Preventive rehabilitation in patients treated with chemoradiation for advanced head and neck cancer
van der Molen, L.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Pre- and posttreatment voice and speech outcomes in patients with advanced Head and Neck cancer treated with chemoradiotherapy: expert listeners’ and patient’s perception

Lisette van der Molen
Maya A. van Rossum
Irene Jacobi
Rob J.J.H. van Son
Ludi E. Smeele
Coen R.N. Rasch
Frans J.M. Hilgers

Journal of Voice, in press
Chapter 8

ABSTRACT

Objectives: Perceptual judgments and patients’ perception of voice/speech after concurrent chemo-radiotherapy (CCRT) for advanced head and neck cancer.

Study design: Prospective clinical trial.

Methods: A standard Dutch text and a diadochokinetic task were recorded. Expert listeners rated voice/speech quality (GBRAS-based) and articulation (overall, [p], [t], [k]) and Comparative Mean Opinion Scores (CMOS) of voice/speech was calculated at 3 assessment points. A structured, study-specific questionnare evaluated patients’ perception pretreatment (N=55), at 10-weeks (N=49) and 1-year posttreatment (N=37).

Results: At 10-weeks, perceptual voice quality is significantly affected; overall voice quality (mean -0.24; p=.008), Strain (mean -0.12; p=.012), Nasality (mean -0.08; p=.009), Roughness (mean -0.22; p=.001), and Pitch (mean -0.03; p=.041). These parameters improved over time, but not beyond baseline levels, except for Asthenia at 1-year (voice is less asthenic than at baseline; mean +0.20; p=.007). Perceptual analyses of articulation showed no significant differences. Patients judge their voice quality as good (score 18/20) at all assessment points, but at 1-year the majority (70%) judge their “voice not as it used to be”. In the 1-year versus 10-weeks posttreatment comparison, the larynx-hypopharynx tumor group were more strained, whereas non-larynx tumor voices were judged less strained (means -0.33 and +0.07, respectively; p=.031). Patients’ perceived changes in voice/speech quality at 10-weeks post- versus pretreatment correlate weakly with expert judgments.

Conclusion: Overall, (perceptual) CCRT effects on voice and speech seem to peak at 10-weeks posttreatment, but level off at 1-year. However, at that assessment point the majority of patients still perceive their voice as different from baseline.
INTRODUCTION

Advanced cancer of the head and neck affects structures necessary for swallowing, speech and voice. Concomitant chemoradiotherapy (CCRT) treatment with curative intent targets the tumor with the aim of organ preservation, unlike surgery [1;2]. Yet, anatomic structures and tissues at and around the tumor site can be negatively affected by this treatment resulting in impaired function, and typical CCRT sequels are dry mucosa, muscle atrophy, and fibrosis [3-5]. Whereas quite a large number of studies have investigated swallowing function before and after CCRT [4-6], studies on voice and speech function are few in number [7;8]. At first glance this seems remarkable, because the same anatomic structures are involved in voice and speech as in swallowing. Moreover, voice and speech influence daily communication, which contributes significantly to a person’s identity, personal well-being and overall quality of life [9]. However, voice and speech dysfunctions are less vital/life threatening than, for example, aspiration. This might explain why voice and speech functions have received less attention.

For speech and voice it is helpful to differentiate between tumors above the hyoid bone, i.e. located in the vocal tract (henceforth non-laryngeal group: tumors in the oral cavity, oropharynx, and/or nasopharynx), and below that anatomical structure, i.e. located in or adjacent to the sound source (henceforth laryngeal group: laryngeal and hypopharyngeal tumors). In tumors above the hyoid one might expect that the tumor as well as the treatment will mainly affect speech. In patients with laryngeal tumors, the tumor and treatment will most probably have a primary negative effect on voicing. Besides this ‘site-effect’, the negative side effects of CCRT also depend on which areas of the vocal tract and/or sound source are located within the radiation field. Unfortunately, outcomes in literature are mostly summarized across tumor sites and many studies are not consistent in separating voice and speech terminology and related issues [8].

A recent systematic review established that most studies investigating speech and/or voice outcomes after CCRT mainly relied on Quality of Life (QOL) questionnaires [8]. Only seldom were perceptual assessment and/or objective measurement tools used to evaluate voice or speech. From the limited information available it seems that voice and speech deteriorate during treatment, but improve progressively after treatment, although normal values for voice and speech are not reached, not even in the long-term [8]. Now organ-preservation protocols have become “routine”, the focus is slowly shifting to (preventive) rehabilitation of function [10;11].

