



## UvA-DARE (Digital Academic Repository)

### Assessment of voice, speech, and related quality of life in advanced head and neck cancer patients 10-years+ after chemoradiotherapy

Kraaijenga, S.A.C.; Oskam, I.M.; van Son, R.J.J.H.; Hamming-Vrieze, O.; Hilgers, F.J.M.; van den Brekel, M.W.M.; van der Molen, L.

**DOI**

[10.1016/j.oraloncology.2016.02.001](https://doi.org/10.1016/j.oraloncology.2016.02.001)

**Publication date**

2016

**Document Version**

Final published version

**Published in**

Oral Oncology

**License**

Article 25fa Dutch Copyright Act

[Link to publication](#)

**Citation for published version (APA):**

Kraaijenga, S. A. C., Oskam, I. M., van Son, R. J. J. H., Hamming-Vrieze, O., Hilgers, F. J. M., van den Brekel, M. W. M., & van der Molen, L. (2016). Assessment of voice, speech, and related quality of life in advanced head and neck cancer patients 10-years+ after chemoradiotherapy. *Oral Oncology*, *55*, 24-30.  
<https://doi.org/10.1016/j.oraloncology.2016.02.001>

**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: <https://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.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



## Assessment of voice, speech, and related quality of life in advanced head and neck cancer patients 10-years+ after chemoradiotherapy



S.A.C. Kraaijenga<sup>a</sup>, I.M. Oskam<sup>a</sup>, R.J.J.H. van Son<sup>a,c</sup>, O. Hamming-Vrieze<sup>b</sup>, F.J.M. Hilgers<sup>a,c,\*</sup>, M.W.M. van den Brekel<sup>a,c,d</sup>, L. van der Molen<sup>a</sup>

<sup>a</sup>The Netherlands Cancer Institute, Department of Head and Neck Oncology and Surgery, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

<sup>b</sup>The Netherlands Cancer Institute, Department of Radiation Oncology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

<sup>c</sup>Institute of Phonetic Sciences, University of Amsterdam, Spuistraat 210, 1012 VT Amsterdam, The Netherlands

<sup>d</sup>Academic Medical Center, Department of Oral and Maxillofacial Surgery, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 3 December 2015

Received in revised form 26 January 2016

Accepted 1 February 2016

Available online 10 February 2016

#### Keywords:

Head and neck cancer

Chemoradiotherapy

Voice quality

Speech

Intelligibility

GRBAS

Perceptual evaluation

Automatic speech recognition

Long-term effects

IMRT

### SUMMARY

**Objectives:** Assessment of long-term objective and subjective voice, speech, articulation, and quality of life in patients with head and neck cancer (HNC) treated with concurrent chemoradiotherapy (CRT) for advanced, stage IV disease.

**Materials and methods:** Twenty-two disease-free survivors, treated with cisplatin-based CRT for inoperable HNC (1999–2004), were evaluated at 10-years post-treatment. A standard Dutch text was recorded. Perceptual analysis of voice, speech, and articulation was conducted by two expert listeners (SLPs). Also an experimental expert system based on automatic speech recognition was used. Patients' perception of voice and speech and related quality of life was assessed with the Voice Handicap Index (VHI) and Speech Handicap Index (SHI) questionnaires.

**Results:** At a median follow-up of 11-years, perceptual evaluation showed abnormal scores in up to 64% of cases, depending on the outcome parameter analyzed. Automatic assessment of voice and speech parameters correlated moderate to strong with perceptual outcome scores. Patient-reported problems with voice (VHI > 15) and speech (SHI > 6) in daily life were present in 68% and 77% of patients, respectively. Patients treated with IMRT showed significantly less impairment compared to those treated with conventional radiotherapy.

**Conclusion:** More than 10-years after organ-preservation treatment, voice and speech problems are common in this patient cohort, as assessed with perceptual evaluation, automatic speech recognition, and with validated structured questionnaires. There were fewer complaints in patients treated with IMRT than with conventional radiotherapy.

© 2016 Elsevier Ltd. All rights reserved.

### Introduction

In patients with advanced head and neck cancer (HNC), both the tumor and its treatment with combined chemoradiotherapy (CRT) can adversely impact voice and speech outcomes. In patients with cancers of the oral cavity and oropharynx, destructive effects of the tumor will mainly affect patients' articulation and/or speech, whereas in laryngeal cancer patients, the tumor often has negative effects on voice quality [1,2]. Treatment effects of (chemo-) radiotherapy on voice quality and speech predominantly depend on radiation doses to the organs at risk surrounding the primary

\* Corresponding author at: Dept. Head and Neck Surgery and Oncology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel.: +31 205122550.

E-mail address: [f.hilgers@nki.nl](mailto:f.hilgers@nki.nl) (F.J.M. Hilgers).

tumor and lymph nodes. When the larynx is included in the radiation field, decreased voice quality may be attributed to impaired vocal fold vibration, incomplete glottic closure, insufficient lubrication/dryness of the laryngeal mucosa, muscle atrophy, fibrosis, hyperaemia, and/or erythema [3]. Patients often complain about increased vocal effort, breathiness, and hoarseness [2]. Radiation treatment for non-laryngeal cancer may also influence voice and speech, even at long-term [4], due to radiation-induced anatomical changes of the vocal tract, e.g. scarring, edema and/or fibrosis of structures in/around the oral cavity or oropharynx [5,6]. Consequently, reduced speech intelligibility and impaired articulation may affect patients' daily life activities and interactions, which can be associated with severe functional and psychosocial problems, and reduced quality of life [7,8].

Previous literature on voice quality and speech following CRT for advanced HNC has proposed the use of prospective, standardized multidimensional voice and speech assessment protocols, based on adequate scientific background with long-term follow-up [1,7,9]. In 2009, Dwivedi and colleagues studied speech outcomes following oral cavity and/or oropharyngeal cancer, and recommended speech evaluation by various modalities, i.e. perceptual evaluation, acoustic evaluation, and structured questionnaires [9]. Also Jacobi et al. and Schuster and Stelzle clarified in their reviews in this area the need for structured, standardized protocols, including baseline assessments and long-term follow-up [1,7].

