Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review

Mathieu Balaguer, Timothy Pommée, Jérôme Farinas, Julien Pinquier, Virginie Woisard, Renée Speyer

Title page:

Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: A systematic review

Authors:
Mathieu Balaguer[1,2], Speech and Language Pathologist, PhD Student,
Timothy Pommée[2], Speech and Language Pathologist, PhD Student,
Jérôme Farinas[2], PhD,
Julien Pinquier[2], PhD,
Virginie Woisard[1,3], MD, PhD,
Renée Speyer[4], Speech and Language Pathologist, Epidemiologist, PhD

(1) CHU Larrey, Toulouse, France
(2) IRIT, CNRS, Université Paul Sabatier, Toulouse, France
(3) Laboratoire Octogone-Lordat, Toulouse, France
(4) Faculty of Educational Sciences, University of Oslo, Oslo, Norway

Address of correspondence:

IRIT Institut de Recherche en Informatique de Toulouse
118 Route de Narbonne
31062 TOULOUSE CEDEX 9

Corresponding author:

Mathieu Balaguer
ORCID : 0000-0003-1311-4501
mathieu.balaguer@univ-tlse3.fr

Compliance with Ethical Standards:

Funding: This study was funded by the Hospitals of Toulouse, and Grant ANR-18-CE45-0008 from The French National Research Agency in 2018 RUGBI project "Improving the measurement of intelligibility of pathological production disorders impaired speech" led by Jérôme Farinas at IRIT.

Conflict of Interest: All authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Keywords: speech; intelligibility; acoustics; head and neck cancer; oncology

ABSTRACT

Background: The development of automatic tools based on acoustic analysis allows to overcome the limitations of perceptual assessment for head and neck cancer patients. The aim of this study is to provide a systematic review of literature describing the effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis.

Methods: Two databases (PubMed and Embase) were surveyed. The selection process, according to the PRISMA statement, led to a final set of 22 articles.

Results: Nasalance is studied mainly in oropharyngeal patients. The vowels are mostly studied using formant analysis and vowel space area, the consonants by means of spectral moments with specific parameters according to their phonetic characteristic. Machine learning methods allow classifying “intelligible” or “unintelligible” speech for T3 or T4 tumors.

Conclusions: The development of comprehensive models combining different acoustic measures would allow a better consideration of the functional impact of the speech disorder.

INTRODUCTION

Head and Neck Cancer (HNC) has major functional repercussions on the upper aerodigestive tract (UAT) (breathing, swallowing and phonation/speech). Because of the sensory-motor impairment related to the presence of the tumor in the anatomical regions involved in the articulation of the speech, a functional impairment at the level of communication is likely to appear [1]. The speech-related quality of life will also be impacted [2, 3].

In this oncological context, various factors can affect the quality of speech, including the treatments, the size of the tumor [4, 5, 6] or its location [7, 8].

With the increasing rate of oropharyngeal cancer incidence [9, 10], the evaluation of speech and its disorders becomes a major issue in the management of patients with HNC.

This evaluation is mainly based on a perceptual assessment: therapists, mainly speech pathologists, assess the quality of the patient's speech production. But these methods have two major limitations. First, most of the tools are intended for voice quality assessment in laryngeal cancers [11], whereas speech disorder is the most common symptom in cancers of the oral cavity and the oropharynx [12]. Second, these measures are known to show great inter- and intra-judge variability. Indeed, the reliability of the perceptual estimates is mostly listener-dependent [13]. The degree of familiarity of the listener with the patient or with the task might increase predictability and improve the functional speech scores given by the rater. The rating by an expert in the pathology field or by a rater that is familiar with the patient can be very different of that by a naive listener. Moreover, the reproducibility of the perceptual assessment is also subject to intra-judge variations. The emotional context or the mental alertness of the judge at the time of the assessment may influence the outcome [14].

Recently, the technology development allows investigating new tools for speech evaluation, based on objective data [15]. For this purpose, acoustic speech analysis is currently a growing field of research.

Review question: The aim of this paper is to provide a systematic review of literature describing the effects of head and neck cancer on speech intelligibility using acoustic analysis. This review will focus on speech intelligibility in adults with oral or oropharyngeal cancer assessed by acoustic measures.

MATERIAL AND METHODS

Protocol and registration

The methodology and reporting on this systematic review were guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and checklist. The PRISMA statement and checklist is designed to guide researchers in the essential and transparent reporting of systematic reviews [16, 17].

Eligibility criteria

To be eligible for inclusion in this systematic review, articles were required to describe the effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis.
Only articles with the following criteria were included:

- Assessment of speech intelligibility,
- Use of acoustics and related terms (such as acoustic analysis, phonetics, signal processing, sound spectrography…),
- Oral or oropharyngeal cancer patients.