With the intention of preventive rehabilitation, a randomized controlled trial (RCT) on advanced head and neck cancer patients was initiated at our institute. The effect of two preventive exercise programs evaluated functional outcomes such as swallowing, mouth opening, pain, nutrition and quality of life at 10-weeks and 1-year post-CCRT compared to baseline measurements;
Chapter 8

The results were published earlier [12;13]. Both exercise programs were similar with respect to the muscle groups trained, incorporating strengthening and stretching exercises to benefit swallowing and mouth opening. No specific voice or speech therapy was provided. Voice and speech data were, however, also collected at all assessment points to assess the effects of CCRT on voice and speech outcomes.

This article reports on the perceptual voice and speech judgments by expert listeners pre-, 10-weeks, and 1-year posttreatment. Additionally, the patients’ perceptions of their voice and speech quality are reported. We also investigated whether the tumor site (primarily larynx or primarily non-larynx) influenced the voice and speech outcomes differently. Finally, the relationship between the perceptual judgments given by expert listeners and the patients’ perception of their voice and speech is reported.

PATIENTS AND METHODS

Patients

Between September 2006 and April 2008, 55 patients treated with concurrent chemoradiotherapy (CCRT) for advanced head and neck cancer participated in a randomized control trial (RCT). The study was approved by the local medical ethical committee, and written informed consent was obtained from all patients before they entered the study. Patients were eligible if they had an inoperable primary tumor of the oral cavity, oropharynx, larynx, hypopharynx, or nasopharynx, stage III or IV [12].

At 10-weeks posttreatment, 49 patients, and at 1-year posttreatment 37 patients were disease free and evaluable. The main reasons for loss to follow-up were death (N=9), progressive disease (N=4), adjuvant local (surgical) treatment (N=3), patient refusal (N=1), and administrative miss (N=1). Ten-weeks posttreatment there were 39 (80%) males and 10 (20%) females with a median age of 57 years, and patients had either stage III (N=16) or stage IV (N=33) tumors. At 1-year, 28 (76%) males and 9 (24%) females with a median age of 58 years were evaluable; 14 patients (38%) with stage III, and 23 (62%) with stage IV disease. Patient data are summarized in Table 1. All patients were treated with CCRT with curative intent. Radiation dose was 70 Gy administered in 35 fractions in 7 weeks with Intensity-Modulated Radiotherapy Therapy (IMRT). Concomitant chemotherapy consisted of 100-mg/m2 Cisplatin as a 40 minutes IV infusion on day 1, 22 and 43 of treatment. Before the start of treatment, all patients were randomized into two comparable preventive swallowing and mouth opening exercise groups (see below). Patients did not receive any specific voice or speech therapy.
Table 1 | Patient characteristics randomized controlled trial (RCT).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Follow-up time</th>
<th>Pretreatment (%)</th>
<th>10-weeks posttreatment (%)</th>
<th>1-year posttreatment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td></td>
<td>55</td>
<td>49</td>
<td>37</td>
</tr>
<tr>
<td>Age in years</td>
<td></td>
<td>57</td>
<td>32.79 years</td>
<td>32.78 years</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>Male</td>
<td>44 (80)</td>
<td>39 (80)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>11 (20)</td>
<td>10 (20)</td>
</tr>
<tr>
<td>T Classification</td>
<td></td>
<td>T1</td>
<td>8 (15)</td>
<td>8 (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2</td>
<td>15 (27)</td>
<td>15 (31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3</td>
<td>21 (38)</td>
<td>19 (39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4</td>
<td>11 (20)</td>
<td>7 (14)</td>
</tr>
<tr>
<td>N Classification</td>
<td></td>
<td>N0</td>
<td>6 (11)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1</td>
<td>15 (27)</td>
<td>14 (29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2</td>
<td>28 (51)</td>
<td>26 (53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N3</td>
<td>6 (11)</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td>III</td>
<td>17 (31)</td>
<td>16 (33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV</td>
<td>38 (69)</td>
<td>33 (67)</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td>Oral cavity/oropharynx/nasopharynx (non-laryngeal group)</td>
<td>36 (65)</td>
<td>31 (60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laryngo/hypopharynx (laryngeal group)</td>
<td>19 (35)</td>
<td>16 (37)</td>
</tr>
<tr>
<td>Exercise program</td>
<td></td>
<td>Standard</td>
<td>28 (51)</td>
<td>25 (51)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experimental</td>
<td>27 (49)</td>
<td>24 (49)</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td>100 mg/m² Cisplatin as a 40 minutes IV infusion on days 1, 22, and 43, and concurrent intensity-modulated radiotherapy (IMRT) of 70 Gy in 35 fractions administered over 7 weeks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 gives an overview of the numbers per tumor site and per exercise group, and which data were not available for voice and speech analyses, because of missing recordings. The patients were equally distributed over exercise groups in terms of tumor site. In the present paper data of both exercise groups was pooled (no comparison between the two exercise groups was made). This decision was based on two important facts, namely, the exercises were directed at swallowing and mouth opening and not specifically at voice and/or speech, and the effects of the two exercise programs in terms of swallowing and mouth opening (as reported in previous papers) were very similar [12;13].