Despite these recommendations, prospectively collected voice and speech data still are scarce [4,10,11], especially at long-term [2]. At the same time, technology is improving, and automated methods of voice and speech evaluation are under development as an alternative and/or adjunct to traditional, time-consuming perceptual evaluation of voice quality and speech [7,12,13]. In particular in research setting, automatic speech recognition is already used, to provide global measures of speech intelligibility and (to a lesser extent) of voice quality [14,15]. However, also in clinical settings automatic speech evaluation can be used to ensure multidimensional assessments, which can be time efficient and fast. The aim of the current study was to report on the long-term objective and subjective voice and speech outcomes, including perceptual evaluation, automatic evaluation, and patient-reported outcomes.

## Material and methods

### *Patient and treatment characteristics*

As part of a randomized controlled clinical trial between 1999 and 2004 at the Netherlands Cancer Institute [16], twenty-two HNC survivors treated with concurrent cisplatin-based radiotherapy were disease-free, evaluable, and willing to participate at long-term (10-years+) post-treatment evaluation. For patients' and treatment characteristics and reasons for exclusion at the long-term assessment point we refer to the recently published paper on dysphagia in the same patient cohort [17]. In summary, the original patient cohort consisted of patients diagnosed with stage IV cancer of the oral cavity, oropharynx, or hypopharynx. Patients were treated with cisplatin as either a standard 100 mg/m<sup>2</sup> intravenous (IV) 40 min infusion on days 1, 22, and 43, or a high-dose, targeted and rapid 150 mg/m<sup>2</sup> intra-arterial (IA) cisplatin injection with intravenous sodium thiosulphate rescue in weeks 1, 2, 3, and 4. The primary tumor area and neck nodes were irradiated with 2 Gy per fraction, in 35 fractions over 7 weeks, starting concurrently with chemotherapy. Ten patients (45%) were treated with intensity-modulated radiotherapy (IMRT), and 12 patients (55%) with conventional radiotherapy. Based on perceptual categorization, three patients were categorized as audibly non-native speakers, whereas the other nineteen were categorized as native (with/without audible regional or dialect variants).

### *Data collection*

Voice, speech, and articulation outcomes were collected at 10-years+ post-treatment from speech recordings consisting of a 189-word Dutch fairy tale with neutral content containing almost all Dutch phonemes (similar to earlier studies in our Institute [10,12]; Appendix A). Patients were asked to read the text aloud at a comfortable loudness and pitch level. All recordings were made in a sound-treated room using a Sennheiser MD421 Dynamic Microphone and an Edirol (Roland) R-1 portable 16-bit (44.1 kHz)

digital wave recorder. The mouth-to-microphone distance was kept constant at approximately 30 cm.

### *Perceptual evaluation*

The stimuli for the listening experiment consisted of two fragments, the first 70 words (A) and the following 68 words (B), from the original 189-word passage read by the patients [12,13]. Thus, each patient was rated twice by each SLP, once on fragment A and once on fragment B. Stimulus material was manually selected by an independent expert, excised, and equalized at 70 decibel with the PRAAT program [18]. Four practice items, a list of words, and sustained/a/vowels were also recorded but not used for the current analysis. During the listening experiment, all recordings were presented over a Sennheiser HD418 headphone.

### *Perceptual rating*

Two experienced speech language pathologists (SLPs), both Dutch native speakers, were asked as expert listeners to rate voice, speech, and articulation parameters independently. The listeners were blinded to patient information. Recordings were presented for evaluation using the Open Source program TEVA [19], which runs as a PRAAT extension [10,15,20]. Semantic scales were used to rate voice quality on computerized Visual Analog Scales (VAS). Included scales were overall grade of voice quality, roughness, breathiness, asthenia, and strain (GRBAS) [21]. Also a number of additional semantic scales were included to rate overall speech intelligibility, the precision of articulation, nasality, and prosody. The GRBAS scale was not used in its standardized form (rating on 0–3), but the descriptors of the GRBAS scale were used to computerize and digitize VAS ratings to scores ranging from 0 ('least similar to normal') to 1000 ('most similar to normal'). The listeners discussed and adjusted scale definitions during the evaluation of 10 practice sessions, with the same recorded text available from a different patient population [10]. The final/experiment recordings were presented in identical order to both listeners one week later. The expert listeners could repeat the stimuli as often as necessary. Approximately 3 min per patient were necessary to complete the full experiment.

### *Reliability and agreement*

Supplement Table 1 lists the intrarater (exact and close) agreement and disagreement for each listener separated per variable converted into ordinal categories, by dividing the visual analog scale into four equal parts labeled 'good' (normal), 'fair', 'moderate', and 'poor' (abnormal) [15]. Agreement occurred in >73% per rater. The strength of the correlation between the individual judgments (test-retest reliability of fragment A compared to fragment B) of each rater on a 0–1000 scale was also quite high (single-measure Intraclass Correlation Coefficient (ICC(3,1)) for [consistency] using a two-way mixed model; see Supplement Table 1 for the corresponding ICC(3,1) values and confidence intervals per variable). Therefore, for further analysis the mean opinion scores were used to define the agreement and disagreement between the two listeners. Supplement Table 2 provides the interrater reliability and agreement of the raters' mean opinion scores. As can be seen, scores were in exact agreement (difference ≤125 points) in 6–21 cases (27–96%), in close agreement (difference ≤250 points) in 1–12 cases (5–55%), and in disagreement in 1–9 cases (5–41%), depending on the variable analyzed. Except for prosody, all variables demonstrated ICC(3,1) values of 0.75 or higher, indicating good reliability. For prosody the ICC(3,1) was 0.60, indicating acceptable reliability [22,23]. Hence, for overall analysis of perceptual evaluation, average scores between the two raters' mean

opinion scores were used to evaluate perceptual voice and speech parameters.

