European Annals of Otorhinolaryngology, Head and Neck diseases 136 (2019) 355–359

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

Original article

Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer

M. Balaguer[a,*], A. Boisguérin[a], A. Galtier[a,b], N. Gaillard[a], M. Puech[a,b], V. Woisard[a,b,c]

[a] CHU de Larrey, 24, chemin de Pouvourville, TSA 30030, 31059 Toulouse cedex 9, France
[b] Unité d'oncoréhabilitation – oncopole, 1, avenue Irène Joliot-Curie, 31059 Toulouse cedex 9, France
[c] Laboratoire Octogone-Lordat, URI Octogone-Lordat (EA4156), maison de la recherche, université de Toulouse II Jean-Jaurès, 5, allée Antonio-Machado, 31058 Toulouse cedex 9, France

ARTICLE INFO

Keywords:
Speech
Oral cavity cancer
Oropharyngeal cancer
Assessment

ABSTRACT

Background: Perceptual evaluation is a means of assessing speech disorder severity in clinical practice. Although limited in reliability and reproducibility, its ease of application makes it very widely used. Choice of assessment criteria and type of speech sample are key points.
Objective: To compare a panel's perceptual evaluations on two tasks with different criteria.
Material and method: The corpus comprised 87 samples from patients treated for oral cavity or oropharynx cancer, assessed by 6 experts on two criteria (impairment of intelligibility and of speech signal) and two kinds of speech sample (semi-spontaneous versus reading speech)
Results: Although strong correlations were found between tasks ($r > 0.8$), the speech signal criterion gave a score distribution providing a better metric. Severity was greater in oral cavity (mean, 5.44 ± 2.47) than oropharyngeal cancer (6.46 ± 2.24). Semi-spontaneous speech tended to show less severity score ceiling effect than reading speech (mean, 6.06/10 for picture description and 6.51/10 for reading).
Conclusion: Speech signal impairment in semi-spontaneous speech seems to be the best clinical measure to assess speech disorder following treatment of oral cavity or oropharynx cancer.

1. Introduction

The difficulty of making a quantitative assessment of speech sequelae after head and neck cancer hinders the integration of functional risk in treatment protocols, while recent developments in telecommunication have made oral language essential in interpersonal communication.

Head and neck is the fifth most frequent cancer region in France (HAS 2009), and often impacts oral communication capacity, impairing the vocal signal or articulation and thus quality of life [1].

Intelligibility is essential to communication, and impairment diminishes the capacity to interact [2]. It represents the degree to which the speaker's message can be understood by the listener: "proportion of speech understood" or, in the assessment context, "proportion of words correctly transcribed" [3–5]. It is thus a matter of perception by others, which makes it difficult to quantify, due to human compensation strategies enhancing signal decoding by cues relying on language structure and meaning [6].

Speech disorder severity is therefore an alternative index in clinical contexts. It is based on degree of intelligibility impairment associated to other speech signal variables such as acoustic-phonetic code emission quality, speech speed and other temporal and/or prosodic parameters relevant to the perceived difficulty [7–9]. Disorder severity is a broader concept, including intelligibility plus compensation strategies.

Speech disorder assessment by speech therapists is based on perceptual evaluation. One of the most widely used instruments in France is the BECD clinical dysarthria assessment battery (*Batterie d'Évaluation Clinique de la Dysarthrie*) [10]. Perceptual assessment may be made at two levels [11]:

- phoneme emission: is the phoneme perceived by the listener the intended one? This concerns acoustic-phonetic coding, or intelligibility;
- oral communication, or "running speech intelligibility", which involves top-down mechanisms.

* Corresponding author.
E-mail address: balaguer.m@chu-toulouse.fr (M. Balaguer).

https://doi.org/10.1016/j.anorl.2019.05.012
1879-7296/

Speech assessment needs to be two-fold: qualitative, to determine the mechanisms underlying the impairment, and quantitative, to score the degree of impairment. Thus two strategies may be used: a grading scale such as a visual analogue scale (from perceived severe disorder to no disorder), or measurements (e.g., rate of error between intended and perceived phonemes).

All such perceptual tests, however, have a drawback: reliability and reproducibility are listener-dependent [11]. The listener's familiarity with the patient or with the task increases predictability, and results may differ between panels of experts and of naïve observers. Reproducibility is likewise subject to variation. Even so, these are the most widely used clinical tests, mainly for reasons of ease of use.