**Preventive exercise programs**

For the interested reader the two exercise programs will be described, even though for the present study these programs are of little consequence, as explained above. Prior to treatment
all patients were randomized into either of two groups: an experimental group that was provided with the TheraBite Jaw Motion Rehabilitation System device, Atos Medical AB, Hörby, Sweden and a group receiving standard intervention (Standard group). In short, both regimes consist of comparable stretch and strength exercises to keep the swallowing musculature active. Patients were encouraged to practice 3 times daily before, during, and up to 1-year post-CCRT (see for a detailed description of the exercises the study of Van der Molen et al. [12]).

Table 2 | Overview of the voice data available at the different follow-up times given separately for the listening experiments and the subjective questionnaire. Subdivided into two tumor location groups and exercise group.

<table>
<thead>
<tr>
<th>Dutch text &amp; patient questionnaire</th>
<th>Pre (N=55)</th>
<th>P1 (N=46)</th>
<th>P3 (N=37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-laryngeal-group</td>
<td>S group</td>
<td>E group</td>
<td>S group</td>
</tr>
<tr>
<td>Laryngeal-group</td>
<td>19</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>(pataka)</td>
<td>E group</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Pre (N=55)</td>
<td>S group</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>E group</td>
<td>10</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>P1 (N=46)</td>
<td>S group</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>E group</td>
<td>8</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>P3 (N=32)</td>
<td>S group</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>E group</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Non-laryngeal-group= oral cavity, oropharynx and nasopharynx tumors; Laryngeal-group= hypopharynx and larynx tumors, Pre= baseline data; P1= 10-weeks posttreatment; P3= 1-year posttreatment, S group= Standard exercise group; E group= Experimental exercise group

Data collection
Patient and voice and speech data were collected prior to treatment (baseline), at 10-weeks after the last radiotherapy, and at 1-year posttreatment by means of speech recordings and a structured trial specific questionnaire. The structured patient questionnaire focused on several functional outcomes, and the questions related to voice and speech were used in the present analyses. The speech recordings were used in two auditory perception experiments.

Speech material and recording procedure
The material consisted of a standard Dutch text with a neutral content, which was used to rate voice and speech quality, and a diadochokinetic task (repeating [pataka] as fast as possible), which was used to rate the sharpness of articulation (especially lip, tip of tongue and tongue base movement). The patients were asked to read all texts aloud at a comfortable loudness and pitch level. All speech material was recorded in a sound-treated room, using a Sennheiser MD421 Dynamic Microphone and an Edirol (Roland) R-1 portable 24bit digital wave recorder.
Perceptual evaluations

Stimulus material

For each speaker, per assessment point (pre-, 10-weeks, and 1-year post), 10 seconds stretches of the read-aloud text (see Appendix E), and minimally 4 repetitions of the diadochokinetic utterance [pataka] were manually selected, excised and equalized with Praat (www.fon.hum.uva.nl/praat/). These speech samples were used in the two listening experiments; the 10 second stretches of read-aloud text in a listening experiment on quality of voice and speech, the diadochokinetic utterance in a second listening experiment on sharpness of articulation. Four practice items were included at the start of both listening experiments, and were not used for analyses.

Procedure of listening experiment 1: rating quality of voice and speech

Six experienced speech pathologists were asked to rate the voice and speech, independently. For the purpose of this study, a semantic rating scale was developed, based on the Grade, Roughness, Breathiness, Asthenia and Strained (GBRAS) scale, but extended with a number of semantic scales deemed appropriate for this study population [14]. The scales included were overall Grade of voice quality, Roughness, Breathiness, Asthenia, and Strain. Extra semantic scales were Pitch, Nasality, overall Intelligibility, and Dry mouth. Pairs of recorded speech samples were presented over Sennheiser HD418 headphones during an online experiment. The samples that constituted a pair were always from the same speaker. The comparisons investigated the following speech samples: 10-weeks posttreatment versus pretreatment, 1-year posttreatment versus pretreatment and 1-year posttreatment versus 10-weeks posttreatment, hence, a pair could be either one of these three combinations (e.g. one sample = pretreatment, the other = 10-weeks posttreatment, or a pair with one sample = 1-year posttreatment, the other 10-weeks posttreatment). The text spoken in the paired samples was identical, so that listeners would not be distracted by differences in the spoken texts. The speech pathologists could listen to and compare the speech samples as often as they wished. For each scale, the speech pathologists were asked how sample B compared to sample A (for instance, was B perceived as much more breathy than A), and mark their judgment on an Equal Appearing Interval Scale (EAI) (Likert, 1932) running from minus 3 (sample B was perceived as poorer) through 0 (no difference between the samples) to plus 3 (sample B was perceived as better). An exception was the Pitch scale; minus meant that the Pitch in sample B was perceived as lower, plus that sample B was perceived as higher than sample A. The listeners were unaware of which measurement points were compared (sample A could be any of the three recordings (pretreatment, 10-weeks or 1-year posttreatment), and the same was true for sample B).