#### Automatic speech recognition

Automatic assessment of voice quality and speech was conducted with the Automatic Speech analysis In Speech Therapy for Oncology (ASISTO) expert system [12,13,24]. The assessment models used in this paper have been developed and tested on speech recordings of a similar group of Dutch speakers with HNC before and after CRT [12,13]. Perceptual variables analyzed were Automatic Voice Quality Index (AVQI) and two different systems for determining Running Speech Intelligibility. These latter two expert systems are developed by the Department of Electronics and Information Systems, University of Gent, Belgium; one for text-aligned (ELIS [25]) and one for alignment-free (ELISALF) evaluation [12,13]. AVQI results ranged from 1 to 8 with 1 meaning 'most similar to normal' and 8 meaning 'least similar to normal'. Similarly, Running Speech Intelligibility results ranged from 0 to 100 with 0 meaning 'no phonemes recognized' and 100 meaning 'all phonemes recognized'.

#### Patient-reported outcomes

Patients' perceived voice and speech impairment and related quality of life was assessed with two validated specific voice and speech related quality of life questionnaires: the Voice Handicap Index (VHI) and the Speech Handicap Index (SHI).

The VHI is a 30-item questionnaire scored on a 0–4 point scale for measuring patients' suffering caused by dysphonia, specified into 3 subscales (physical, functional, emotional) identified with 10 items each. The total VHI score can range from 0 to 120 with a higher score corresponding to a higher degree of patient-reported vocal handicap (VHI score 0–30: minimal handicap; 31–60: moderate handicap; 60–120: significant and serious handicap) [26,27]. A cut-off score of 15 points (97% sensitivity and 86% specificity) has been established to identify patients with HNC and voice problems in daily life [28].

Based on the VHI, the SHI has been developed as a valid speech assessment tool for patients with HNC, to provide insight into the nature and severity of patients' speech complaints. Instructions and grading are identical to the VHI, but now adapted to speech-related problems in daily life [29,30]. The total SHI score is calculated by summing the scores on all 30 items (score range 0–120), with a higher score indicating a higher level of speech-related problems. A cut-off score of 6 or higher (95% sensitivity and 90% specificity) has been established for speech problems in daily life, and a difference score of 12 points or higher has been proposed as criterion for clinical significance in-group comparisons [31]. Furthermore, there are two SHI subscales: psychosocial function (14 items, score range 0–56) and speech function (14 items, score range 0–56). The questionnaire also includes a global question "how is your speech today", with 4 response categories ('good', 'reasonable', 'poor', and 'severe').

#### Statistical analysis

Descriptive statistics were generated for all continuous outcome measures at the 10-years+ assessment point. Data were summarized as medians with associated range. Spearman's rank correlation was used to determine significant associations between perceptual, automatic and/or patient-reported outcome variables. The Mann-Whitney U test was used to compare outcome variables between two unpaired groups (i.e. IMRT vs. conventional radiotherapy). Pearson's Chi-Square test was used to test associations or differences in proportion between two or more groups. All data

were collected and analyzed in SPSS (Chicago, Illinois; version 23.0), and a significance level of  $p < 0.05$  was used.

## Results

At 10-years+ post-treatment (median 134 months; range 109–165 months), 22 patients (13 male, 9 female; current mean age: 62 years, range 42–74) were evaluable. All patients were in complete remission. The majority of patients (82%) had a primary tumor located in the oropharynx. The clinical patients' and tumor characteristics of the analyzed cohort at 10-years+ post-treatment ( $n = 22$ ) and the original patient cohort at baseline ( $n = 207$ ) recently have been extensively described [17]. There were no significant differences in proportion between these two groups with respect to gender, tumor site, stage, or treatment ( $p > .05$ ). In Table 1 the perceptual, automatic, and patient-reported voice and speech outcome parameters in 22 patients with HNC at 10-years+ post-treatment are demonstrated.

#### Perceptual evaluation

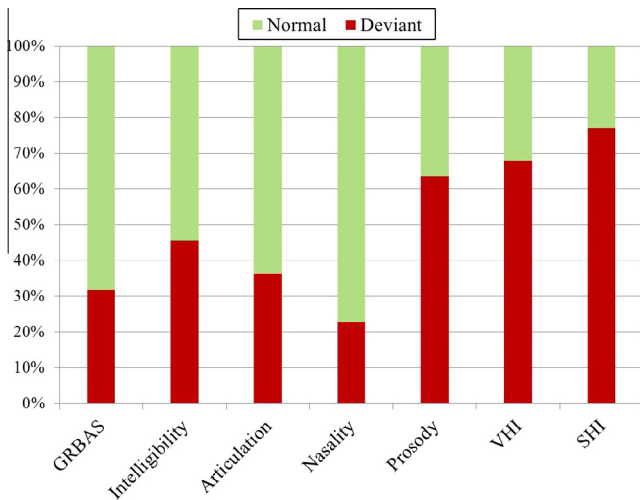
For perceptual evaluation by the SLPs, mean scores (Table 1) were also converted into a four-point ordinal scale 'good', 'fair', 'moderate', and 'poor', whereby the top 25% was labeled as 'normal', and the remainder as 'deviant' (Fig. 1). As can be seen, prosody was most frequently judged as deviant (in 64% of cases), followed by intelligibility (46%), articulation (36%), and voice quality (one or more deviant parameter(s) of the GRBAS; 32%). In total 18/22 patients (82%) showed impairments (deviant scores) on one or more of the outcome parameters. Except for overall grade of voice quality and breathiness, which were significantly more deviant in patients with hypopharyngeal tumors (Mann-Whitney U test; grade:  $p = .040$ ; breathiness:  $p = .005$ ), no correlations between perceptual outcome variables and tumor characteristics were found. Speech intelligibility strongly correlated with articulation ( $r = 0.93$ ;  $p < .001$ ), and nasality ( $r = 0.67$ ,  $p = .001$ ), whereas overall grade of voice quality significantly correlated with