In this study, speech intelligibility is defined as the level which a message can be understood by a listener [18], the proportion of understood speech [19], or the correctly transcribed word rate [20]. Speech intelligibility impairment is described as the functional speech deficit decreasing the ability to interact with someone else [21].
Exclusion criteria were:

- The absence of the original larynx (exclusion of total or partial laryngectomies, larynx prostheses…),
- Studies addressing children populations,
- Papers that were not original articles, such as abstracts, conference proceedings and reviews,
- Case studies,
- Articles not published in English.

Data sources and search strategies

A literature search was performed in two different electronic databases, to gather relevant literature: PubMed and Embase. These two databases were selected based on the subject of this research. Note that a third database, Web of Science, did not retrieve any new reference.
All publications dated up to December 4, 2018, were included, with no limitations regarding the publication dates.

The search terms are listed in Table 1.

Table 1: Database and Search Terms (Subject Headings and Free Text Words)

Database	**Search Terms (subject headings and free text words)**	**Number of records**
PubMed:	(("Speech"[Mesh] OR "Speech Sound Disorder"[Mesh] OR "Speech Disorders"[Mesh] OR "Articulation Disorders"[Mesh] OR "Voice"[Mesh] OR "Voice Quality"[Mesh] OR "Voice Disorders"[Mesh] OR "Hoarseness"[Mesh] OR "Aphonia"[Mesh] OR "Dysphonia"[Mesh] OR "Phonation"[Mesh]] OR "Speech Intelligibility"[Mesh]) OR (intelligibil*[Title/Abstract] OR Comprehensibil*[Title/Abstract] OR understandabil*[Title/Abstract])) AND ("Acoustics"[Mesh] OR "Speech Acoustics"[Mesh] OR "Speech Production Measurement"[Mesh] OR "Phonetics"[Mesh] OR "Signal Processing, Computer-Assisted"[Mesh] OR "Fourier Analysis"[Mesh] OR "Sound Spectrography"[Mesh] OR "Sound"[Mesh] OR "Signal-To-Noise Ratio"[Mesh] OR "Noise"[Mesh]) AND ("Pharyngeal Neoplasms"[Mesh] OR "Mouth Neoplasms"[Mesh] OR "Oropharyngeal Neoplasms"[Mesh] OR "Facial Neoplasms"[Mesh] OR "Head and Neck Neoplasms"[Mesh] OR "Laryngeal Neoplasms"[Mesh] OR "Hypopharyngeal Neoplasms"[Mesh])	296
Embase:	((speech/ OR speech sound disorder/ OR voice/ OR dysphonia/ OR aphonia/ OR voice disorder/ OR hoarseness/ OR phonation/ OR speech intelligibility/) OR (Intelligibil*.ab. OR Intelligibil*.ti. OR comprehensibil*.ab. OR comprehensibil*.ti. OR understandabil*.ab. OR understandabil*.ti.)) AND (voice analysis/ OR voice onset time/ OR voice parameter/ OR acoustics/ OR speech analysis/ OR acoustic analysis/ OR sound analysis/ OR phonetics/ OR signal processing/ OR fourier analysis/ OR sound detection/ OR sound/ OR frequency/ OR frequency analysis/ OR pitch/ OR noise/ OR signal noise ratio/) AND ("head and neck cancer"/ OR "head and neck tumor"/ OR oropharynx tumor/ OR pharynx tumor/ OR oropharynx cancer/ OR oropharynx carcinoma/ OR pharynx carcinoma/ OR oropharynx squamous cell carcinoma/ OR pharynx cancer/ OR pharynx tumor/ OR mouth cancer/ OR mouth tumor/ OR salivary gland tumor/ OR tongue tumor/ OR tonsil tumor/ OR mouth carcinoma/ OR "head and neck carcinoma"/ OR mouth squamous cell carcinoma/ OR salivary gland carcinoma/ OR tongue carcinoma/ OR tonsil carcinoma/ OR face tumor/ OR face cancer/ OR larynx cancer/ OR larynx tumor/ OR larynx carcinoma/ OR hypopharynx cancer/ OR hypopharynx tumor/ OR hypopharynx carcinoma/ OR hypopharynx squamous cell carcinoma/)	262

All abstracts were reviewed by two independent raters. Differences of opinion about the eligibility of articles were settled by consensus. A flowchart of the selection process according to PRISMA [16] is shown in Figure 1.