Procedure of listening experiment 2: rating sharpness of articulation

The diadochokinetic utterance in a second listening experiment on sharpness of articulation. Four practice items were included at the start of both listening experiments, and were not used for analyses.

Procedure of listening experiment 3: rating quality of voice and speech

Six experienced speech pathologists were asked to rate the voice and speech, independently. For the purpose of this study, a semantic rating scale was developed, based on the Grade, Roughness, Breathiness, Asthenia and Strained (GBRAS) scale, but extended with a number of semantic scales deemed appropriate for this study population [14]. The scales included were overall Grade of voice quality, Roughness, Breathiness, Asthenia, and Strain. Extra semantic scales were Pitch, Nasality, overall Intelligibility, and Dry mouth. Pairs of recorded speech samples were presented over Sennheiser HD418 headphones during an online experiment. The samples that constituted a pair were always from the same speaker. The comparisons investigated the following speech samples: 10-weeks posttreatment versus pretreatment, 1-year posttreatment versus pretreatment and 1-year posttreatment versus 10-weeks posttreatment, hence, a pair could be either one of these three combinations (e.g. one sample = pretreatment, the other = 10-weeks posttreatment, or a pair with one sample = 1-year posttreatment, the other 10-weeks posttreatment). The text spoken in the paired samples was identical, so that listeners would not be distracted by differences in the spoken texts. The speech pathologists could listen to and compare the speech samples as often as they wished. For each scale, the speech pathologists were asked how sample B compared to sample A (for instance, was B perceived as much more breathy than A), and mark their judgment on an Equal Appearing Interval Scale (EAI) (Likert, 1932) running from minus 3 (sample B was perceived as poorer) through 0 (no difference between the samples) to plus 3 (sample B was perceived as better). An exception was the Pitch scale; minus meant that the Pitch in sample B was perceived as lower, plus that sample B was perceived as higher than sample A. The listeners were unaware of which measurement points were compared (sample A could be any of the three recordings (pretreatment, 10-weeks or 1-year posttreatment), and the same was true for sample B).
Chapter 8

The speakers and order of measurements were randomized. Listeners needed approximately 3 hours to complete the online experiment (http://www.fon.hum.uva.nl/rob/Lisette/index).

Procedure of listening experiment 2: rating of articulation

The second listening experiment was carried out to evaluate possible changes in articulation over time. Seven experienced speech pathologists/phoneticians were asked to rate the articulation independently during an online experiment. For this purpose, four articulation scales were included: overall articulation, articulation of the [p], articulation of the [t], and articulation of the [k]; based on the recorded [pataka] speech samples. The procedure was exactly the same as in listening experiment 1. There were again paired speech samples (sample A and sample B) and the samples that constituted a pair were again always from the same speaker. The listeners were unaware of which recording was A and which was B (baseline, 10-weeks, or 1-year posttreatment records). For each scale, the listeners were asked to mark on the EAI whether sample B was poorer, the same or better than sample A. The listeners could listen to and compare the speech samples as often as they wished. The speakers and order of measurements were randomized. Listeners needed approximately 1.5 hours to complete the online experiment (http://www.fon.hum.uva.nl/rob/Lisette2/index).

Quality of life (QOL) questionnaire

To investigate how patients perceive their own voice and speech 10-weeks and 1-year after CCRT, a structured study-specific QOL questionnaire was used. This questionnaire's validity was substantiated during earlier studies and includes detailed and symptom-specific questions for patients treated for head and neck cancer [15]. There were six questions related to speech and/or voice outcomes (see Appendix D, questions B. section d). Five questions concerned Intelligibility (face to face and over the telephone), loudness, Pitch, and speech rate. Patients had to rate how they perceive these aspects of their speech and/or voice on a scale from 1 to 4 (1= poor, 2= moderate, 3= reasonable and 4= good), and one question asked whether patients perceive their voice "as it used to be before their illness" (yes/no).

