck cancer. *Head & Neck* 17, 503-515.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 17, 97-110.

Boersma, P. & Weenink, D.J.M. (1996). *Praat, a system for doing phonetics by computer, version 3.4.* Report of the Institute of Phonetic Sciences Amsterdam 132, 182 pp..
(Also on Internet at *http://fonsg3.let.uva.nl/paul/praat.htm*).

Bjordal, K., Kaasa, S. & Mastekaasa, A. (1994). Quality of life in patients treated for head and neck cancer: a follow-up study 7 to 11 years after radiotherapy. *International Journal of Radiation Oncology, Biology, Physics* 28, 4, 847-856.

Browman, G.P., Levine, M.N., Hodson, I., Sathya, J., Russell, R., Skingley, P., Cripps, C., Eapen, L. & Girard, A. (1993). The head and neck radiotherapy questionnaire: a morbidity/quality-of-life instrument for clinical trials of radiation therapy in locally advanced head and neck cancer. *Journal of Clinical Oncology* 11, 5, 863-872.

Browman, G.P., Wong, G. & Hodson, I. (1993). Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. *New England Journal of Medicine* 328, 3, 159-163.

Coebergh, J.W.W., Heijden, L.H. van der & Janssen-Heijnen, M.L.G. (eds) (1995). *Cancer incidence and survival in the Southeast of the Netherlands 1955-1994. A report from the Eindhoven Cancer Registry 1995.* Integraal Kanker Centrum Zuid.

Colton, R.H., Sagerman, R.H., Chung, C.T., Yu, Y.W. & Reed, G.F. (1978). Voice change after radiotherapy. *Radiology* 127, 821-824.

Colton, R.H. & Conture, E.G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice* 4, 10-24.

Cordes, A.K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research* 37, 264-278.

Dagli, A.S., Mahieu, H.F. & Festen, J.M. (1997). Quantitative analysis of voice quality in early glottic laryngeal carcinomas treated with radiotherapy. *European Archives of Otorhinolaryngology* 254, 78-80.

Deal, R. & Emanuel, F.W. (1978). Some waveform and spectral features of vowel roughness. *Journal of Speech and Hearing Research* 21, 250-264.

Deliyski, D.D. (1993). Acoustic model and evaluation of pathological voice production. *Proceedings of the 3rd International Conference on Speech Communication and Technology, EUROSPEECH '93, Berlin, Germany.*

Emanuel, F.W. & Smith, W.F. (1974). Pitch effects on vowel roughness and spectral noise. *Journal of Phonetics* 2, 247-253.

Erp, A.J.M. van (1991). *The phonetic basis of personality ratings, with specific reference to cleft-palate speech.* Dissertation, PTT Research, Leidschendam.

Fagel, W., Herpt, L. van & Boves, L. (1983). Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation. *Speech Communication* 2, 315-326.

Fant, G. (1960). *The acoustic theory of speech production.* (2nd edition, 1970). Mouton, 's Gravenhage.

Feijoo, S. & Hernandez, C. (1990). Short-term stability measures for the evaluation of vocal quality. *Journal of Speech and Hearing Research* 33, 324-334.

Ferlito, A. (1993). The world health organization's revised classification of tumours of the larynx, hypopharynx, and trachea. *Annals of Otology, Rhinology and Laryngology* 102, 666-669.

Ferlito, A., Carbone, A., DeSanto, L.W., Barnes, L., Rinaldo, A., D'Angelo, L. & Devaney, K.O. (1996). "Early" cancer of the larynx: the concept as defined by clinicians, pathologists, and biologist. *Annals of Otology Rhinology and Laryngology* 105, 245- 250.

Fex, S. & Henriksson, B. (1970). Phoniatric treatment combined with radiotherapy of laryngeal cancer for the avoidance of radiation damage. *Acta Oto-laryngologica* 263, 128-129.

Fourcin, A. J. (1993). Normal and pathological speech: phonetic, acoustic, and laryngographic aspects. In: Shingh, W. & Soutar, D.S. (eds). *Functional surgery of the larynx and pharynx.* Butterworth-Heineman, Oxford.