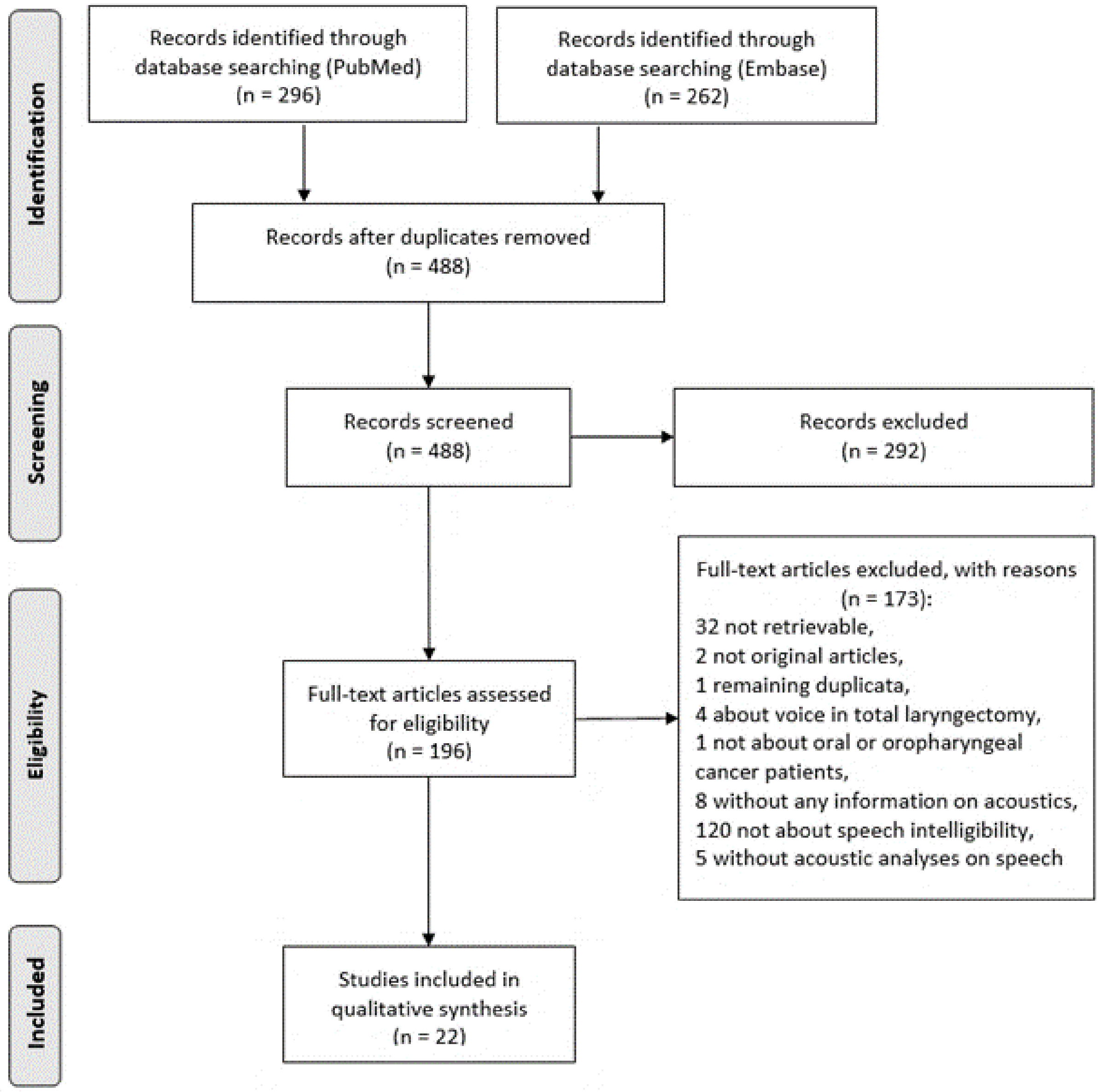

Figure 1: Flow diagram of the review process according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Adapted from Moher and al., 2009 [17]

Methodological quality and level of evidence

The National Health and Medical Research Council (NHMRC) Evidence Hierarchy was used to assess the level of evidence, from I ("Systematic reviews") to IV ("Case series") [22]. The Qualsyst critical appraisal tool by Kmet et al. [23] provides systematic, reproducible and quantitative means of assessing the methodological quality of research over a broad range of study designs. A Qualsyst score higher than 80% was interpreted as strong quality, 60–79% as good quality, 50–59% as adequate quality, and lower than 50% as poor methodological quality. Studies with poor methodological quality were excluded from further analysis.

Data extraction

After assessment of methodological quality, data from all remaining articles were extracted for the following categories: number of participants in the study and their characteristics (age, diagnosis and language spoken), acoustic parameters (and their definitions), comparison criterion/a, speech sample, and authors' main conclusions.

Additionally, geographic bibliometric data was extracted using the Netscity tool[1] (by the Netscience project of the Labex SMS, Toulouse, France).

Results

Study selection

A total of 488 records were retrieved from the 2 electronic databases. Two independent reviewers screened all records and assessed 196 full-text articles for eligibility. A final total of 22 articles met the inclusion criteria and were included in this review (see Figure 1 above).

Quality assessment

The overall quality of the studies, as assessed by the QualSyst tool, ranged from “good” to “strong”, with four studies ranked as “good” and 18 as “strong”.
Based on the NHMRC Evidence Hierarchy [23], 20 studies were classified as level III evidence (14 as III-3: “Comparative studies with 2 or more single-arm studies”; six as III-2: “Comparative studies with concurrent controls and allocation not randomized (cohort studies), or case control studies”), and two as level IV evidence (“Case series”). No article of a low level of evidence had to be excluded. The ratings of all 22 included articles are listed in Table 2.