Patients filled out the QOL questionnaire in the presence of the researcher (LvdM). It was explained that the patients should answer how they perceive their own speech and/or voice, and not how others had said they experienced the patient's speech and/or voice. Every question and all possible answers were presented verbally to the patients by the researcher, and they had to choose the answer that responded the closest to how they experienced their speech/voice. If a question was unclear, or if patients misunderstood the question, it was repeated or additional explanation was given by the researcher.
To allow proper comparison with the perceptual voice and speech judgments of the expert listeners, difference values were calculated for the patients’ perceptions as well (for instance a patient’s answer to a question pretreatment was subtracted from the answer given posttreatment).

**Statistical analysis**

All statistical analyses were conducted in SPSS v 16.01.

**Listening experiments**

The perceptual evaluation was quantified using the Comparative Mean Opinion Score (CMOS) (ITU-P.800, 1996), a standard technique also employed by the telecommunications industry to compare auditory quality speech signals. A one-sample t-test was used to determine whether the mean value differs significantly from zero, where a positive score means sample B was judged to be better than sample A, and a negative score means sample B was judged to be worse than sample A (except the “Pitch” scale in listening experiment 1, as explained above).

**Patient’s judgments: questionnaire**

The 6 speech-related items (intelligibility, volume, pitch, speech rate, intelligibility on the phone, and voice change over time) of the trial specific structured questionnaire were combined into a more limited set of multiple-item scales according to Likert’s method of summated ratings. A Cronbach’s Alpha was calculated. A value >.700 indicates good reliability. Statistical analyses primarily included descriptive analyses and tabulations. Difference values (see explanation above) were compared using Mann-Whitney U. A p-value of <.05 was taken to indicate statistical significance.

**Difference between tumor groups (laryngeal versus non-laryngeal)**

For the perceptual judgments, the non-parametric Mann-Whitney U was used to compare whether differences exist between the non-laryngeal and the laryngeal group.

**Correlations between patients’ judgments and expert listener ratings**

Statistical associations were calculated by Pearson’s correlation coefficient.
Chapter 8

RESULTS

The results of the two listening experiments and the structured study-specific questionnaire will be presented separately. First, the overall voice and speech judgments will be given and in addition the differences between tumor sites will be specified. Finally, the results of the listening experiments and the structured study-specific questionnaire will be compared.

Listening experiment 1: rating quality of voice and speech

Table 3 gives an overview of the Comparative Mean Opinion Score (CMOS) of the perceptual analyses of voice and speech quality. Compared to pretreatment, approximately half of the voice parameters were judged significantly worse at 10-weeks post-CCRT: Overall voice quality (mean -0.24; p=.008), Strain (mean -0.12; p=.012), Nasality (mean -0.08; p=.009), and Roughness (mean -0.22; p=.001). Moreover, Pitch (mean -0.03; p=.041) was perceived as significantly lower at 10-weeks posttreatment than pretreatment.

At 1-year only Asthenia (mean +0.20; p=.007) was less (indicating that the voice was perceived as more powerful) compared to the pretreatment measurements. None of the other scales showed statistically significant differences.

Comparing the 1-year posttreatment voice and speech results to 10-weeks posttreatment, significant improvements could be seen on the scales Overall voice quality (mean +0.19; p=.038), Roughness (mean +0.15; p=.033), and Breathiness (mean +0.16; p=.013). In Fig. 1 the bar graphs of these means are given.

<table>
<thead>
<tr>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall voice</td>
<td>-0.24</td>
<td>1.485</td>
<td>.008*</td>
<td>+0.02</td>
<td>1.452</td>
<td>.019</td>
<td>+0.19</td>
<td>1.248</td>
</tr>
<tr>
<td>Roughness</td>
<td>-0.15</td>
<td>1.056</td>
<td>.000**</td>
<td>-0.14</td>
<td>1.004</td>
<td>.015</td>
<td>+0.15</td>
<td>0.803</td>
</tr>
<tr>
<td>Breathiness</td>
<td>-0.05</td>
<td>0.572</td>
<td>0.454</td>
<td>+0.01</td>
<td>0.414</td>
<td>.016</td>
<td>+0.16</td>
<td>0.430</td>
</tr>
<tr>
<td>Asthenia</td>
<td>+0.02</td>
<td>0.799</td>
<td>0.773</td>
<td>+0.20</td>
<td>0.768</td>
<td>.007**</td>
<td>+0.10</td>
<td>0.684</td>
</tr>
<tr>
<td>Strain</td>
<td>+0.12</td>
<td>1.136</td>
<td>0.032**</td>
<td>+0.06</td>
<td>1.137</td>
<td>0.034</td>
<td>+0.02</td>
<td>0.860</td>
</tr>
<tr>
<td>Pitch</td>
<td>+0.05</td>
<td>0.311</td>
<td>0.041*</td>
<td>+0.06</td>
<td>0.326</td>
<td>0.008</td>
<td>+0.09</td>
<td>0.543</td>
</tr>
<tr>
<td>Nasality</td>
<td>+0.06</td>
<td>1.055</td>
<td>0.008**</td>
<td>-0.03</td>
<td>1.103</td>
<td>0.001</td>
<td>+0.01</td>
<td>0.943</td>
</tr>
<tr>
<td>Intelligibility</td>
<td>+0.03</td>
<td>0.951</td>
<td>0.038**</td>
<td>+0.03</td>
<td>0.866</td>
<td>0.116</td>
<td>+0.05</td>
<td>0.775</td>
</tr>
<tr>
<td>Dry mouth</td>
<td>-0.04</td>
<td>0.730</td>
<td>0.287</td>
<td>-0.04</td>
<td>0.688</td>
<td>0.224</td>
<td>+0.03</td>
<td>0.583</td>
</tr>
</tbody>
</table>