Franken, M.C., Boves, L., Peters, H.F.M. & Webster, R.L. (1995). Perceptual rating instrument for speech evaluation of stuttering treatment. *Journal of Speech and Hearing Research* 38, 280-288.

Gelfer, M. P. (1995). Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. *Journal of Speech and Hearing Research* 38, 1189-1198.

Gobl, C. & Ni Chasaide, A.N. (1988). The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross-language study. *Speech Transmission Laboratory-Quarterly Progress and Status Report* 2-3, 23-59. Stockholm: Department of Speech Communication & Music Acoustics, Royal Institute of Technology.

Goepfert, H. (1994). Debate on the preferred therapy for T1 laryngeal cancer. In: Smee, R. & Bridger, C.P. (eds). *Laryngeal Cancer. Proceedings of the 2nd world congress on laryngeal cancer, Sydney*. Elsevier, 308-310.

Gray, S.D., Hirano, M. & Sato, K. (1993). Molecular and cellular structure of vocal fold tisuue. In: Titze, I.R. (ed.). *Vocal fold physiology, frontiers in basic science.* Singular Publishing Group, Inc., San Diego, California.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. & Wendin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-laryngologica* 90, 442-451.

Hammarberg B. (1986). *Perceptual and acoustic analysis of dysphonia.* Thesis of the Department of Logopedics and Phoniatrics, Huddinge University Hospital, Stockholm.

Hammarberg, B. & Gauffin, J. (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In: Fujimura, O. & Hirano, M. (eds). *Vocal fold physiology, voice quality control.* Singular Publishing Group, Inc., San Diego, California.

Hammarberg, B. (1997). *Perceptual evaluation of dysphonic voices.* Proceedings of Larynx97, 11-18.

Hardcastle, W.J. (1976). *Physiology of speech production.* Academic Press, London.

Harrison, L.B., Solomon, B., Miller, S., Fass, D.E., Armstrong, J. & Sessions, R.B. (1990). Prospective computer-assisted voice analysis for patients with early stage glottic cancer: a preliminary report of the functional result of laryngeal irradiation. *International Journal of Radiation Oncology, Biology, Physics* 19, 123-127.

Heeneman, H., Leeper, H.A., Hawkins, S. & Nordick, K.A. (1994). Vocal function following radiotherapy for early (T1) laryngeal cancer: a prospective study. In: Smee, R. & Bridger, C.P. (eds). *Laryngeal cancer. Proceedings of the 2nd world congress on laryngeal cancer, Sydney.* Elsevier, 355-359.

Henton, C. & Bladon, A. (1985). Breathiness in normal female speech: inefficiency versus desirability. *Language and Communication* 5, 221-227.

Hill, R.P. & Tannock, I.F. (1987). Introduction: cancer as a cellular disease, In: Tannock, I.F. & Hill, R.P. (eds). *The basic science of oncology*, Pergamon Press, New York.

Hill, R.P. (1990). Experimental radiotherapy. In: Tannock, I.F. & Hill, R.P. (eds). *The basic science of oncology.* Pergamon Press, New York.

Hillenbrand, J. & Houde, R.A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research* 29, 311-321.

Hirano, M. (1981). Clinical examination of voice. *Disorders of human communication* 5. Springer-Verlag, New York.

Hirano, M., Kurita, S. & Sakaguchi, S. (1989). Ageing of the vibratory tissue of human vocal folds. *Acta Oto-laryngologica* 107, 428-433.

Hirano, M. & Bless, D.M. (1993). *Videostroboscopic examination of the larynx.* Singular Publishing Group, Inc., San Diego.

Hirano, M., Mori, K. & Iwashita, H. (1994). Voice in laryngeal cancer. In: Smee, R. & Bridger, C.P. (eds). *Laryngeal cancer. Proceedings of the 2nd world congress on laryngeal cancer, Sydney*. Elsevier, 54-64.

Hirano, M. & Mori, M. (1996). Transoral laser resection in team practice for early cancer of larynx. *Proceedings of the Fourth International Conference on Head and Neck Cancer, Toronto*, 277-283.