[1] https://www.geotests.net/netscitypg/index.php

Table 2: Level of evidence and methodological quality ratings for the 22 included articles using the Qualsyst critical appraisal tool by Kmet et al. [23] *and National Health and Medical Research Council (NHMRC) level* [22]

Reference	Qualsyst score * (%)	Methodology quality	NHMRC Level of evidence †
[24] **Acher & Fougeron, 2014**	15/20 (75%)	Good	IV
[25] **Chung et al., 2011**	20/22 (91%)	Strong	III-3
[26] **De Bruijn et al., 2009**	21/24 (88%)	Strong	III-2
[27] **De Bruijn et al., 2011**	20/22 (91%)	Strong	III-2
[28] **de Carvalho-Teles, Ubijara Sennes, & Gielow, 2008**	18/22 (82%)	Strong	III-3
[29] **Dwivedi et al., 2016**	24/24 (100%)	Strong	III-2
[30] **Fang, Li, Ma, & Zhang, 2017**	14/20 (70%)	Good	III-2
[31] **Ha et al., 2016**	20/22 (91%)	Strong	III-2
[32] **Jacobi, Rossum, Molen, Hilgers, & Brekel, 2013**	19/20 (95%)	Strong	III-3
[33] **Kazi et al., 2007**	20/20 (100%)	Strong	III-2
[34] **Kim, Rao, & Clements, 2014**	19/22 (86%)	Strong	III-3
[35] **Knuuttila, Pukander, & Ma, 1999**	16/20 (80%)	Strong	III-3
[36] **Kraaijenga & Molen, 2014**	19/22 (86%)	Strong	III-3
[37] **Kumar, Jain, Thakar, & Aggarwal, 2013**	16/22 (73%)	Good	III-3
[38] **Laaksonen, Rieger, Harris, & Seikaly, 2011**	21/22 (95%)	Strong	III-3
[39] **Laaksonen, Rieger, Happonen, Harris, & Seikaly, 2010**	17/20 (85%)	Strong	III-3
[40] **Markkanen-leppa et al., 2005**	22/22 (100%)	Strong	III-3
[41] **Moerman, Vermeersch, Lierde, Fahimi, & Van Cauwenberge, 2003**	14/22 (64%)	Good	IV
[42] **Seikaly et al., 2003**	17/20 (85%)	Strong	III-3
[43] **Takatsu, Hanai, & Suzuki, 2016**	19/20 (95%)	Strong	III-3
[44] **Wakumoto et al., 1996**	18/20 (90%)	Strong	III-3
[45] **Yoshida et al., 2000**	19/20 (95%)	Strong	III-3

* Methodological quality: strong > 80%; good 60–79%; adequate 50–59%; poor < 50%.

† NHMRC hierarchy: Level 1 Systematic reviews; Level II Randomized control trials; Level III–1 Pseudo-randomized control trials; Level III–2 Comparative studies with concurrent controls and allocation not randomized (cohort studies), case control studies, or interrupted time series with a control group; Level III–3 Comparative studies with historical control, 2 or more single-arm studies, or interrupted time series without a control group; Level IV Case series.

The full outcome table on the 22 retained articles can be found in Appendix A.

Bibliometric data

The field of acoustic parameters in speech analysis in patients treated for head and neck cancer mainly concerns teams located in three geographical areas: Western Europe (mainly the Netherlands), North America and the Far East (Japan and South Korea). Some collaborations between teams are noted: between Finland and Canada, and between South Korea and the United States (see Figure 2).
This will have an influence on the languages of the study speech samples.