CMOS = Comparative Mean Opinion Score, SD = Standard Deviation, Pre = baseline speech samples, P1 = 10-weeks posttreatment speech, samples, P3 = 1-year after treatment speech samples, * = significant <.05

Listening experiment 2: rating of articulation

Table 4 gives the Comparative Mean Opinion Score (CMOS) of the perceptual analyses of articulation. None of these results were significant (Fig. 2).
Structured study-specific questionnaire

For the patients’ experience regarding the voice characteristics summed Likert scale were used. The Cronbach’s Alpha scores of the 5-items scale was good (> .700), Table 5. In this respect, a higher score on the structured questionnaire presents a better voice. The analysis showed that the voice characteristics are rated good (overall means scores around 18, with a highest score of 20) at all measurement points in time. At 1-year posttreatment 100% of the patients rated their voice reasonable (5/32; 16%) to good (27/32; 84%). Mann-Whitney U analyses revealed no significant differences over time.

Table 4 | Results of second listening experiment: Overview of the Comparative Mean Opinion Score (CMOS): articulation.

<table>
<thead>
<tr>
<th></th>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
<th>CMOS</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall articulation</td>
<td>-0.03</td>
<td>1.025</td>
<td>.139</td>
<td>-0.10</td>
<td>1.063</td>
<td>.195</td>
<td>-0.11</td>
<td>1.066</td>
<td>.108</td>
</tr>
<tr>
<td>Articulation (a)</td>
<td>-0.05</td>
<td>0.916</td>
<td>.299</td>
<td>-0.10</td>
<td>0.864</td>
<td>.087</td>
<td>-0.10</td>
<td>1.006</td>
<td>.102</td>
</tr>
<tr>
<td>Articulation (b)</td>
<td>-0.06</td>
<td>0.791</td>
<td>.341</td>
<td>-0.03</td>
<td>0.762</td>
<td>.520</td>
<td>-0.05</td>
<td>0.726</td>
<td>.500</td>
</tr>
<tr>
<td>Articulation (c)</td>
<td>-0.03</td>
<td>0.810</td>
<td>.450</td>
<td>-0.03</td>
<td>0.871</td>
<td>.542</td>
<td>-0.10</td>
<td>0.863</td>
<td>.066</td>
</tr>
</tbody>
</table>

CMOS = Comparative Mean Opinion Score, SD = Standard Deviation, Pre = baseline speech samples, P1 = 10-weeks posttreatment speech samples, P3 = 1-year after treatment speech samples.

Figure 1 | Bar graphs of the Comparative Mean Opinion Score (CMOS): overall voice quality, Roughness, Breathiness, Asthenia, Strain, Pitch, Nasality, Intelligibility, and Dry mouth at the different measurement pairs (pre = pretreatment; p1 = 10-weeks posttreatment; p3 = 1-year posttreatment). * = significant difference
In answer to the question: “Is your voice as it used to be?” approximately half of the patients answered ‘no’ at baseline 46 % (21/46) and at 10-weeks 50 % (23/46) after treatment. At 1-year posttreatment this number remarkably increased to 70% of the patients (26/37), with 14 of those patients (38%), who had not experienced any voice changes at 10-weeks posttreatment.

Results per tumor site
For the voice and speech quality judgments – listening experiment 1 – there was a significant difference between the two tumor groups when comparing 1-year posttreatment to 10-weeks posttreatment. The larynx group was considered more strained, whereas the non-larynx group was considered less strained (means -0.33 and +0.07, respectively; p=.031). There were no differences between the tumor groups for the articulation ratings, or with regard to the patients’ judgments.