We therefore sought to identify the most useful tests, feasible in everyday clinical practice, for perceptual evaluation of speech to provide reference values for progression assessment.

The aim was to compare results on perceptual tests of disorder severity and intelligibility impairment between two situations: reading, and semi-spontaneous speech.

The study hypothesis was that severity assessed on semi-spontaneous picture description is more clinically relevant than intelligibility assessed on reading.

2. Material and methods

2.1. Corpus

The study protocol was designed as part of the Carcinologic Speech Severity Index (C2SI) project [12], the aim of which is to measure the impact of head and neck (oral and pharyngeal) cancer treatment on speech production by automated speech processing compared to perceptual methods.

The corpus was built up from patients seen in follow-up after oral cavity or oropharyngeal cancer treatment in 2015 and 2016 at the Institut Universitaire du Cancer Oncopole cancer center in Toulouse, France.

Inclusion criteria comprised: 6 months post-treatment, and clinical remission so that speech disorder would be as stable as possible, whether perceptible to the naked ear or not (so as to include the mildest deficits). The study was thus conducted in a context of chronic speech disorder.

Exclusion criteria comprised speech disorder potentially associated with some other pathology such as stroke or fluency disorder (stammering).

2.2. Tests

Patients' speech was recorded on a reading task (READ) and a picture description task (DESC).

All recordings were made in the oncorehabilitation unit of the Institut Universitaire du Cancer Toulouse – Oncopole. During follow-up consultation, patients entered a soundproof recording booth for speech assessment tasks. Audio files were recorded in WAVE format on a digital recorder with microphone and popshield filter, to optimizer recording quality and minimize measurement bias (misassessment due to recording issues).

2.2.1. Reading task (READ)

Patients read aloud the first paragraph of La chèvre de Monsieur Seguin by Alphone Daudet:

> "Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon. Un beau matin, elles cassaient leur corde, s'en allaient dans la montagne, et là-haut le loup les mangeait. Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait. C'était, paraît-il, des chèvres indépendantes voulant à tout prix le grand air et la liberté".

This text was chosen for its length and balanced phonetic makeup; it is widely used in clinical assessment in France [6].

2.2.2. Picture description task (DESC)

Patients randomly chose one out of several photographs, all showing a sea-front with boats.

They were asked to describe the picture so that a listener some distance away could copy it. The description task lasted at least 1 minute, to give the listening panel sufficient material.

2.2.2.1. Listening panel. The completed corpus of recordings was analyzed by a listening panel comprising 6 speech therapists, working in the ENT department and thus considered to be experts.

2.3. DESC task analysis

Individual recordings were rendered anonymous and presented to the panel, who listened to them via headphones in a calm setting.

Panelists were informed that they would hear recordings of subjects describing a photograph related to the sea. After listening, they were to make two assessments:

- impairment of intelligibility and of speech signal, on a 0–10 scale (the lower the score, the greater the perceived disorder) by unit steps (discrete scoring). The two terms were defined in the instructions as follows: "Intelligibility is the comprehensibility of the message carried by the signal; signal impairment severity is the degree of overall vocal signal impairment.";
- the degree of difficulty of understanding the information given by the speaker, on a visual analogue scale (VAS), placing a bar along a line going from "I understood nothing" to "I understood everything", via "I understood some words but not the message" and "I understood the message as a whole but not all of the words". From this, a continuous quantitative score was derived.

Thus, on the DESC task, there were 3 global scores per patient: signal impairment ("sig"), intelligibility ("intell"), and degree of difficulty of understanding ("intelldeg").

2.4. READ task analysis

The reading task was analyzed 6 months after the picture description task. The perceptual assessment procedure was similar except for the absence of "Inteldeg" assessment on VAS.

For the READ task, there were thus 2 scores per panel per patient; signal impairment and intelligibility.

2.4.1. Data analysis

Description of distributions, correlations between variables and differences in score shed light on the metric quality of the various parameters.

Population characteristics were described. Perceptual assessment results were panel-dependent, and inter-observer reliability was assessed on interclass correlation coefficients (ICC), corresponding to the proportion of variance associated with inter-patient variability free of measurement error: i.e., the proportion of variance accounted for by a given factor. In the present case, ICC assessed the proportion of variance in perceptual score due to inter-observer variability compared to overall variation in perceptual score: ICC = (inter-observer variation)/(inter-observer variation + random error). Thus, the higher the ICC, the lower the random error and the more homogeneous the inter-observer scores. Normal values for interpretation purposes have been

Table 1
Interclass correlation coefficients per test.

