interspeech 2008 incorporating SST 08

Phonetic-acoustic and feature analyses by a neural network to assess speech quality in patients treated for head and neck cancer

Marieke de Bruijn[1], *Irma Verdonck- de Leeuw*[1], *Louis ten Bosch*[2], *Joop Kuik*[1], *Hugo Quené*[3], *Lou Boves*[2], *Hans Langendijk*[4], *René Leemans*[1]

[1] Department of Otolaryngology/Head and Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands
[2] Department of Language and Speech, University of Nijmegen, Nijmegen, The Netherlands
[3] Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands
[4] Department of Radiation Oncology, University of Groningen Medical Center, Groningen, The Netherlands
m.debruijn@vumc.nl, im.verdonck@vumc.nl

10.21437/Interspeech.2008-483

Abstract

Subjective speech evaluation is the gold standard to assess speech quality of head and neck cancer patients. This study investigates if conventional acoustic-phonetic and novel feature analysis contribute to the development of a multidimensional speech assessment protocol. Speech recordings of 51 patients 6 months post-treatment and of 18 control speakers were subjectively evaluated for intelligibility, nasal resonance and articulation. Self-evaluation of speech problems was assessed by the EORTC QLQ-H&N35 speech subscale. Feature analysis was performed to assess objectively nasality in vowels /a,i,u/. Results revealed that size of the vowel triangle, pressure release of /k/ and nasality in /i/ predict best intelligibility, articulation and nasal resonance and differentiated best between patients and controls. Within patients, /k/ and /x/ differentiated tumour site and tumour classification. Various objective variables were related to speech problems as reported by patients.

Index Terms: head and neck cancer, reconstruction, speech quality, nasality, intelligibility, feature analysis, objective evaluation, subjective evaluation

1. Introduction

A tumour in the oral cavity or oropharynx and its treatment may damage various anatomic structures involved in speech production. Functional impairment of the tongue, velum, and other speech organs is common, which leads to speech impairment difficulties related to mastication, swallowing, articulation interfering with communication, and social activities. These problems may ultimately result in a lower quality of life [1].

Speech quality depends on volume and most notably on the mobility of residual speech organs [2, 3]. Speech sounds may be affected, substituted, or omitted due to tumour growth and treatment.

Surgical techniques have improved significantly during the past two decades. Reconstruction techniques for larger defects have been developed using free skin flaps, e.g., the radial forearm flap. Thin, pliable skin flaps yield optimal speech results [4-5]. These flaps have proven to be very reliable with limited donor site morbidity.

Speech assessment can be performed in various ways. Subjective evaluation of speech by trained or untrained raters. Speech problems in daily life can be evaluated by patients themselves using questionnaires. Objective techniques of assessing speech quality are usually performed by phonetic-acoustic analysis of speech signals or indicators such as a nasometer or ElectroPalatoGram. Results of earlier studies revealed that after glossectomy, patients' intelligibility is often deteriorated. Although speech therapy can improve speech quality, some patients require a secondary pasty using a skin graft resulting in better mobility of the residual tongue and better intelligibility [2, 3, 7]. Another study revealed that speech quality of patients with a tumour in the oral cavity or oropharynx is related to tumour classification and location. Patients with larger tumours often have increased speech difficulty. Patients with an oropharyngeal tumour experience more difficulties regarding velopharyngeal closure resulting in nasal resonance caused by air escaping from the nose [6].

The most common methods to assess speech outcome in head and neck cancer patients are subjective ratings and self-evaluation. The aim of the present study is to investigate construct and predictive validity of objective speech parameters.

2. Methods

2.1. Speakers

The study population consisted of 51 patients treated at the department of Otolaryngology-Head and Neck Surgery at VU University Medical Center in Amsterdam, the Netherlands from 1998 to 2001. Patients underwent surgery for advanced oral or oropharyngeal carcinoma and soft tissue transfer for the reconstruction of surgical defects. All free flaps were successful. Patients received radiotherapy in case of larger (T3-4) tumours, contaminated surgical margins or nodes or extranodal spread. The primary site received a dose of 56 Gy in total.

Patients were included in the study after written informed consent. All patients were under 75 years of age and were able to participate in functional tests. A control group consisted of 18 age and gender matched participants.