Horiot, J.C., Bontemps, P., Bogaert, W. van den, Le Fur, R., Weijngaert, D. van den, Bolla, M., Bernier, J., Lusinci, A., Stuschke, M., Lopez-Torrecilla, J., Begg, A., Pierart, M. & Collette, L. (1997). Accelerated fractionation (AF) compared to conventional fractionation (CF) improves loco-regional control in the radiotherapy of advanced head and neck cancers: results of the EORTC 22851 randomized trial. *Radiotherapy & Oncology* 44, 111-121.

Hoyt, D.J., Lettinga, J.W., Leopold, K.A. & Fisher, S.R. (1992). The effect of head and neck radiation therapy on voice quality. *Laryngoscope* 102, 477-480.

Karim, A.B.M.F., Snow, G.B., Siek, H.T.H. & Njo K.H. (1983). The quality of voice in patients irradiated for laryngeal carcinoma. *Cancer* 51, 47-49.

Karim, A.B.M.F., Kralendonk, J.H., Njo, K.H., Tabak, J.M. & Gort, G. (1989). A critical look at the TNM classification for laryngeal carcinoma. *Cancer* 65, 1918-1922.

Keane, T.J. (1994). Controversies in the radiotherapy of T1 and T2 glottic cancer. In: Smee, R. & Bridger, C.P. (eds). *Laryngeal cancer. Proceedings of the 2nd world congress on laryngeal cancer, Sydney*. Elsevier, 311-314.

Kiebert, G.M. (1995). *Choices in oncology: patients' valuations of treatment outcomes in terms of quality and length of life*. Ph.D.thesis University of Leiden.

Kirk, R.E. (1982). *Experimental design: procedures for the behavioral sciences.* Brooks/Cole Publishing Company, Monterey.

Kleinsasser, O. (1992). Revision of classification of laryngeal cancer, is it long overdue (proposals for an improved TN-classification). *Journal of Laryngology and Otology* 106, 197-204.

Klingholz, F. (1990). Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels. *Journal of the Acoustical Society of America* 87, 2218-2224.

Kojima, H., Gould, W., Lambiase, A. & Ishiki, N. (1980). Computer analysis of hoarseness. *Acta Oto-laryngologica* 89, 545-554.

Kreiman, J., Gerratt, B.R., Precoda, K. & Berke, G.S. (1992). Individual differences in voice quality perception. *Journal of Voice* 35, 512-520.

Kreiman, J., Gerrat, B., Kempster, G., Erman, A. & Berke, G. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research* 36, 21-40.

Krom, G. de (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *Journal of Speech and Hearing Research* 37, 985-1000.

Krom, G. de (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research* 38, 794-811.

Kuwabara H. & Sagisaka, Y. (1995). Acoustic characteristics of speaker individuality: control and conversion. *Speech Communication* 16, 165-173.

Lablance, G.R., Maves, M.D., Scialfa, T.M., Eitnier, C.M. & Steckol, K.F. (1992). Comparison of electroglottographic and acoustic analysis of pitch perturbation. *Otolaryngology-Head and Neck Surgery* 107, 5, 617-621.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press, Cambridge.

Laver, J., Wirz, S., Mackenzie, J. & Hiller, S.M. (1981). A perceptual protocol for the analysis of vocal profiles. *Edinburgh University Department of Linguistics, Work in Progress* 14, 139-155.

Laver, J. (1991). *The gift of speech*. Edinburgh University Press, Edinburgh.

Laver, J. & Hanson R. (1981). Describing the normal voice. In: Darby, J. (ed.). *Speech Evaluation in Psychiatry*. Grune and Stratton, New York.

Leeuw, I.M. de (1991). Perceptual evaluation of voice quality before and after radiotherapy of patients with early glottic cancer and of normal speakers. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 15, 109-120.

Lehman, J.J., Bless, D.M. & Brandenburg, J.H. (1986). An objective assessment of voice production after radiation therapy for stage I squamous cell carcinoma of the glottis. *Otolaryngology-Head and Neck Surgery* 98, 2, 121-129.