	ICC	95% confidence interval
Signal impairment (DESC)	0.69	0.62–0.77
Intelligibility (DESC)	0.77	0.70–0.83
Degree of intelligibility (DESC)	0.74	0.67–0.82
Signal impairment (READ)	0.69	0.61–0.77
Intelligibility (READ)	0.43	0.32–0.55

reported, indicating moderate reliability for ICC values between 0.5 and 0.75, and good reliability for values > 0.75 [13].

Panel scores were reported as mean per panel. Normal distribution was checked on Shapiro-Wilk test, and residual homoscedasticity on Breusch-Pagan test. The various test results were reported on box-plots. Comparisons and correlations used non-parametric tests as appropriate: Wilcoxon and Spearman correlation tests, respectively. Bonferroni correction was applied on multiple tests, to take first-order risk inflation into account.

Analysis used Stata 14.2 software.

3. Results

3.1. Population

Eighty-seven patients treated for oral cavity or oropharyngeal cancer had voice recordings: 51 male, 36 female; mean age, 65.8 years (range, 36–87 years).

Thirty-five patients (40.2%) had oral cavity cancer: 16 oral floor, 8 tongue, 6 retromolar and 5 mandibular. Fifty-two (59.8%) had oropharyngeal cancer: 26 tonsillar, 13 tongue-base, 4 velar and 9 unspecified oropharyngeal.

Tumor stages were predominated by T2 and T4: respectively, 33 and 31 (38% and 35%). Twelve were T3 (14%), and 11 T1 (13%) (The higher the T stage on the TNM classification, the greater the tumor volume).

Seventy-three patients (83.9%) underwent surgery with associated radiation therapy, chemotherapy or neck dissection. Fourteen patients received either exclusive radiation therapy ($n = 5$) or chemoradiotherapy ($n = 9$).

Mean treatment-to-recording time was 64.2 months (range, 6–239; median, 39).

3.2. Inter-observer reliability

Table 1 shows ICC values for perceptual evaluations.

Agreement was moderate to good (>0.69) on all tasks except reading intelligibility (0.43).

The panel of 6 speech therapists was thus homogeneous.

3.3. Results per score

Table 2 shows mean score results.

Table 3 shows mean scores according to tumor size and location.

Mean scores were lower for signal impairment than for intelligibility. There was also a tumor size effect, with lower scores on all tests in advanced (T3, T4) than early tumor (T1, T2: Dwivedi, 2012). Scores were higher in oropharyngeal than oral cavity cancer.

Distribution analysis showed a ceiling effect for no or slight disorder, which was stronger on intelligibility than signal impairment scores. Scatter was greater for signal impairment than intelligibility (Fig. 1).

Results for "degree of intelligibility" (on picture description, with continuous values) and for "intelligibility" (discrete values from 0 to 10) were similar. Even so, comparison of means for matched series on non-parametric Wilcoxon test showed a significant difference ($P = 0.001$), with a higher mean value for degree of intelligibility than for intelligibility (7.96 vs. 7.61).

3.4. Correlations between mean signal impairment and intelligibility scores

Correlations between scores were very high, with Spearman rho systematically > 0.72, and P-values systematically < 0.001 after Bonferroni correction.

For a given task, correlations between signal impairment and intelligibility were at least 0.90 (DESC: 0.91; READ: 0.90).

Signal impairment on READ and DESC showed 0.89 correlation, and intelligibility 0.88.

On different tasks, correlations between signal impairment and intelligibility were between 0.80 and 0.87.

Degree of intelligibility on picture description showed 0.81 correlation with signal impairment on description and 0.89 with intelligibility on description. Correlations were slightly lower but still strong between intelligibility and signal impairment on reading (between 0.73 and 0.74).

3.5. Comparison between READ and DESC task scores

Although signal impairment scores were close in absolute terms between the two tasks (Fig. 2), the differences were significantly different.

Mean scores were significantly higher on READ than on DESC, both for signal impairment ($P = 0.014$) and for intelligibility ($P < 0.001$): 6.51 vs. 6.06 and 8.69 vs. 7.61, respectively.