Correlation perceptual and patients’ judgments
Comparing 10-weeks to baseline, patients perceived a worse voice and speech quality, and this correlated with the following expert listeners judgments: Overall voice quality (r 0.547; p=.000), Roughness (r 0.466; p=.001), Breathiness (r 0.473; p=.001), Asthenia (r 0.476; p=.001), Strain (r
0.411; p=.005), Pitch (r 0.326; p=.029), Intelligibility (r 0.354; p=.018), and Dry mouth (r 0.482; p=.001); overall articulation (r 0.447; p=.002), articulation of the [t] (r 0.361; p=.014), articulation of the [k] (r 0.426; p=.003), and articulation of the [p] (r 0.325; p=.027). No correlations were found for 1-year posttreatment versus pretreatment or versus 10 week posttreatment.

Table 5 | Likert scales and scores regarding the voice characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Pretreatment (N=46)</th>
<th>10-weeks posttreatment (N=46)</th>
<th>1-year posttreatment (N=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s Alpha</td>
<td>.872</td>
<td>.734</td>
<td>.818</td>
</tr>
<tr>
<td>Voice characteristics (5 questions with 4 points each): Mean Range</td>
<td>Score 17.8 7-20</td>
<td>Score 18.0 10-20</td>
<td>Score 18.2 12-20</td>
</tr>
<tr>
<td>Is your voice as it used to be?</td>
<td>Yes 25 (54%) 21 (46%)</td>
<td>Yes 23 (50%) 23 (50%)</td>
<td>Yes 22 (69%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0-5 = poor; 6-10 = moderate; 11-15 = reasonable; 16-20 = good

DISCUSSION

This is one of few clinical studies that prospectively assessed the effects of organ preserving concurrent chemo-radiotherapy (CCRT) on voice and speech, using perceptual analyses and a structured study specific questionnaire. The study shows that in patients with advanced head and neck cancer treated with CCRT expert listeners judge the overall voice quality, as well as Roughness, Strain and Nasality significantly worse at 10-weeks posttreatment compared to pretreatment measurements. Moreover, Pitch is judged to be lower than pretreatment. At 1-year posttreatment overall voice and speech quality, Roughness and Breathiness have improved significantly compared to 10-weeks posttreatment. One year posttreatment most perceptual voice parameters again approach baseline levels, except that the voice was perceived as less asthenic (more powerful) than pretreatment. In contrast, the outcome of the question “Is your voice as it used to be?” revealed that many patients judged their voice to have deteriorated. Pretreatment and 10-weeks posttreatment roughly half of the patients judge their voice as different from before their illness, whereas this percentage increased to 70 at 1-year posttreatment.

It is difficult to compare these results with the literature, because very few studies used clinician-based voice quality ratings in CCRT treated patients and to our knowledge there is no study that also used the Comparative Mean Opinion Score (CMOS) to directly compare assessment points [7,8]. A close match is the study by Carrara-de Angelis et al. [16], in which
3 trained listeners prospectively analyzed voice quality of 43 hypopharyngeal and laryngeal cancer patients treated with CCRT using the GBRAS scale. Normal voice was perceived in 1 (7%) of the 15 patients, mild dysphonia in 4 (27%), moderate in 6 (40%), and severe in 4 (27%) at 2 to 9 months after CCRT. The vocal parameters of Roughness or Breathiness were the most frequent abnormalities (87% and 78%, respectively), 3 patients (20%) presented with strain and/or instability in vocal quality. This is comparable with the results found in the present study. In the present study, no significant results for the articulation ratings were found. However, it must be noted that these ratings are only based on judgments of [p], [t], and [k] of the diadochokinetic utterance [pataka], and in addition, this could also be a result of comparing the effects of the tumor with the effects of the treatment. At baseline the tumor could already have influenced the voice and speech quality, and after treatment the negative side-effects of chemoradiation could have influenced the voice and speech quality. So, if no differences were found, it does not imply that there are no voice and speech problems.

We used study-specific questions, which have proven their reliability in our earlier studies [17-19], to get a more precise idea of how patients perceive their own voice and speech. In the present study, patients did not notice any significant changes in their voice or speech. In the literature voice and speech outcomes are mostly assessed by standardized Quality of Life questionnaires [7,8]. Psyrri et al., for example, collected posttreatment quality of life data by using the PSS H&N, and reported persistent Roughness in 2 out of 18 patients, whereas 3 patients reported that their voice quality had improved [20]. Since the PSS H&N has only one question on speech and no item on voice, the origin of the voice outcome remains unclear. Generally, treatment effects seem to have lessened in the long-term or are not perceived very clearly. However, as already mentioned before, the patients in the present study changed their opinion, suggesting that the voice at 1-year posttreatment was different to the voice at 10-weeks. Patients probably judge their swallowing problems as more important, until these have stabilized and they can focus on other changes. Hence, it seems that questions eliciting a comparative judgment over time yield more information than a moment specific judgment.