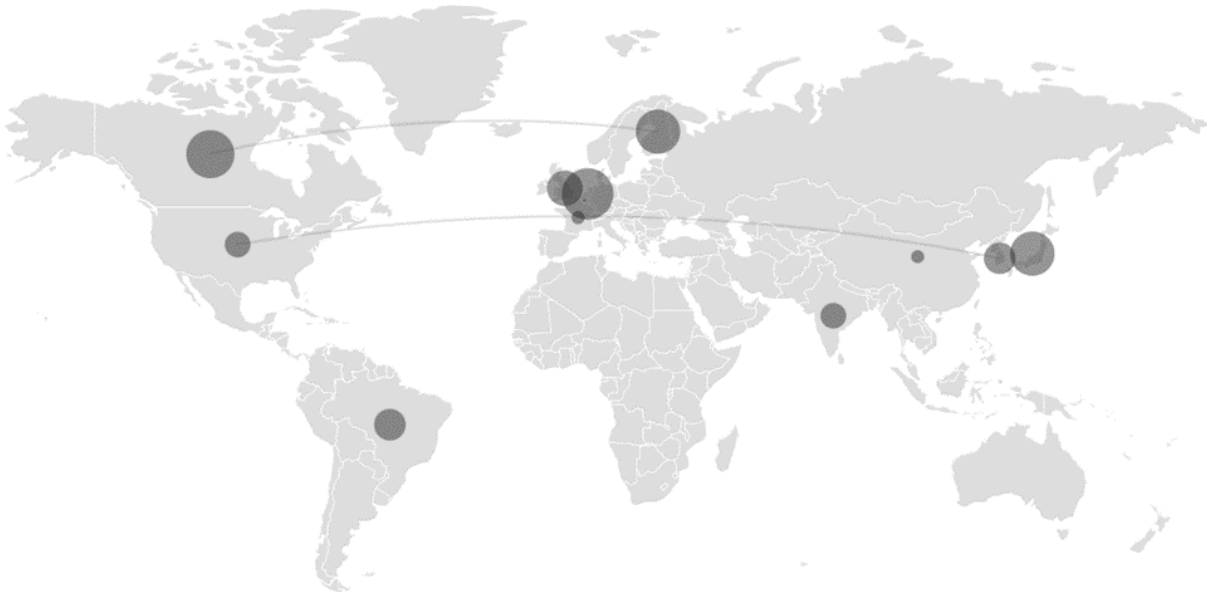

Figure 2: World location of authors' affiliations, and collaborations between research teams

Most of the studies selected in this review have been published since 2010 (13/22, 59%). The use of cepstral coefficients and of machine learning tools in speech assessment in an oncological context started around 2010 (see Figure 3). The field of speech acoustic analysis is therefore growing, due to the recent use of new acoustic measures.

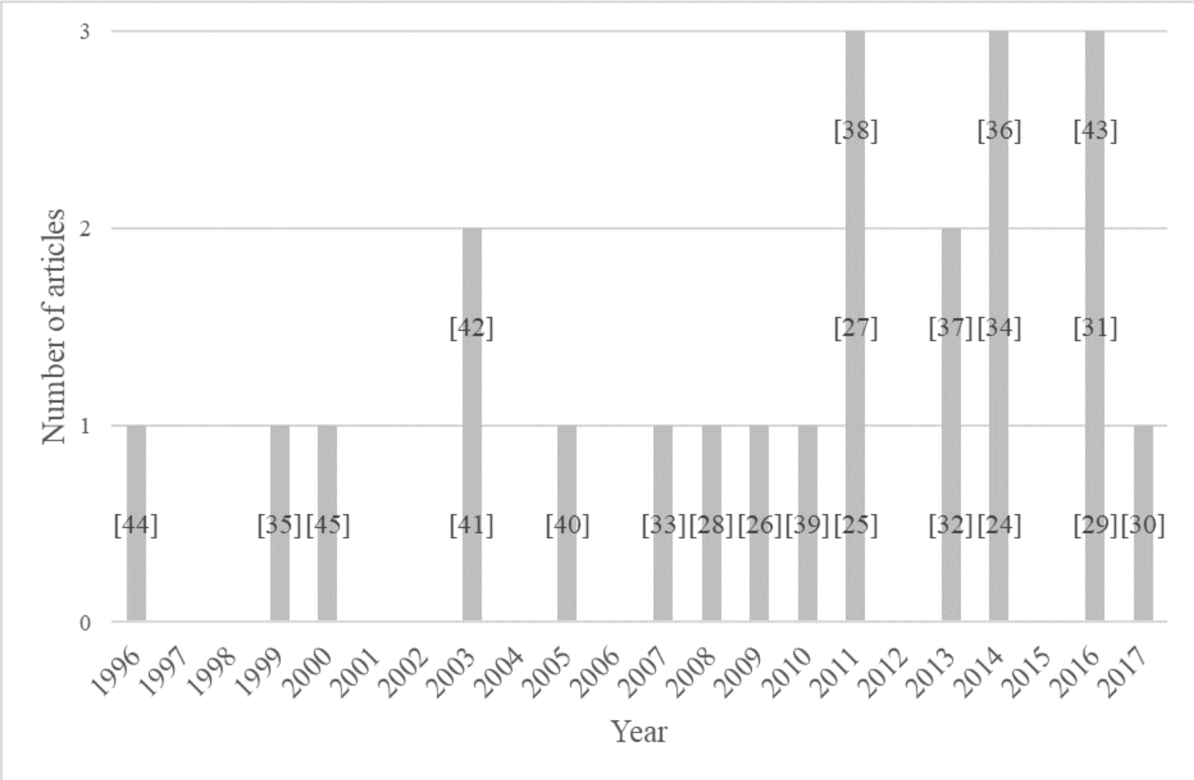

Figure 3: Number of articles selected per year (the numbers inside the bar charts are the reference of the article)

Participants

Among the 22 studies, 10 include more than 20 patients [25, 26, 27, 28, 29, 32, 33, 34, 40, 43], and also 10 include between two and 18 subjects [24, 31, 35, 37, 38, 39, 41, 42, 44, 45]. Two articles do not report the number of subjects involved in the acoustic analysis [30, 36]. Details are given in Table 3. Note that with the exception of two studies [24, 45], the subjects included are mostly men.