4. Discussion

There have been few studies of intelligibility and speech disorder severity in oral cavity and oropharyngeal cancer [14,15]. The present results confirm impact on speech production, notably related to initial tumor size and location: oral cavity location and T4 stage were associated with greater signal impairment and poorer intelligibility. However, data are still lacking to predict treatment impact, and notably on speech-related quality of life.

Regarding speech production disorder severity assessment in clinical practice, the present results showed that all test results, whether reading or semi-spontaneous, correlated in terms both of signal impairment and of intelligibility.

On the other hand, inter-observer reliability was a limitation, better for picture description than reading, and better for

Table 2
Scores attributed by panels.

	Mean	Median	Minimum	Maximum	Interquartile range
Signal impairment (DESC)	6.06	6.33	0.58	9.67	4.2–8
Intelligibility (DESC)	7.61	8.67	1.2	10	6.83–9.5
Degree of intelligibility (DESC)	7.96	9.26	1.16	10	6.7–9.9
Signal impairment (READ)	6.51	7.02	1.03	10	5.03–8.1
Intelligibility (READ)	8.69	9	3.5	10	8.4–9.5

Standard deviation not shown, as none of the distributions were normal: P-value (Shapiro-Wilk test) < 0.01.

Table 3
Mean scores according to tumor size and location, with standard deviations in brackets.

	T stage (TNM classification)				Tumor location	
	T1	T2	T3	T4	OC	OP
Signal impairment (DESC)	6.77 (1.71)	6.95 (2.01)	5.62 (2.83)	5.02 (2.39)	5.44 (2.47)	6.46 (2.24)
Intelligibility (DESC)	8.71 (1.03)	8.55 (1.85)	6.95 (3.14)	6.47 (2.71)	6.83 (2.70)	8.12 (2.23)
Degree of intelligibility (DESC)	9.13 (1.04)	9.05 (1.40)	7.62 (3.29)	6.49 (2.86)	7.35 (2.78)	8.36 (2.31)
Signal impairment (READ)	6.67 (1.79)	7.53 (1.69)	5.88 (2.64)	5.62 (2.16)	6.10 (2.31)	6.77 (2.03)
Intelligibility (READ)	8.99 (0.56)	9.23 (0.60)	8.15 (1.56)	8.21 (1.71)	8.40 (1.60)	8.87 (1.10)

T: T stage on TNM classification; OC: oral cavity; OP: oropharynx.

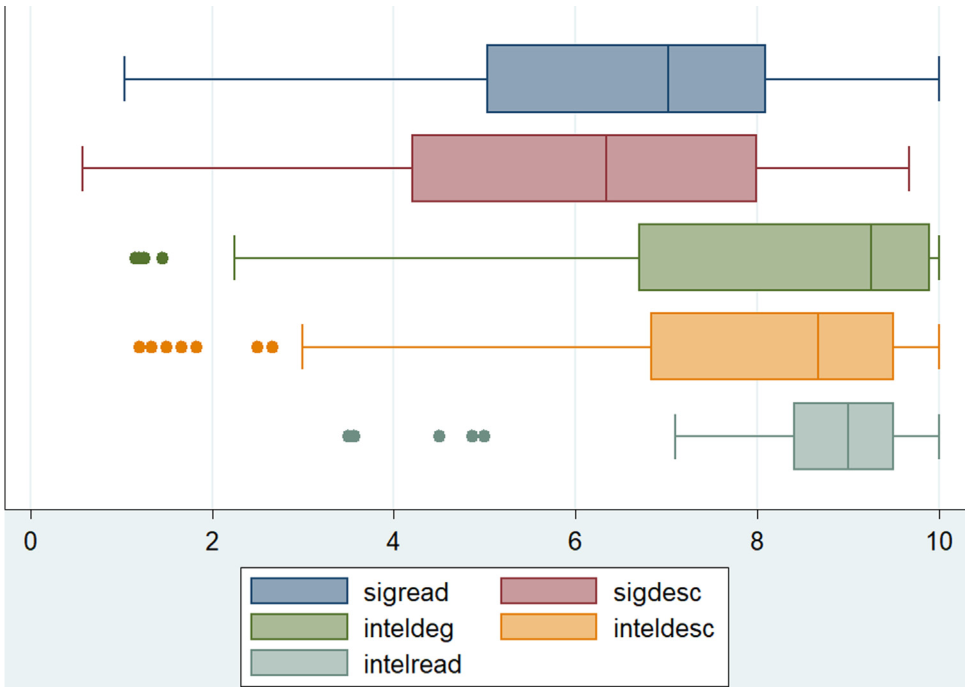

Fig. 1. Box-plot of scores. Sigread: signal impairment on reading; sigdesc: signal impairment on picture description; inteldeg: degree of intelligibility on picture description on visual analog scale; inteldesc: intelligibility on picture description; intelread: intelligibility on reading.