List, M.A., Ritter-Sterr, C.A., Baker, T.M., Colangelo, L.A., Matz, G., Pauloski, B.R. & Logemann, J.A. (1996). Longitudinal assessment of quality of life in laryngeal cancer patients. *Head & Neck* 18, 1-10.

Llewllyn-Thomas, H.A., Sutherland, H.J., Hogg, S.A., Ciampi, A., Harwood, A.R., Keane, T.J., Till, J.E. & Boyd, N. (1984). Linear analogue self-assessment of voice quality in laryngeal cancer. *Journal of Chronic Disorders* 37, 12, 917-924.

Löfqvist, A. & McGowan, R.S. (1991). Voice source variations in running speech. In: Gauffin, J. & Hammarberg, B. (eds). *Vocal fold physiology. Acoustic, perceptual, and physiological aspects of voice mechanisms.* Singular Publishing Group, Inc., San Diego.

Maier, H., Dietz, A., Heller, W.D. & Weidauer, H. (1992). Tobacco and alcohol and the risk of head and neck cancer. *Clinical Investigation* 70, 320-327.

Martin, D., Fitch, J. & Wolfe, V. (1995). Pathological voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research* 38, 765-771.

McGuirt, W.F., Blalock, D., Koufman, J.A., Feehs, R.S., Hilliard, A.J., Greven, K. & Randall, M. (1994). Comparative voice results after laser resection or irradiation of T1 vocal cord carcinoma. *Archives of Otoalaryngology-Head and Neck Surgery* 120, 951-955.

Miller, S., Harrison L.B., Solomon, B. & Sessions, R.B. (1990). Vocal changes in patients undergoing radiation therapy for glottic carcinoma. *Laryngoscope* 100, 603-606.

Million, R.R., Mendenhall, W.M., Parsons, J.T., Fein, D.A., Cassisi, N.J. & Stringer, S.P. (1994). Comparison of radiation therapy and surgery in the management of T1-T2 vocal cord carcinoma. In: Smee, R. & Bridger, C.P. (eds). *Laryngeal cancer. Proceedings of the 2nd world congress on laryngeal cancer, Sydney.* Elsevier, 303-307.

Mori, K., Haji, T., Kojima, H. & Honjo, I. (1988). Voice characteristics of aged persons. *Studia Phonologica* XXII, 39-48.

Morton, R.P. (1996). Quality-of-life: current status and future directions. *Proceedings of the 4th International Conference on Head and Neck Cancer, Toronto*, 119-125.

Murry, T., Bone, R.C. & Essen, C. von (1974). Changes in voice production during radiotherapy for laryngeal cancer. *Journal of Speech and Hearing Disorders* 39, 194-201.

Murty, G.E., Carding, P.N. & Kelly, P.J. (1991). Combined glottographic changes in the elderly. *Clinical Otolaryngology* 16, 532-534.

Muscat, J.E. & Wynder, E.L. (1991). Tobacco, alcohol, asbestos, and occupational risk factors for laryngeal cancer. *Cancer* 69, 9, 2244-2251.

Nieboer, G.L.J., De Graaf, T. & Schutte, H.K. (1988). Esophageal voice quality judgements by means of the semantic differential. *Journal of Phonetics* 16, 417-436.

Olsen, K.D., Perry, W.C., Rohe, D.E. & Keith, R.L. (1995). Quality of life after surgical treatment of cancer of the larynx. *Annals of Otology, Rhinology and Laryngology* 104, 763-769.

Orlikoff, R.F. & Baken, R.J. (1989). The effect of heartbeat on vocal fundamental frequency perturbation. *Journal of Speech and Hearing Research* 32, 576-582.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. (1957). *The measurement of meaning*. University of Illinois Press, Urbana.

Ott, S., Klingholz, F., Willich, N. & Kastenbauer, E. (1992). Die Bestimmung der Qualität der Sprechstimme nach Therapie von T1- und T2-Stimmlippen–Karzinomen. *Laryngologie, Rhinologie, Otologie* 71, 236-241.

Pabon, J.P.H. & Plomp, R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *Journal of Speech and Hearing Research* 31, 710-722.