Table 3: Number of participants in the included studies

Number of participants	Number of studies (%)
2 to 5 participants	3 (14%)
6 to 10 participants	2 (9%)
11 to 20 participants	5 (23%)
21 to 50 participants	4 (18%)
51 to 62 participants	6 (27%)
Not reported	2 (9%)

Two studies [30, 34] use patient data from retrospective corpora.

All participants in the 22 studies had cancers of the oral cavity or of the oropharynx at the time of the study.

In total, 11 studies (50%) address patients treated for cancer of the oral cavity only. The anatomical sites mainly (9/11) involve the tongue (treated by total [28, 35, 38, 39, 43, 44], or partial glossectomy [24, 31, 33]). The remaining two studies investigate maxillary tumors [37, 45].

Six studies (27%) include both patients treated for cancer of the oral cavity and patients treated for oropharynx cancer [26, 27, 29, 30, 34, 40].

Only five (23%) include only patients with an oropharyngeal tumor location. Two addressed patients with a tumor extension to the soft palate [41, 42]. The other three relate to the tonsil, alone [25] or in comparison to the area of the base of the tongue [32, 36].

The distribution of the tumor locations is illustrated in Figure 4.

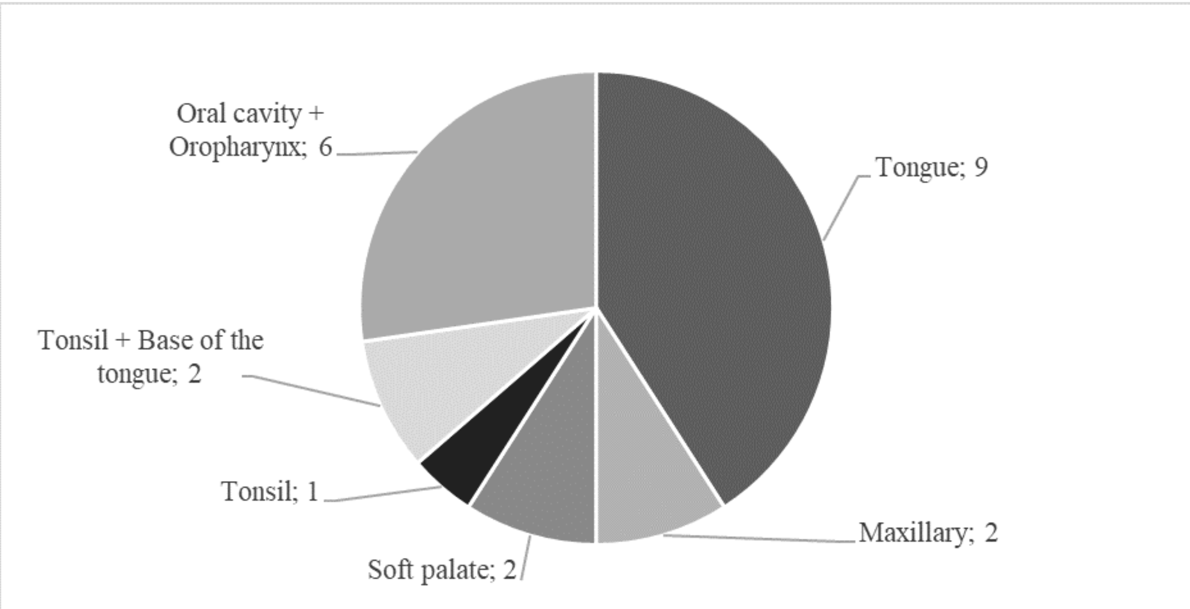

Figure 4: Tumor locations

Regarding the size of the tumor, 12 studies (54%) include smaller tumors (T1 + T2) than large ones [25, 26, 27, 29, 31, 32, 35, 38, 39, 40, 43, 44]. Three studies (14%) focus on larger tumors (T3 + T4) [30, 34, 42]. One study (5%) includes as many subjects with small T1 + T2 tumors than with larger T3 + T4 tumors (T3 + T4) [41]. Lastly, six studies (27%) do not report the size of the tumor of participants [24, 28, 33, 36, 37, 45].

Figure 5 shows the detailed proportion of the tumor sizes across the studies that reported these sizes.