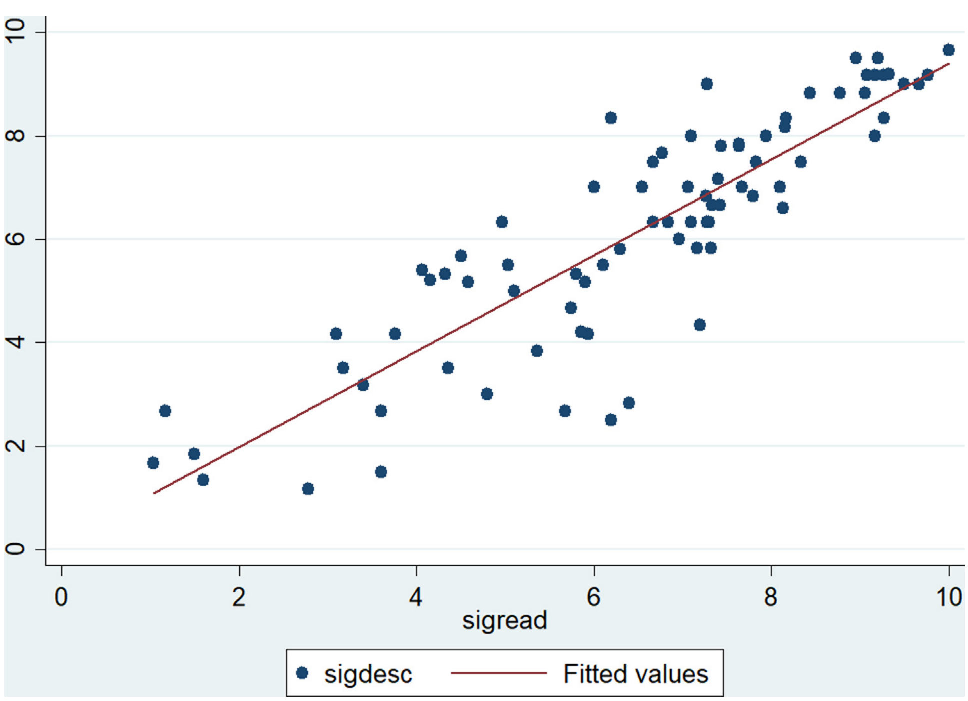

Fig. 2. Scatter diagram of results for signal impairment on picture description (sigdesc) versus signal impairment on reading (sigread): $r = 0.89$ ($P < 0.001$).

intelligibility than signal impairment severity on picture description (0.77 vs. 0.69).

That mean scatter was wider for signal impairment than intelligibility, whether on discrete or continuous variables, argued for use of signal impairment assessment.

The picture description test involved semi-spontaneous speech, directed only by the picture and the instructions, and gave perceptual scores with less ceiling effect for absence of disorder than the reading test. Mean scores were lower for semi-spontaneous speech, as seen in Fig. 1, and scatter was less.

The use of a graduated intelligibility scale, unlike in previous studies by the same team in a population with more varied pathology including neurologic dysarthria, has not been validated [16]; the ceiling effect was greater, with less scatter.

Thus, the present results as a whole confirmed the study hypothesis: assessing speech disorder severity from semi-spontaneous speech based on picture description is more clinically contributive than from reading.

Although the correlations between the various perceptual scores for head and neck cancer patients showed that their mean values all represent the same information, intelligibility showed a ceiling effect, increasing the scores by more than 1.5 points compared to signal impairment scores. Inter-observer reliability was better for intelligibility on the picture description task, but this was the only exception to the general superiority of measuring signal impairment rather than intelligibility in description tests. The higher ICCs associated with intelligibility assessments may be due to the ceiling effect, with panelists tending more readily to give a maximum score of 10/10. Moreover, an ICC value of 0.69 for perceptual scores of speech signal impairment on picture description is enough for this parameter to be chosen [17].