2.2. Speech assessment

Speech data were collected at 6 months post-treatment. Patients and controls read aloud a Dutch text in a sound-proof room. A 30 cm mouth to microphone distance was used for recordings. Speech was digitized using Cool Edit PRO 1.2 (Adobe Systems Incorporated, San Jose, California, USA) with 22kHz sample frequency and 16-bit resolution. For each speaker the recording level was adjusted to optimize signal-to-noise ratio.
Two types of subjective assessments and two types of objective assessments were used to determine speech quality.

In the present study phonemes were objectively analysed by two methods: objective feature analysis by a neural network and acoustic-phonetic analysis. As for objective feature analysis, the amount of distinctive speech features is still debated, ranging up to 25 –such as fricative, plosive, nasal and voice [8, 9]. Detecting features in speech sounds was performed by a neural network that was trained on normal speech as input. The algorithm calculated the presence of features - on a scale from 0 to 1 –of each 10 ms time frame. In the present study, the focus was on the feature nasality. Due to structural changes in the vocal tract of patients, it is likely that they experience velar closure impairment, resulting in an overall nasal speech quality. The algorithm measured the feature nasality in the cardinal vowels /a,i,u/ that were segmented manually. The algorithm was trained on healthy speech of two speakers (male and female) from the IFA-corpus[1]. Feature detection analysis on speech of two other test speakers yielded a good result: 86% of the features were correctly identified at frame-level.
Acoustic-phonetic analysis was performed by the speech processing program PRAAT[2]. Vowels /a,i,u/ and velar consonants /k,x/ were manually segmented from running speech. For each speech sound (/a,i,u,k,x/), two utterances were used for acoustic-phonetic and feature analysis. The formant values of F1 and F2 (in Hz) of the vowels and the size of the vowel triangle (see below), proportion of pressure release of the /k/ as percentage of total duration of /k/ and the steepness of the spectral slope of the /x/ were measured.
Vowels were used because, compared to consonants, they are relatively easy to identify in the speech signal, and easier to analyse acoustically. The formants F1 and F2 of the vowels /a,i,u/ represent the vowel triangle or the most extended positions of the vocal tract and tongue. Moreover, vowel formant analyses proved to be valid measures of speech quality in patients treated for oral cancer in earlier studies [5, 10]. The size of the vowel triangle was measured in Hz². The consonants /k/ and /x/ were chosen because earlier research revealed that patients with an oral or oropharyngeal tumour often have difficulty with the production of velar speech sounds [6, 11]. Of each speech sound, two utterances were segmented from running speech and analysed (i.e. /k/1 and /k/2).

Subjective assessment of pathological speech comprised overall intelligibility by two speech therapists on a 1-10 scale, ranging from worst to best intelligibility. Intelligibility is a result constructed by multiple speech factors of which quality of articulation and nasal resonance have been determined in this study. Articulation and nasal resonance were rated on a 1-4 scale, ranging from normal to deviant speech quality. To enable subjective speech evaluation, a computer program was developed to perform blinded randomized listening experiments and to score automatically, intelligibility, nasality and articulation. Interrater agreement (Cronbach's alpha) for subjective assessment of intelligibility was high: 0.86. Intrarater agreement was high with 100% equal scores between the ratings on the first and second, repeated speech fragments on articulation and nasal resonance. Next to listener judgments, patients valuated their own speech quality, guided by selected items of the speech subscale of the quality of life questionnaire EORTC H&N-35 [12].

2.3. Statistical analyses

Construct validity of objective speech analyses was tested by means of univariate Pearson correlation coefficients between subjective intelligibility, articulation and nasal resonance and objective parameters (formants of the cardinal vowels /a,i,u/, size of the vowel triangle, spectral slope of /x/ and duration of pressure release of /k/ and nasality on the vowels /a,i,u/). To obtain insight into the value of objective parameters in predicting subjective speech evaluation, multivariate regression analyses were performed. Multivariate linear regression analyses (stepwise backward) were performed with intelligibility and self-evaluation as the dependent variables and the objective parameters as the independent variables. Logistic regression analyses (stepwise forward-Wald) were performed with articulation and nasal resonance (both recoded into a binary scale) as the dependent variables and the objective parameters as the independent variables. Mann-Whitney tests were performed to determine the validity of the objective speech parameters regarding known group differences: patient vs. controls, smaller (T2) vs. larger (T3-4) tumours, and tumour location (oral vs. oropharyngeal) .