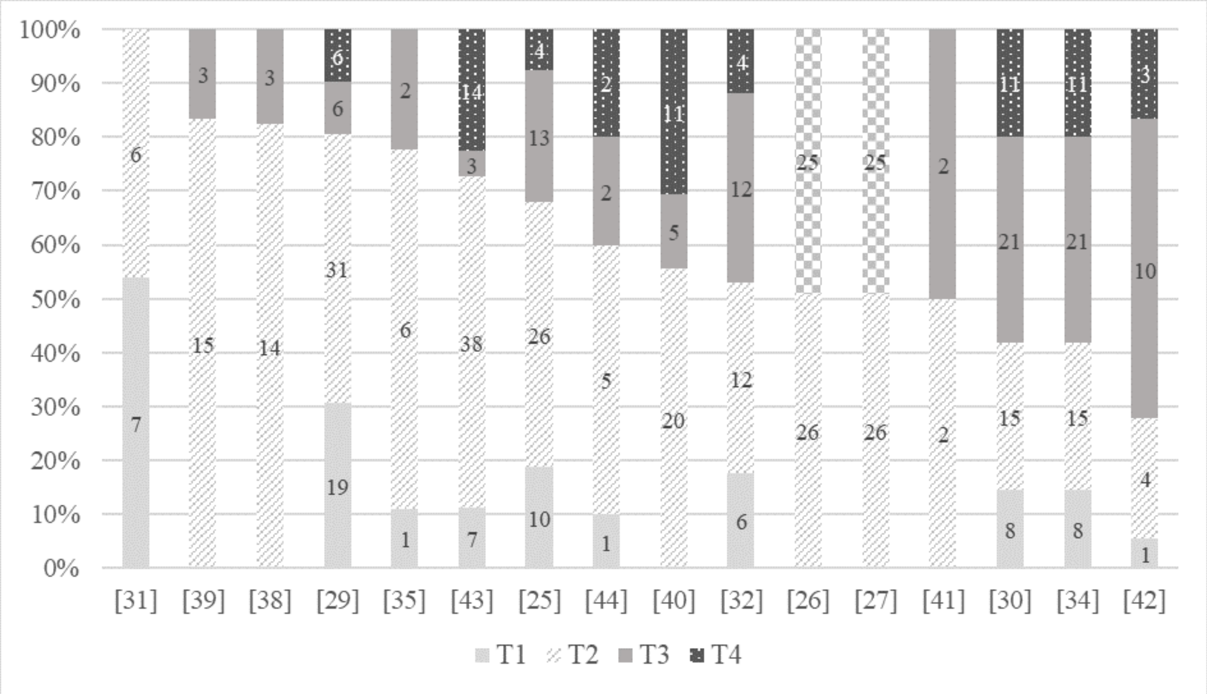

Figure 5: Detailed proportion of the tumor size (T classification) in the retained articles (note that articles 26 and 27 do not differentiate T3 and T4 sizes in the 25 participants)

Of the 22 included studies, 20 (91%) address surgically treated patients. Among them, surgery was carried out exclusively (with no reported information about complementary treatment) in 14 studies, including seven surgical reconstructions [25, 26, 27, 28, 29, 31, 33, 37, 38, 39, 40, 43, 44, 45]. Surgery was combined with other treatment methods such as radiotherapy or chemotherapy in six studies [24, 30, 34, 35, 41, 42]. Finally, a radiochemotherapy without surgical treatment was performed for the participants of two studies (9%) [32, 36].

The main languages spoken by the subjects and thus constituting the speech sample are English ([29, 31, 38, 39], including American English and Canadian English) and Dutch in four studies, respectively [26, 27, 32, 34]. The remaining studies are all carried out in different languages: French [24], Portuguese [28], German [30], Finnish [35], Hindi [37], Japanese [42] and Korean [25]. Seven studies do not report the language [33, 36, 40, 41, 42, 44, 45].

Comparison outcomes

The different comparison outcomes used in the studies are shown in Table 4.

Table 4: Comparison outcome chosen by the authors of included studies

Comparison outcome	**Number of studies (%)**
Perceptual assessment	6 (27%)
Global intelligibility	1
Specific parameters: articulation, nasality, “weakness”	4
Percent correct identification of consonants	1
Difference analysis	5 (23%)
Formants	1
Spectral parameters	2
Parameters or existing norms from software	2
Same parameters before / after treatment	3 (14%)
Same parameters in subjects and controls	8 (36%)

Six studies (27%) compare acoustic measures with a perceptual outcome. The latter is an intelligibility score assigned by judges using a Likert-type ordinal scale, either globally [24] or on specific parameters such as articulation, nasality or “weakness” [26, 27, 29, 45]. One study uses the percentage of correct identified consonants [25].

Five studies (23%) investigate the performance of acoustic scores either by analyzing differences between the investigated parameters or by comparing the results with existing data: comparison of formants [28], comparison of the performance of two spectral parameters [30, 34] and comparison with the same parameters from other software or with existing norms [37, 41].

Three studies (14%) compare acoustic parameters before and after treatment [32, 35, 36]. Eventually, eight studies (36%) compare the same parameters between a subject and a control group [31, 33, 38, 39, 40, 42, 43, 44].