5. Conclusion

Speech disorder is a frequent sequela in oral cavity and oropharyngeal cancer. Measurement is a challenge to be faced in seeking to improve management. The instruments used in everyday clinical practice are inherently subjective, depending on a listener's perception in assessing speech production deficit. The present results are in favor of using a measurement of signal impairment in a semi-spontaneous speech task, with severity defined as the degree of overall impairment of the vocal signal and the speech sample being based on a photograph randomly chosen from a set of 10 pictures all showing sea-fronts with one or several boats. The inter-observer reliability and metric properties of this parameter are compatible with clinical application in consultation for persons with experience of this kind of disorder. Its limitations, however, argue for developing a disorder index based on automated speech analysis, to avoid intra- and inter-individual measurement variability.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgments

The present study was part of the Carcinologic Speech Severity Index (C2SI) project, funded by the Institut National du Cancer (funding n° 2014-135) as a free research project in Human and Social Science, Epidemiology and Public Health. The principal investigator is Virginie Woisard (Hôpital Larrey, Toulouse University Hospital).

We thank the medical and paramedical teams of the Institut Universitaire du Cancer Oncopole de Toulouse for recordings, and the listening panelists and patients who took part, and the Institut de Recherche en Informatique de Toulouse (Toulouse Information Technology Research Institute, Paul Sabatier Toulouse III University) which was also involved.

References

[1] Mlynarek AM, Rieger JM, Harris JR, O'Connell DA, Al-Qahtani KH, Ansari K, et al. Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. J Otolaryngol Head Neck Surg 2008;37(1):2–10.
[2] Lindblom B. On the communication process: speaker-listener interaction and the development of speech. Augment Altern Commun 1990;6(4):220–30.
[3] Kent RD, Weismer G, Kent JF, et Rosenbeck J. Toward phonetic intelligibility testing in dysarthria. J Speech Hear Disord 1989;54(4):482–99.
[4] Keintz CK, Bunton K, et Hoit JD. Influence of visual information on the intelligibility of dysarthric speech. Am J Speech Lang Pathol 2007;16(3):222–34.
[5] Hustad KC. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. J Speech Lang Hear Res 2008;51(3):562–73.
[6] Ghio A, Pouchoulin G, Teston B, Pinto S, Fredouille C, De Looze C, et al. How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? Speech Commun 2012;54:664–79.
[7] Kent R. Intelligibility in speech disorders. Amsterdam/Philadelphia: John Benjamins; 1992.
[8] Yorkston KM, Beukelman DR, Tice TR. Sentence intelligibility test. Lincoln, NE: Tice technology services; 1996.
[9] Auzou P. Les objectifs du bilan de la dysarthrie. In: Auzou P, Rolland Monnoury V, Pinto S, Öszancak C, editors. Les dysarthries. Marseille: Solal; 2007. p. 189–95.
[10] Auzou P, Rolland-Monnoury V. Batterie d'évaluation de la dysarthrie. 1st ed. Isbergues: Ortho Ed.; 2006.
[11] Middag C. Automatic analysis of pathological speech. Ghent, Belgium: Ghent University, Department of Electronics and information systems; 2012.
[12] Astesano C, Balaguer M, Farinas J, Fredouille C, Gaillard P, Ghio A, et al. Carcinologic speech severity index project: a database of speech disorders productions to assess quality of life related to speech after cancer. Miyazaki Japan: LREC; 2018.
[13] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63, http://dx.doi.org/10.1016/j.jcm.2016.02.012.
[14] Plisson L, Pillot-Loiseau C, Crevier-Buchman L. Intelligibilité de la parole après le traitement d'un cancer de l'oropharynx : étude descriptive chez sept patients en pré-traitement et en post-traitement précoce; 2017 [7e Journées de phonétique clinique (JPC7), Jun 2017, Paris, France].
[15] Breil L, Crépin H, Smadja M, Crevier-Buchman L. Quality of life after oropharyngectomy. Rev Laryngol Otol Rhinol (Bord) 2006;127(5):305–14.
[16] Dittner J, Lepage B, Woisard V, Kergadallan M, Boisteaux K, Robert E, et al. Elaboration et validation d'un test quantitatif d'intelligibilité pour les troubles pathologiques de la production de parole. Rev Laryngol Otol Rhinol 2010;131(1):9–14.
[17] Elie C, Colombet I. Méthodes d'estimation de la reproductibilité. Note méthodologique. Sang Thrombose Vaiss 2011;23(3):138–45.