3. Results

3.1. Patient characteristics

Patient characteristics regarding gender, tumour site and T-classification are shown in table 1.

Table 1. *Characteristics of 51 patients included in the study.*

	n	(%)
Gender		
Male	28	(55)
Female	23	(45)
Tumour site		
Oral cavity	21	(41)
Oropharynx	30	(59)
T-classification		
2	26	(51)
3-4	25	(49)

3.2. Construct validity

Univariate correlations between subjective evaluations and objective parameters were assessed with Pearson correlation coefficients.
Acoustic-phonetic analyses of /k/ and /i/ were related to subjective evaluation of intelligibility (/k/1: r=.50, p<.001,

[1] http://www.fon.hum.uva.nl/IFAcorpus.
[2] www.praat.org.

/k/2: r=.36, p=.003) (/i/ F2: r=.35, p=.003), of articulation (/k/1: r=.40, p=.001, /k/2: r=.25, p=.040) (/i/ F2: r=.36, p=.002), and of nasal resonance (/k/1: r=.25, p=.042, /k/2: r=.39, p=.001) (/i/ F1: r=-.42, p<.001). Acoustic-phonetic analysis of /u/ was related to subjective nasal resonance (/u/: r=-.37, p=.003), but not to intelligibility and articulation. The size of the vowel triangle was related to intelligibility (r=.39, p=.002) and articulation (r=.42, p<.001) but not to nasal resonance.

Significant correlations were found between feature analysis of nasality of /i/ and subjective evaluation of overall intelligibility (/i/: r=-.29, p=.017) and articulation (/i/: r=-.28, p=.021). The total amount of nasality identified –the mean of measured nasality in all frames- in the full read-aloud text was related to all three subjective evaluations (intelligibility: r=-.24, p=.048, articulation: r=-.30, p=.012, nasal resonance: r=-.26, p=.032).

To obtain insight into which combination of objective parameters predicts subjective (self-)assessments, multiple regression analyses were performed.

Subjective evaluation of overall intelligibility was predicted significantly (R^2=.55, p<.001) by acoustic-phonetic analysis of /k/ (t=3.99, p<.001), /i/ (t=3.92,p<.001) and the size of the vowel triangle (t=-3.93,p<001), as well as by objective feature analysis of nasality of /u/ (t=-2.80,p=.007) and /a/ (t=-2.16, p=.035).

Subjective evaluation of articulation was predicted significantly (R^2=.78) by acoustic-phonetic analysis of /k/ (Wald=6.59, p=.010), F1 of /i/ (Wald=8.54, p=.003), F2 of /i/ (Wald=4.20, p=.040), the size of the vowel triangle (Wald=6.56, p=.010) and objective feature analysis of nasality in running speech (Wald=7.66, p=.006).

Subjective assessment of nasal resonance was predicted significantly (R^2=.52) by acoustic-phonetic analyses of /x/ (Wald=9.34, p=.002), /k/ (Wald=7.32, p=.007) and F1 of /i/ (Wald=7.67, p=.006).

Finally, the predictive power of objective parameters was determined for patients self-assessments concerning difficulty of speech in daily life. The steepness of the spectral slope of /x/ (t=3.45, p=.001), F1 of /a/ (t=-2.45, p=.019), F2 of /i/ (t=-3.39, p=.002), the size of the vowel triangle (t=2.09, p=.044), objective feature analysis of nasality on running speech (t=-2.66, p=.012) and subjective evaluation of nasal resonance (t=-3.83, p<.001) predicted significantly (R^2=.45) patients self-assessments.

These results reveal adequate construct validity of objective speech analyses.

3.3. Known group differences

Regarding known group differences, Mann-Whitney tests were performed on patients vs. controls and within patients according to tumour classification and tumour site (table 2).