**Table 1**

Descriptive statistics and distribution by domain of perceptual, automatic, and patient-reported voice and speech variables in 22 head and neck cancer patients at 10-years+ post-treatment.

| Variable (score)             | Min–Max | Median | Mean $\pm$ SD |
|------------------------------|---------|--------|---------------|
| <i>Perceptual evaluation</i> |         |        |               |
| Grade                        | 105–993 | 832    | 743 $\pm$ 245 |
| Roughness                    | 179–995 | 936    | 822 $\pm$ 223 |
| Breathiness                  | 387–999 | 995    | 934 $\pm$ 145 |
| Asthenia                     | 687–999 | 987    | 961 $\pm$ 71  |
| Strain                       | 360–998 | 969    | 888 $\pm$ 186 |
| Nasality                     | 6–991   | 877    | 794 $\pm$ 284 |
| Prosody                      | 293–998 | 721    | 693 $\pm$ 214 |
| Speech intelligibility       | 113–987 | 771    | 689 $\pm$ 256 |
| Articulation                 | 94–983  | 842    | 722 $\pm$ 270 |
| <i>Automatic evaluation</i>  |         |        |               |
| Voice quality (AVQI)         | 3.7–6.1 | 4.7    | 4.9 $\pm$ 0.6 |
| Intelligibility (ELIS)       | 62–94   | 83     | 82 $\pm$ 9    |
| Intelligibility (ELISALF)    | 67–92   | 85     | 82 $\pm$ 8    |
| <i>Subjective evaluation</i> |         |        |               |
| Voice Handicap Index         | 0–57    | 21     | 22 $\pm$ 18   |
| Physical domain              | 0–22    | 10     | 10 $\pm$ 8    |
| Functional domain            | 0–19    | 6.5    | 7 $\pm$ 6     |
| Emotional domain             | 0–18    | 3      | 5 $\pm$ 5     |
| Speech Handicap Index        | 0–65    | 21.5   | 24 $\pm$ 20   |
| Speech domain                | 0–38    | 13.5   | 16 $\pm$ 12   |
| Psychosocial domain          | 0–26    | 5      | 7 $\pm$ 8     |

Abbreviations: Min = minimum; Max = maximum; SD = standard deviation; AVQI = Automatic Voice Quality Index; ELIS = text-aligned Running Speech Intelligibility [25]; ELISALF = alignment-free Running Speech Intelligibility.



**Fig. 1.** Percentages of patients ( $n = 22$ ) with 'normal' or 'deviant' perceptual and patient-reported voice and speech parameters. Note: for perceptual scores the top 25% was labeled as 'normal', and the remainder as 'deviant'. For patient-reported outcome parameters 'deviant' scores were based on validated cut-offs [28,31].

roughness ( $r = 0.94$ ;  $p = .000$ ), and strain ( $r = 0.89$ ;  $p = .000$ ). Patients treated with IMRT (45%) showed significant better intelligibility scores compared to patients treated with conventional radiotherapy (55%; see Table 2).

#### Automatic evaluation

Table 1 shows the descriptive statistics at 10-years+ post-treatment for automatic assessment of voice quality (AVQI) and speech intelligibility. AVQI scores ranged from 3.66 to 6.08 (with 1 meaning 'most similar to normal' and 8 meaning 'least similar to normal'). A trend was seen for a moderate correlation between AVQI and perceptual voice quality scores by the SLPs ( $r = 0.42$ ;  $p = .051$ ; see Fig. 2). Patients with a tumor located in the

hypopharynx showed significantly worse AVQI scores ( $n = 3$ ; mean 5.77; range 5.47–6.08) compared to the patients with a tumor located in the oral cavity/oropharynx ( $n = 19$ ; mean 4.72; range 3.66–5.95; Mann–Whitney  $U$  test;  $p = .009$ ). Regarding (ELIS) speech intelligibility, scores ranged from 62.21 to 93.87 (Table 1). There was a significant correlation with perceptual scores of speech intelligibility ( $r = 0.74$ ;  $p = .000$ ; see Fig. 2).

#### Patient-reported outcomes

Voice Handicap Index (VHI) and Speech Handicap Index (SHI) scores were used to assess patients' perspective and related quality of life of voice and speech dysfunction. In Table 1 the distribution of the various subdomains at 10-years+ post-treatment are shown. Patients with a physical voice disability mainly reported problems such as increased vocal effort, breathiness, and unpredictable/varying clarity of voice, resulting in functional disabilities such as poor understandability by others, in particular during phone calls or in noisy rooms. Patients with speech problems instead more often complained about unpredictably/varying intelligibility and unclear articulation. Overall, deviant SHI scores (SHI > 6) were present in 77% of patients (17/22), whereas 68% (15/22) showed voice problems (VHI > 15). In the psychosocial voice and speech domains hardly any disabilities were reported (median scores 3 and 5, respectively; see Table 1). Patients treated with IMRT (45%) showed significant better scores on all domains compared to patients treated with conventional radiotherapy (55%; see Table 2). Correlation with perceptual and automatic outcome measures (i.e. overall grade of voice quality, speech intelligibility) was poor ( $r < 0.4$ ), except for the question "how is your speech today", which significantly but moderately correlated with automatically assessed speech intelligibility ( $r = 0.46$ ,  $p = .032$ ).