In the present study, patients' perceived changes in voice and speech 10-weeks posttreatment versus pretreatment correlated weakly with the expert judgments. Woodson et al. analyzed the voices of 15 head and neck cancer patients with an overall voice quality scale and patient interviews [21]. Those authors reported that both perceptual analysis and communication score were concordant with patient self-rating of voice and communication ability. Since, similar but weak relations were found in a previous study [22] this might indicate that the two measures are complementary.

As mentioned in the introduction, one might expect more speech problems in the non-laryngeal group and more voice problems in the laryngeal group. However, there was only one difference
between the two tumor groups (non-laryngeal or laryngeal): expert listeners considered the voices in the larynx group to be more strained at 1-year than at 10-weeks posttreatment. An explanation could be that the radiation fields of all advanced head and neck tumors regardless of their origin cover a large part of the pharyngeal-laryngeal tract. As noted by Eisbruch et al. patients with a tumor in the larynx will also receive a high radiation dose to the oropharynx area [23]. Patients with an advanced primary tumor in the oral cavity, on the other hand, will often receive a high radiation dose to pathological nodes in the neck area. Also, the oral cavity is a rather large area and radiation doses to different parts can vary dramatically between patients, so that type and extent of articulatory problems might differ between patients. The relationship between radiation dose and radiation sites should be taken much more into account when evaluating functionality and planning (preventive) rehabilitation. We anticipate that one of our next studies will focus on this relationship. However, more patients might be needed to better differentiate tumor site-related voice and articulation problems.

In comparison to other studies that only rely on one or two general voice questions, the present study included both the patients’ and the experts’ judgments. Optimally, one would prefer a multidimensional assessment including acoustic, stroboscopic and articulation testing, so that a more accurate picture emerges. This is also what Orlikoff et al. [24] emphasized in their discussion, after they found a discrepancy between acoustic, clinician-based, and patient-based data. Woodson et al. [21] found a significant correlation between the acoustic measures (that were abnormal in patients), and jitter scores with listener ratings of perceived voice quality. For the present study population, separately acoustic analyses are being carried out. Preliminary acoustic results, based on sustained /a/ recordings not used in the present study, show that in the laryngeal group voice quality measures are significantly more affected, whereas in the non-laryngeal group both articulation and Nasality are more impaired [25]. Apparently, these effects were too subtle for the listeners in the present study.

Since all patients received similar strength and stretch exercises (not specifically aimed at improving voice and speech), we could not assess the effects of exercises on voice and speech outcome. Comparing the results of the present study with those from a preceding trial in our Institute (with the same study design and CCRT treatment schedule, but without any preventive exercises), lower Likert scale values than in the present study were found (83% versus 100%, respectively, reporting a reasonable to good voice at 1-year [18]). However, since this comparison is a historical one, this somewhat better outcome in favor of the exercise program has to be interpreted with care. On the other hand, the phenomenon of neuroplasticity, as e.g. found for Parkinson’s disease, which implies that the rehabilitation of swallowing is a neuroplastic event, might be an explanation for the surprisingly limited residual voice and speech problems we found at 1-year. Sapir et al., e.g., found that specific voice training also may improve swallowing.
in Parkinson’s and other patients with neurological disorders [26]. Our data would point to the reverse phenomenon, i.e. training of swallowing function may also improve voice and speech. This would coincide with current data on neural plasticity, as discussed by Robbins et al. [27], showing that there is ample evidence to suggest that treatment of a set of muscles for the purpose of improving swallowing would bear some effect on voice and speech, as well. In addition (preventive) rehabilitation specifically addressing voice and speech also could be valuable. A study by Van Gogh et al. [28], e.g., showed that patients did perceive some benefit from voice therapy after radiotherapy.

In conclusion, overall, (perceptual) CCRT side effects on voice and speech seem to peak at 10-weeks posttreatment, but level off at 1-year. However, at that assessment point the majority of patients still perceive their voice as different from baseline.

ACKNOWLEDGEMENT

The authors would like to express gratitude to Corina van As-Brooks, Brigitte Boon, Marjolein Freier, Marieke Hakkesteegt, Eeke Hallo, Ineke Hartgring, Vivienne van de Kamp, Anne Kornman, Merel Latenstein, and Marika Voerman, SLP’s, and Irene Jacobi, and Petra Jongmans, phoneticians, for their willingness to carry out the various (time consuming) perception experiments.
Chapter 8
Reference List

Chapter 8