Significant differences between patients and controls on acoustic-phonetic parameters revealed that patients (P) have a shorter pressure release for /k/ than controls (C) (/k/1: P:28,5%, sd=16. C: 43,5%, sd= 18) and (k2: P: 26,4%, sd=19 C:49,8%, sd=17). Patients have a higher F1 of /i/ than controls (F1 /i/: P: 334 Hz, sd=54. C: 296 Hz, sd=49) but a lower F2 of /i/ than controls (F2/i/: P:2105 Hz, sd=363. C: 2325 Hz, sd=248). The size of the vowel triangle is significantly smaller for patients than for controls (P: .143 Hz^2, sd=.12. C: .213 Hz^2, sd=.11). Patients have significantly more nasality than controls on /a/ (/a/: P: .04 ,sd=.08. C: .00, sd=.01) and on /i/ (/i/1: P: .05, sd=.07. C: .01, sd=.02) and (/i/2: P: .12, sd=.19, C: .03, sd=.04).

Objective feature analysis of nasality did not differentiate between populations with different tumour site and tumour classification. Acoustic-phonetic analysis did differentiate between those groups. Only /x/ distinguished between tumour location: patients with an oropharyngeal tumour had a steeper spectral slope than patients with an oral tumour (/x/: oral: -13, sd=6, oropharynx: -17, sd=6). Patients with smaller tumours had a longer pressure release compared to patients with a larger tumour (/k/: T2: 33%, sd=17. T3-4: 23%, sd=14). There is no difference in articulation rate between patients and control speakers (t= -,412, p<.682).

Table 2. *Significant differences between objective feature of nasality and acoustic-phonetic variables measured on vowels and consonants between pathological and control speakers, and regarding tumour site and tumour stage.*

	Pathological vs. control speakers
nasality feature analysis	
/a/	Z=-2.77, p=.001
/i/1	Z=-2.20, p=.028
/i/2	Z=-2.31, p=.021
acoustic-phonetic analysis	
/k/1	Z=-2.77, p=.006
/k/2	Z=-4.15, p<.001
F1 /i/	Z=-2.36, p=.018
F2 /i/	Z=-2.42, p=.016
size Δ	Z=-2.42, p=.015
	Oral tumour vs. oropharynx tumour
/x/	Z=-2.24, p=.025
	T2 tumour vs. T3-4 tumour
/k/	Z=-2.09, p=.037

4. Discussion

This study presents an inventory of speech performance six months after treatment in a well-defined head and neck cancer patient group. Speech quality was determined using a newly introduced method to objectively measure the feature nasality in pathological speech, objective acoustic-phonetic analysis and commonly used subjective evaluations.

The first aim of the present study was to investigate which objective parameters contribute to the prediction of subjective (self-) evaluations of pathological speech. Especially acoustic-phonetic parameters (/k/, /i/ and the size of the vowel triangle) best predicted subjective assessment of overall intelligibility, articulation, nasal resonance and self-evaluation of speech. Also, the objective feature nasality in the vowels (/a/ and /u/) attributed to the prediction of subjective evaluations. Surprisingly, objectively identified nasality did not attribute to the prediction of subjective evaluation of nasal resonance. This is most probably due to the fact that the perceptual judgement of nasality was based on longer stretches of speech. The method applied in this paper focussed on isolated vowels. To the best of our knowledge this study is the first to apply objective feature analysis of nasality in speech of patients with head and neck cancer. Further research including a larger cohort to train the neural network underlying the feature analyses may provide more detailed insight into the behaviour of the algorithm in pathological speech.

As for the ability to reveal known group differences, feature detection differentiated between the patient and control group regarding the vowels /a/ and /i/: speech of patients was more nasal compared to controls. This result can be explained by structure alterations after tumour involvement and the treatment. Patients have more difficulty with proper velar closure, resulting in air flow through the nose. Further, acoustic-phonetic analysis differentiated between patients and controls regarding pressure release of /k/ and the size of the vowel triangle. Difficulty with production of /k/ originates from velar function difficulties as well. The size of the vowel triangle is mainly caused by deviant formant values of /i/ and is in concordance with earlier studies [13-15]. Acoustic-phonetic analysis revealed differences between tumour classification (/k/) and tumour site (/x/). Like /k/, /x/ is a velar consonant, which is problematic for this patient population. These results are in agreement with earlier research [11].

5. Conclusion

Speech quality of patients with oral or oropharyngeal carcinoma was investigated. Objective feature analysis regarding nasality and acoustic-phonetic analysis proved to be valid. The presented results contribute to further development of a speech analysis protocol to be used in clinical practice. Further research is needed taking into account a larger variation of speech sounds and including larger patient cohorts.