#### Discussion

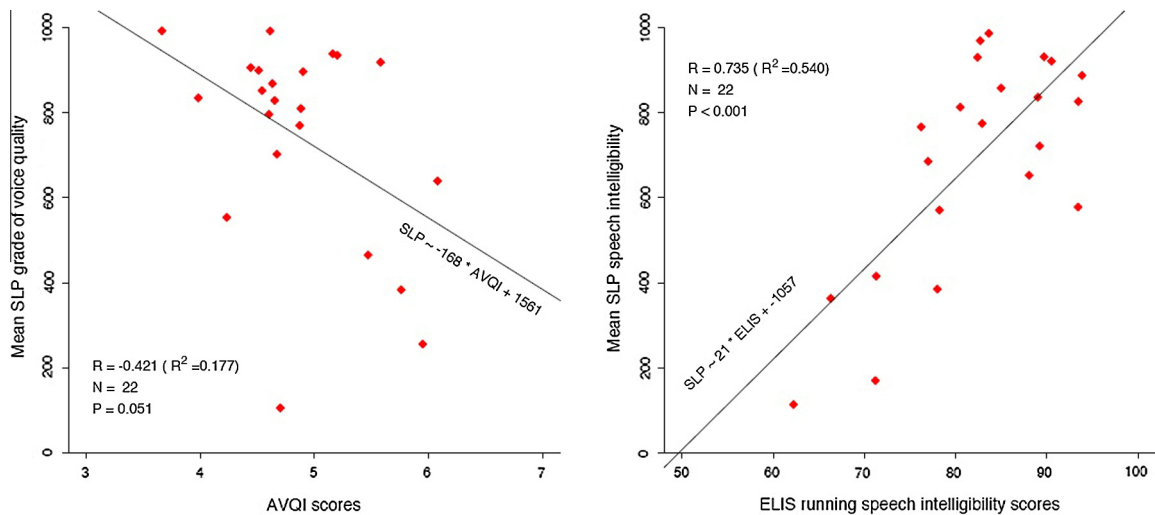
This study assessed long-term (10-years+) objective and subjective voice and speech outcomes following organ-preservation

**Table 2**

Perceptual, automatic, and patient-reported voice and speech variables in 22 patients with HNC at 10-years+ post-treatment, divided by radiotherapy treatment (Intensity-Modulated Radiotherapy [IMRT] versus conventional radiotherapy [CONV]).

| Variable (score)                         | RTx  | N valid | Min–Max | Median | Mean $\pm$ SD   | Statistic                    |
|--|------|---------|---------|--------|-----------------|------------------------------|
| Perceptual voice quality (Grade)         | IMRT | 10      | 465–993 | 875    | 797 $\pm$ 180   | $p = .38$                    |
|  | CONV | 12      | 105–993 | 813    | 698 $\pm$ 288   |                              |
| Automatic voice quality (AVQI)           | IMRT | 10      | 3.7–6.1 | 4.9    | 4.9 $\pm$ 0.7   | $p = .82$                    |
|  | CONV | 12      | 4.0–6.0 | 4.7    | 4.8 $\pm$ 0.5   |                              |
| Voice Handicap Index                     | IMRT | 10      | 0–49    | 2      | 12.5 $\pm$ 17.1 | <b><math>p = .021</math></b> |
|  | CONV | 12      | 9–57    | 26     | 30.2 $\pm$ 14.3 |                              |
| Physical domain                          | IMRT | 10      | 0–22    | 1.5    | 6.6 $\pm$ 8.6   | <b><math>p = .050</math></b> |
|  | CONV | 12      | 3–22    | 16     | 13.7 $\pm$ 6.3  |                              |
| Functional domain                        | IMRT | 10      | 0–16    | 0.5    | 3.5 $\pm$ 5.2   | <b><math>p = .007</math></b> |
|  | CONV | 12      | 0–19    | 8.5    | 9.6 $\pm$ 5.3   |                              |
| Emotional domain                         | IMRT | 10      | 0–14    | 0      | 2.4 $\pm$ 4.5   | <b><math>p = .011</math></b> |
|  | CONV | 12      | 0–18    | 6.5    | 6.9 $\pm$ 5.4   |                              |
| Perceptual speech intelligibility        | IMRT | 10      | 416–987 | 873    | 828 $\pm$ 171   | <b><math>p = .006</math></b> |
|  | CONV | 12      | 113–922 | 616    | 574 $\pm$ 263   |                              |
| Running speech intelligibility (ELIS)    | IMRT | 10      | 71–94   | 83     | 84 $\pm$ 6.4    | $p = .82$                    |
|  | CONV | 12      | 62–93   | 79     | 81 $\pm$ 10.5   |                              |
| Running speech intelligibility (ELISALF) | IMRT | 10      | 69–92   | 86     | 83 $\pm$ 8.4    | $p = .50$                    |
|  | CONV | 12      | 67–91   | 82     | 81 $\pm$ 8.7    |                              |
| Speech Handicap Index                    | IMRT | 10      | 0–53    | 5.5    | 14.0 $\pm$ 18.5 | <b><math>p = .021</math></b> |
|  | CONV | 12      | 10–65   | 27.5   | 31.4 $\pm$ 18.2 |                              |
| Speech domain                            | IMRT | 10      | 0–33    | 5.5    | 9.9 $\pm$ 11.7  | <b><math>p = .030</math></b> |
|  | CONV | 12      | 7–38    | 21     | 20.8 $\pm$ 10.6 |                              |
| Psychosocial domain                      | IMRT | 10      | 0–20    | 0      | 4.0 $\pm$ 7.0   | <b><math>p = .017</math></b> |
|  | CONV | 12      | 1–26    | 6      | 10.3 $\pm$ 8.5  |                              |

Abbreviations: RTx = radiotherapy treatment; Min = minimum; Max = maximum; SD = standard deviation; IMRT = Intensity-Modulated Radiotherapy; CONV = conventional radiotherapy; AVQI = Automatic Voice Quality Index. Note:  $p$ -value according to Mann–Whitney  $U$  test; significance level at  $p < 0.05$ .