Pahn, J. & Pahn, E. (1976). Vorschlag zur Einschätzung des Schweresgrades von Stimmstörungen. *Logopedie en Foniatrie* 48, 49-55.

Philips, R.A. (1987). The genetic basis of cancer. In: Tannock, I.F. & Hill, R.P. (eds). *The basic science of oncology*. Pergamon Press, New York.

Pruyn, J.F.M., Jong, P.C. de, Bosman, L.J., Poppel, J.W.M. van, Borne, H.W. van den, Ryckman, R.M. & Mey, K. de (1986). Psychosocial aspects of head and neck cancer - a review of the literature. *Clinical Otolaryngology* 11, 469-474.

Rabinov, C.R., Kreiman, J., Gerrat, B.T. & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research* 38, 26-32.

Ravasz, L.A. & Batterman, J.J. (1989). Radiotherapie bij het hoofd-hals-carcinoom - algemene aspecten. In: Hordijk, G.J. & Ravasz, L.A. (eds). *Het hoofd-halscarcinoom*. Bohn, Scheltema & Holkema, Utrecht.

Rietveld, T. & Van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Mouton de Gruyter, Berlin.

Sato, K. & Hirano, M. (1995). Age-related changes of the macula flava of the human vocal fold. *Annals of Otology, Rhinology and Laryngology* 104, 839-844.

Scherer, R.C., Vail, V.J. & Guo, C.G. (1995). Required number of tokens to determine representative voice perturbation values. *Journal of Speech and Hearing Research* 38, 1260-1269.

Södersten, M. & Lindestad, P.A. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research* 33, 601-611.

Stoicheff, M.L. (1975). Voice following radiotherapy. Part I: Questionnare study; Part II: Speaking fundamental frequency of adult males following radiotherapy for carcinoma of the larynx: a preliminary report. *Laryngoscope* 85, 608-618.

Stoicheff, M.L., Ciampi, A., Passi, J.E. & Fredrickson, J.M. (1983). The irradiated larynx and voice: a perceptual study. *Journal of Speech and Hearing Research* 26, 482-485.

Sulter, A.M., Wit, H.P., Schutte, H.K. & Miller, D.G. (1994). A structured approach to voice range profile (phonetogram) analysis. *Journal of Speech and Hearing Research* 37, 1076-1085.

Takahaski, H. & Koike, Y. (1975). Some perceptual dimensions and acoustic correlates of pathological voices. *Acta Oto-laryngologica, Supplement 338*, 2-24.

Thawley, S.E., Panje, W.R., Batsakis, J.G. & Lindberg, R.D. (1987). *Comprehensive management of head and neck tumors*. W.B. Saunders Company, Philadelphia.

Tielen, M.T.J. (1992). *Male and female speech. An experimental study of sex-related voice and pronunciation characteristics*. Ph.D.-thesis, University of Amsterdam.

Titze, I.R. (1994). Toward standards in acoustic analyses of voice. *Journal of Voice* 8, 1-7.

Titze, I.R., Wong, D., Milder, M.A., Hensley, S.R. & Ramig, L.O. (1995). Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *Journal of Speech and Hearing Research* 38, 526-535.

Tortora, G.J. & Reynolds Grabowski, S (1993). *Principles of anatomy and physiology*. HarperCollins College Publishers, New York.

Union Internationale Contre le Cancer (UICC) (1987). *Manual of clinical oncology*. Springer-Verlag, Berlin.

Verdonck-de Leeuw, I.M., Koopmans-van Beinum, F.J., Hilgers, F.J.M. & Keus, R. (1997). The development of an instrument to analyse voice quality of male patients with early glottic cancer. *Proceedings Larynx97*, 29-32.

Verdonck-de Leeuw, I.M. & Boersma, P. (1996). The effect of radiotherapy measured by means of the harmonics-to-noise ratio. In: Powell, Th.W. (ed.). *Pathologies of speech and language: contributions of clinical phonetics and linguistics*. ICPLA, New Orleans, 203-210.