6. References

[1] Borggreven, P.A., Aaronson, N.K., Verdonck- de Leeuw, I.M., Muller, M.J., Heiligers, M.L.C.H., de Bree, R., Langendijk, J.A., Leemans, C.R. (2007). "Quality of life after surgical treatment for oral and oropharyngeal cancer: a prospective longitudinal assessment of patients reconstructed by a microvascular flap oral oncology", Oral Oncology, 10, 1034-1042.

[2] Terai, H. & M. Shimahara. (2004). Evaluation of speech intelligibility after a secondary dehiscence operation using an artificial graft in patients with speech disorders after partial glossectomy. British Journal of Oral and Maxillofacial Surgery, 42, 190-194.

[3] Michi, K. –I., S. Imai, Y. Yamashita & N. Suzuki. (1989). Improvement of speech intelligibility by a secondary operation to mobilize the tongue after glossectomy. Journal of cranio-maxillectomy-facial Surgery, 17, 162-166.

[4] Su, W.F., Hsia, Y.J., Chang, Y.C., Chen, S.G. & H. Sheng, (2003). "Functional comparison after reconstruction with a radial forearm free flap or a pectoralis major flap for cancer of the tongue", Otolaryngol. Head Neck Surgery., 128:412-418

[5] Verdonck-de Leeuw, I.M., ten Bosch, L., Chao, L.Y., Rinkel, R.N.P.M., Borggreven, P.A., Boves, L., & R.C. Leemans. (2007). Speech quality after major surgery of the oral cavity and oropharynx with microvascular soft tissue reconstruction.

[6] Markkanen-Leppänen, M., E. Isotalo, A. Mäkitie, E. Suominen, S. Asko-Seljavaara & M-L. Haapanen. (2005). Speech aerodynamics and Nasality in oral cancer patients treated with microvascular transfers. The Journal of Craniofacial Surgery, 6, 990-995.

[7] Furia, C., L. Kowalski, M. Latorre, E. Angelis, N. Martins, A. Barros & K. Ribeiro. (2001). Speech intelligibility after glossectomy and speech rehabilitation. Archives of otolaryngology head & neck surgery, 127, 877-883.

[8] Jakobson, R., Fant, G.M.C. & M. Halle. (1952). Preliminaries to speech analysis: the distinctive features and their correlates. MIT press.

[9] Chomsky, N. & M. Halle. (1968). *The sound pattern of English*. MIT Press.

[10] Chao, LY., Verdonck-de Leeuw, IM., ten Bosch, L. & T. Rietveld. (2007). Akoestisch-fonetische analyses van spraak van patiënten behandeld voor een tumour in de mondholte of de orofarynx. MA-thesis, Radboud Universiteit Nijmegen.

[11] Borggreven, P., I. Verdonck-de Leeuw, J. Langendijk, P. Doornaert, M. Koster, R. de Bree & C. Leemans. (2005). Speech outcome after surgical treatment for oral and oropharyngeal cancer: a longitudinal assessment of patients reconstructed by a microvascular flap. Head & Neck, 9, 785-793.

[12] Aaronson NK, Ahmedzai S, Bergman B, Bullinger, M., Cull, A., Duez, NJ., Filiberti, A., Flechtner, H., Fleishman, SB., de Haes, JCJM., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, PB., Schraub, S., Sneeuw, K., Sullivan, M., & F. Takeda. (1993). The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. J Natl Cancer Inst;85:365-76.

[13] Yoshida, H., Furuya, Y., Shimodaira, K., Kanazawa, T., Kataoka, R. & K. Takahashi. (2000). Spectral characteristics of hypernasality in maxillectomy patients. Journal of Oral Rehabilitaion, 27, 723-730.

[14] Sumita, Y., S. Ozawa, H. Mukohyama, T. Ueno & T. Ohyama. (2002). Digital acoustic analysis of five vowels in maxillectomy patients. Journal of Oral Rehabilitation, 29, 649-656.

[15] Whitehill, T., V. Ciocca, J. Chan & N. Samman. (2006). Acoustic analysis of vowels following glossectomy. Clinical Linguistics & Phonetics, 20, 135-140.