**Fig. 2.** Relationship between automatic evaluation of voice quality (AVQI scores) and perceptual evaluation of voice quality by the SLPs (left), and between automatic text-aligned evaluation of running speech intelligibility (ELIS scores) and perceptual evaluation of speech intelligibility by the SLPs (right).

treatment for advanced HNC. Results of the 22 evaluable patients showed considerable functional deficits in this respect. Perceptual evaluation by the SLPs, rating overall speech intelligibility, the precision of articulation, the GRBAS criteria, prosody, and nasality, revealed that 86% of patients showed impairments on one or more of the outcome parameters. The automatic expert system ASISTO, rating automatic voice quality index (AVQI) and running speech intelligibility, seemed to support the perceptual evaluation results of the SLPs, since there were significant, moderate to strong correlations with overall grade of voice quality and with speech intelligibility. Subjective voice and speech complaints were evaluated in the present patient cohort with (sub) total VHI and SHI scores, and revealed moderate but clinically relevant disabilities, that were present in 68% and 77% of patients, respectively.

Other studies evaluating patient-reported voice and speech outcomes after treatment for HNC also demonstrated decreased voice quality following CRT [11,32], with impact on quality of life and psychosocial function [33]. One of the first VHI evaluations after CRT for stage III–IV HNC was performed by Keereweer and colleagues. Mild to severe voice impairment was found in all of the 20 participating patients, who were at least 2.5 years after treatment [32]. In the study of Vainshtein and colleagues, almost 20% of patients reported further voice worsening at 18- and 24-months follow-up after chemo-IMRT for stage III–IV oropharyngeal cancer, most commonly due to worsening vocal clarity [11]. Speech problems were also found in recent studies that evaluated post-treatment SHI scores [8,31]. Rinkel et al. reported impaired speech in daily life (SHI > 6) in 55% of patients with primary HNC (all subsites and stages included), whereas in our study this was 77%. The higher prevalence of disabilities in the current study might be attributable to the more advanced tumor stage with only stage IV tumors included. Furthermore, the follow-up time in the current study was considerably longer (11 years versus a maximum of 5 years in the other studies), which might reflect a further deterioration post CRT over time, as recently also was found for dysphagia issues [17,34].

Interestingly, the problems were predominantly related to radiation technique, because patients treated with IMRT showed significantly less voice and speech problems on the various domains compared to patients treated with conventional radiotherapy. This is in line with other studies that found correlations between radiation dose to the glottis and voice quality worsening or speech impairment after IMRT [11,35]. In the literature, it has been found that radiation dose to the larynx correlates with

laryngeal edema severity, resulting in vocal cord dysfunction and thus poor voice quality [5,6]. This might explain why the patients with a hypopharynx tumor in the current cohort showed more voice problems compared to the others, because high doses to the larynx are unavoidable in these patients, although this concerned only three patients. For non-laryngeal HNC, IMRT may reduce the radiation dose to the pharynx [36], resulting in less edema, fibrosis, and structural alteration of the vocal tract, and thus better speech intelligibility [35]. Ongoing clinical trials in HNC are currently trying to optimize the IMRT process to further improve outcomes [37].

Relation to radiation technique was previously also found for dysphagia and quality of life issues [17,38]. It is therefore not unlikely that the patients who developed both functional deficits (dysphagia and voice/speech problems were significantly correlated in the current cohort; results not published) received higher radiotherapy doses on the muscles or structures critical to these functions. Besides, none of the patients had participated in a preventive rehabilitation program, which has been associated with better post-treatment functional outcomes [2].

Although perceptual evaluation is currently a widely used assessment tool for voice and speech evaluation, we also performed automatic assessment of voice quality and speech intelligibility with the expert system ASISTO [24]. This system has previously been shown to be as accurate as SLPs ( $n = 13$ ) for evaluation of patients treated for HNC [12]. To our knowledge, this is the first practical/clinical application of automatic assessment of voice quality and speech in a HNC patient population with considerable functional deficits following organ-preservation treatment. Additionally, the system was used to evaluate possible bias/subjectivity within perceptual evaluation. The ASISTO scores for speech intelligibility correlated strongly with perceptual mean opinion scores of speech intelligibility, while this correlation was only moderate and borderline significant for voice quality. Possibly, some bias can be blamed here, since only two SLPs participated as listeners in the present study, and they rated voice quality as less severe compared to the system in 15/22 (68%) of patients (Fig. 2). This indicates that their judgement might have been somewhat 'colored' and thus overrated by their extensive experience with patients with HNC. Intelligibility results correlated well, and thus were probably not overrated, which is conceivable because it is easier to score whether one understands something than to rate voice quality, as was found in previous studies [12,39].

Despite the acceptable correlations, it is obvious that perceptual evaluation by SLPs is still not identical to that of a computer program. With regards to radiation technique, minor differences between groups can be statistically significant in one evaluation and just not anymore in the other, especially when numbers are small as in the current study. Moreover, our ASR has not been trained/calibrated on the severest pathological voices in HNC patients, and earlier research with this tool has shown that very low perceptual scores are somewhat more difficult to predict [12,39]. This might have obscured the RT-induced perceptual difference found for SLP assessment. Nevertheless, these differences in outcomes between the two evaluation methods thus have to be interpreted with caution.

We did not measure other acoustic voice parameters (e.g. voicedness, fundamental frequency), since multiple studies have demonstrated that these modalities (independently) have no clear role in the management of patients with cancers of the oral cavity and oropharynx, due to lack of reproducible results, poor correlation with other speech assessment methods (e.g. perceptive or subjective evaluation), and absence of standard protocols [40,41]. In fact, automatic evaluation with ASISTO could also apply as such 'acoustic' parameter, since AVQI is a weighted combination of acoustic parameters [42], and running speech intelligibility is the recognition result of a phoneme recognizer based on the audio signal [12]. Unfortunately, because standardized procedures of objective voice and speech assessments do not exist, yet, results are difficult to compare with other studies performed at different clinics or centers [7].

## Conclusion

Ten years after organ-preservation treatment, functional voice and speech problems are common in this patient cohort, as assessed with perceptual evaluation automatic speech recognition, and with validated structured questionnaires. There were fewer complaints in patients treated with IMRT than with conventional radiotherapy.