Verdonck-de Leeuw, I.M. & Koopmans-van Beinum F.J. (1995). The effect of radiotherapy on various acoustical, clinical and perceptual pitch measures. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm* 4, 610-613.

Verdonck-de Leeuw, I.M. & Koopmans-van Beinum F.J. (1995). Voice quality before and after radiotherapy: acoustical, clinical, and perceptual pitch measures. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 19, 1-9.

Vieira, M.N., McInnes, F.R. & Jack, M.A. (1997). Comparative assessment of electroglottographic and acoustic measures of jitter in pathological voices. *Journal of Speech, Language and Hearing Research* 40, 170-182.

Werner-Kukuk, E., Leden, H. von & Yanaghihara, N. (1968). The effects of radiation therapy of laryngeal function. *Journal of Laryngology and Otology* 82, 1-15.

Wijngaarden, H.A. van, Leeuwen, J.P.P.M. van & Hordijk, G.J. (1988). De stem na larynxbestraling. *Logopedie en Foniatrie* 60, 157-161.

Wilkinson, L. (1989). *Systat, the system for statistics*. Systat Inc., Evaston.

Wilson, D.K. (1987). *Voice problems of children*. Williams and Wilkins, Baltimore.

Wolfe, V. & Steinfatt, T.M. (1987). Prediction of vocal severity within and across voice types. *Journal of Speech and Hearing Research* 30, 230-240.

Wolfe, V., Cornell, R. & Palmer, C. (1991). Acoustic correlates of pathologic voice types. *Journal of Speech and Hearing Research* 34, 509-516.

Wolfe, V., Fitch, J. & Cornell, R. (1995). Acoustic prediction of severity in commonly occurring voice problems. *Journal of Speech and Hearing Research* 38, 273-279.

Wynder, E.L., Covey, L.S., Marbuchi, K., Johnson, J. & Muschinsky, M. (1976). Environmental factors in cancer of the larynx, a second look. *Cancer* 38, 1591-1601.

Yumoto, E., Sasaki, Y. & Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *Journal of Speech and Hearing Research* 27, 2-6.

Irma M. Verdonck-de Leeuw

VOICE CHARACTERISTICS FOLLOWING RADIOTHERAPY: THE DEVELOPMENT OF A PROTOCOL

Irma Verdonck-de Leeuw (1963) studied Dutch language and literature at the University of Leyden as well as phonetics at the University of Amsterdam. She graduated in 1988 at the Institute of Phonetic Sciences in Amsterdam. Subsequently, she was trained as a speech and language therapist and exercised her profession in a practice in The Hague from 1990 until 1994. In 1990, she became a Ph.D. student (AIO) at the University of Amsterdam, where she carried out the research project which is reported on in this thesis. In 1996 and 1997, she was employed as a speech and language therapist at the Academic Hospital Leyden. In 1997 she became an editor of 'Logopedie en Foniatrie', the journal of the Dutch Society of Logopedics and Phoniatrics. Currently, she is working as a researcher at the Department of Otorhinolaryngology of the Academic Hospital of the Free University in Amsterdam. Irma de Leeuw is married to Maarten Verdonck; they have two (twin) sons: Juriaan and Casper.

The research described in this thesis is part of a clinical trial that is carried out at the Netherlands Cancer Institute/Antoni van Leeuwenhoek hospital. That trial aims to determine the optimal radiation dose for small glottic tumours. One of the complications of radiotherapy can be a deterioration of voice characteristics. The aim of this research project is to obtain parameters that can describe these voice characteristics. Voice analyses were carried out of a longitudinal group of ten patients before, as well as six months and two years after radiotherapy, and of five separate groups of ten patients each before and six months up to ten years after radiation, and of twenty control speakers. A deterioration of voice characteristics was assessed for 45% of patients six months to seven years after radiotherapy. Stripping the vocal fold for initial diagnosis and smoking after treatment have a negative effect on voice characteristics following radiation. Findings in the present thesis underscore the need of a multidimensional approach in investigating voice characteristics and show that an analysis protocol should comprise voice quality, vocal function, and vocal performance aspects.

printed by
ICG printing - Dordrecht