## Conflict of interest statement

This study was made possible by grants provided by Atos Medical (Sweden), "Stichting de Hoop" (The Netherlands), and the "Verweilius Foundation" (The Netherlands).

## Acknowledgements

Catherine Middag and Jean-Pierre Martens (Department of Electronics and Information Systems, University of Gent, Belgium) are greatly acknowledged for their collaboration regarding ASISTO; Irene Jacobi (PhD, The Netherlands Cancer Institute) is acknowledged for her help with the speech recordings; Klaske van Sluis (SLP, The Netherlands Cancer Institute) is acknowledged for her collaboration with the perceptual analysis. Wilma van Heemsbergen, epidemiologist and clinical researcher (The Netherlands Cancer Institute), is greatly acknowledged for her support and advice in the statistical analysis.

## Appendix A. Excerpt from 'De vijvervrouw' by Godfried Bomans (in Dutch)

### A.1. Fragment A

Er leefden eens een koning en een koningin en die hadden maar één kind. Dat was de prins. De prins was erg verwend. Toen hij nog in de wieg lag, kreeg hij al een gouden rammelaar. Hij at van een

gouden bordje en hij dronk uit een gouden bekertje. Al zijn speelgoed was van goud, en het werd steeds moeilijker om hem iets te geven, wat hij al niet had.

### A.2. Fragment B

En toen hij achttien jaar werd, had hij alles wat hij maar bedenken kon en het was allemaal van zuiver goud. Maar hij was toch jarig en er moest hem iets gegeven worden. De prins stond bij het raam, toen zijn ooms en tantes binnenkwamen. Zij hadden ieder een cadeautje in de hand, maar ze waren erg verlegen, want ze begrepen wel dat de prins het al had.

## Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.oraloncology.2016.02.001>.

## References

- [1] Jacobi I, van der Molen L, Huiskens H, van Rossum MA, Hilgers FJ. Voice and speech outcomes of chemoradiation for advanced head and neck cancer: a systematic review. *Eur Arch Otorhinolaryngol* 2010;267:1495–505.
- [2] Kraaijenga SA, van der Molen L, Jacobi I, Hamming-Vrieze O, Hilgers FJ, van den Brekel MW. Prospective clinical study on long-term swallowing function and voice quality in advanced head and neck cancer patients treated with concurrent chemoradiotherapy and preventive swallowing exercises. *Eur Arch Otorhinolaryngol* 2015;272(11):3521–31.
- [3] Lazarus CL. Effects of chemoradiotherapy on voice and swallowing. *Curr Opin Otolaryngol Head Neck Surg* 2009;17:172–8.
- [4] Paleri V, Carding P, Chatterjee S, Kelly C, Wilson JA, Welch A, et al. Voice outcomes after concurrent chemoradiotherapy for advanced nonlaryngeal head and neck cancer: a prospective study. *Head Neck* 2012;34:1747–52.
- [5] Fung K, Yoo J, Leeper HA, Hawkins S, Heeneman H, Doyle PC, et al. Vocal function following radiation for non-laryngeal versus laryngeal tumors of the head and neck. *Laryngoscope* 2001;111:1920–4.
- [6] Hamdan AL, Geara F, Rameh C, Hussein ST, Eid T, Fuleihan N. Vocal changes following radiotherapy to the head and neck for non-laryngeal tumors. *Eur Arch Otorhinolaryngol* 2009;266:1435–9.
- [7] Schuster M, Stelzle F. Outcome measurements after oral cancer treatment: speech and speech-related aspects—an overview. *Oral Maxillofac Surg* 2012;16:291–8.
- [8] Lazarus CL, Husaini H, Hu K, Culliney B, Li Z, Urken M, et al. Functional outcomes and quality of life after chemoradiotherapy: baseline and 3 and 6 months post-treatment. *Dysphagia* 2014;29:365–75.
- [9] Dwivedi RC, Kazi RA, Agrawal N, Nutting CM, Clarke PM, Kerawala CJ, et al. Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers. *Cancer Treat Rev* 2009;35:417–24.
- [10] van der Molen L, van Rossum MA, Jacobi I, van Son RJ, Smeele LE, Rasch CR, et al. Pre- and posttreatment voice and speech outcomes in patients with advanced head and neck cancer treated with chemoradiotherapy: expert listeners' and patient's perception. *J Voice* 2012;26(664):e25–33.
- [11] Vainshtein JM, Griffith KA, Feng FY, Vineberg KA, Chepeha DB, Eisbruch A. Patient-reported voice and speech outcomes after whole-neck intensity modulated radiation therapy and chemotherapy for oropharyngeal cancer: prospective longitudinal study. *Int J Radiat Oncol Biol Phys* 2014;89(5):973–80.
- [12] Clapham R, Middag C, Hilgers F, Martens J-P, van den Brekel M, van Son R. Developing automatic articulation, phonation and accent assessment techniques for speakers treated for advanced head and neck cancer. *Speech Commun* 2014;59:44–54.
- [13] Middag CC, van Son R, Martens JP. Robust automatic intelligibility assessment techniques evaluated on speakers treated for head and neck cancer. *Comput Speech Lang* 2014;28:467–82.
- [14] Kitzing P, Maier A, Ahlander VL. Automatic speech recognition (ASR) and its use as a tool for assessment or therapy of voice, speech, and language disorders. *Logoped Phoniatri Vocol* 2009;34:91–6.
- [15] Clapham RP, van As-Brooks CJ, van Son RJ, Hilgers FJ, van den Brekel MW. The relationship between acoustic signal typing and perceptual evaluation of tracheoesophageal voice quality for sustained vowels. *J Voice* 2015;29(4):23–9.
- [16] Rasch CR, Hauptmann M, Schornagel J, Wijers O, Buter J, Gregor T, et al. Intra-arterial versus intravenous chemoradiation for advanced head and neck cancer: results of a randomized phase 3 trial. *Cancer* 2010;116:2159–65.
- [17] Kraaijenga SA, Oskam IM, van der Molen L, Hamming-Vrieze O, Hilgers FJ, van den Brekel MW. Evaluation of long term (10-years+) dysphagia and trismus in patients treated with concurrent chemo-radiotherapy for advanced head and neck cancer. *Oral Oncol* 2015;51(8):787–94.

- [18] Free downloadable at <http://www.praat.org>.
- [19] Open Source program TEVA; available at [http://www.fon.hum.uva.nl/IFASpokenLanguageCorpora/NKlcorpora/NKl\\_TEVA/](http://www.fon.hum.uva.nl/IFASpokenLanguageCorpora/NKlcorpora/NKl_TEVA/).
- [20] Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.05.
- [21] Hirano M. Clinical examination of voice. New York: Springer-Verlag; 1981.
- [22] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
- [23] Portney LG, Watkins MP. Foundations of clinical research: applications to practice; Appleton & Lange; 1993.
- [24] ASISTO expert system; available at <http://asisto.elis.ugent.be/>.
- [25] ELIS: 'Elektronica en Informatie Systemen'; available at <https://elis.ugent.be/>.
- [26] Jacobsen BJA, Grywalski C, Silbergleit A, Jacobsen G, Benninger M. The Voice Handicap Index (VHI): development and validation. *Am J Speech-Lang Pathol* 1997;6:66–70.
- [27] Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, Guimaraes I, Holmberg EB, Nawka T, et al. Validation of the voice handicap index by assessing equivalence of European translations. *Folia Phoniatr Logop* 2008;60:173–8.
- [28] Van Gogh CD, Mahieu HF, Kuik DJ, Rinkel RN, Langendijk JA, Verdonck-de Leeuw IM. Voice in early glottic cancer compared to benign voice pathology. *Eur Arch Otorhinolaryngol* 2007;264:1033–8.
- [29] Rinkel RN, Verdonck-de Leeuw IM, van Reij EJ, Aaronson NK, Leemans CR. Speech Handicap Index in patients with oral and pharyngeal cancer: better understanding of patients' complaints. *Head Neck* 2008;30:868–74.
- [30] Dwivedi RC, St Rose S, Roe JW, Chisholm E, Elmiyeh B, Nutting CM, et al. First report on the reliability and validity of speech handicap index in native English-speaking patients with head and neck cancer. *Head Neck* 2011;33:341–8.
- [31] Rinkel RN, Verdonck-de Leeuw IM, Doornaert P, Buter J, de Bree R, Langendijk JA, et al. Prevalence of swallowing and speech problems in daily life after chemoradiation for head and neck cancer based on cut-off scores of the patient-reported outcome measures SWAL-QOL and SHI. *Eur Arch Otorhinolaryngol* 2015. June 14 [Epub ahead of print].
- [32] Keereweer S, Kerrebijn JD, Al-Mamgani A, Sewnaik A, Baatenburg de Jong RJ, van Meerten E. Chemoradiation for advanced hypopharyngeal carcinoma: a retrospective study on efficacy, morbidity and quality of life. *Eur Arch Otorhinolaryngol* 2012;269:939–46.
- [33] Rinkel RN, Verdonck-de Leeuw IM, van den Brakel N, de Bree R, Eerenstein SE, Aaronson N, et al. Patient-reported symptom questionnaires in laryngeal cancer: voice, speech and swallowing. *Oral Oncol* 2014;50:759–64.
- [34] Hutcheson KA, Lewin JS, Barringer DA, Lisec A, Gunn GB, Moore MW, et al. Late dysphagia after radiotherapy-based treatment of head and neck cancer. *Cancer* 2012;118:5793–9.
- [35] Nguyen NP, Abraham D, Desai A, Betz M, Davis R, Sroka T, et al. Impact of image-guided radiotherapy to reduce laryngeal edema following treatment for non-laryngeal and non-hypopharyngeal head and neck cancers. *Oral Oncol* 2011;47(9):900–4.
- [36] Roe JW, Carding PN, Dwivedi RC, Kazi RA, Rhys-Evans PH, Harrington KJ, et al. Swallowing outcomes following Intensity Modulated Radiation Therapy (IMRT) for head & neck cancer – a systematic review. *Oral Oncol* 2010;46:727–33.
- [37] Tejjal G, Jaiprakash A, Susovan B, Ghosh-Laskar S, Murthy V, Budrukkar A. IMRT and IGRT in head and neck cancer: have we delivered what we promised? *Indian J Surg Oncol* 2010;1:166–85.
- [38] Rathod S, Gupta T, Ghosh-Laskar S, Murthy V, Budrukkar A, Agarwal J. Quality-of-life (QOL) outcomes in patients with head and neck squamous cell carcinoma (HNSCC) treated with intensity-modulated radiation therapy (IMRT) compared to three-dimensional conformal radiotherapy (3D-CRT): evidence from a prospective randomized study. *Oral Oncol* 2013;49:634–42.
- [39] Van Nuffelen G, Middag C, De Bodt M, Martens JP. Speech technology-based assessment of phoneme intelligibility in dysarthria. *Int J Lang Commun Disord* 2009;44:716–30.
- [40] Finizia C, Dotevall H, Lundstrom E, Lindstrom J. Acoustic and perceptual evaluation of voice and speech quality: a study of patients with laryngeal cancer treated with laryngectomy vs irradiation. *Arch Otolaryngol Head Neck Surg* 1999;125:157–63.
- [41] Dwivedi RC, St Rose S, Chisholm EJ, Clarke PM, Kerawala CJ, Nutting CM, et al. Acoustic parameters of speech: lack of correlation with perceptual and questionnaire-based speech evaluation in patients with oral and oropharyngeal cancer treated with primary surgery. *Head Neck* 2014. December 18 [Epub ahead of print].
- [42] Maryn Y, de Bodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord* 2010;43:161–74.