Speech samples

Fourteen studies measure acoustic parameters in isolated phonemes. Specifically, eight analyze sustained vowels [24, 28, 29, 33, 35, 36, 37, 45], three analyze phonemes extracted from a read text [26, 27, 38] and two from isolated words [31, 39]. One study analyzes both sustained vowels and words (formants and their transitions [43]).

The speech sample of one study is composed of syllables [44], and another study recorded diadochokinesis [32].

One study carries out analyzes at the sentence-level [40], and four use a more global analysis on a read text [25, 30, 34, 41].

One study does not report the composition of its speech sample [42].

These results are shown in Table 5.

Table 5: Constitution of the speech samples

Speech sample	**Number of studies (%)**
Isolated phonemes	14 (64%)
Sustained vowels	8
Extracted from a read text	3
Extracted from isolated words	2
Combination of sustained vowels and phonemes in words	1
Syllables and diadochokinesis	2 (9%)
Sentences	1 (5%)
Read text	4 (17%)
Not reported	1 (5%)

Acoustic measures

The acoustic parameters analyzed in the included studies, reported below, are shown in Appendix B. Figure 6 represents the distribution of the units of analysis in the articles.

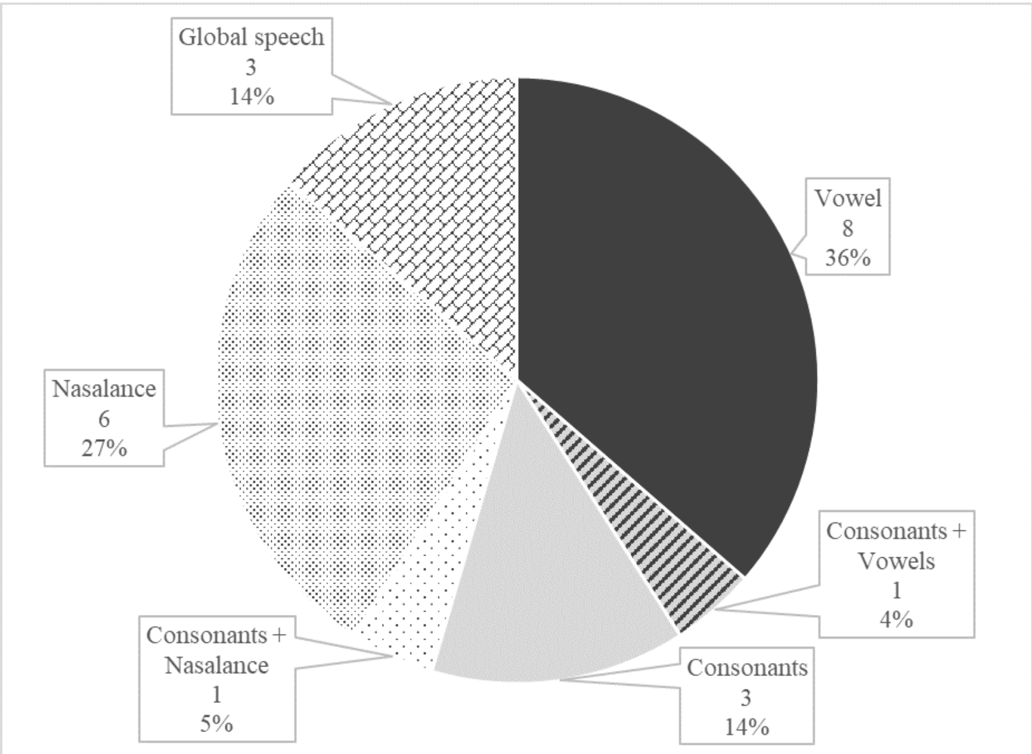

Figure 6: Number of studies analyzing the categories of acoustic parameters

Nasalance (7 articles)

Seven articles focus on the analysis of nasality. Three studies carried out the nasalance analysis on vowels [32, 36, 37], one on sentences [40], and two on a read text [25, 41]. One study does not report the speech unit used [42].

Most of the studies compute a nasality score by using dedicated software (Praat [36], Dr. Speech [37]) or nasometers [40, 41, 42]. The ratio of the acoustic energy emerging from the nasal and from the oral cavity is calculated in two studies [25, 32].

Nasalance score present a significant association with perceptual assessment, in extended resection or reconstruction of the soft palate [25]. Four other studies show an increased nasalance after treatment [32, 36, 37, 42]. One study shows that oral cavity tumors do not have a significant impact on the nasality in contrast to oropharyngeal tumors [40].

Vowels (9 articles)

Nine articles study the first and second formants (F1 and F2) of vowels [26, 28, 29, 31, 33, 35, 39, 43, 45]. Of these, three also study F3 [28, 31, 33] and one analyzes formants up to F12 [45]. The vowel space area (VSA) is used in two studies [26, 43] and the transition slope is only found in one [43].

Four studies investigate acoustic differences before and after treatment. After tongue surgery, significant differences are found in F1 and F2 [35, 39], with F1 generally being increased and F2 being lowered. The acoustic measures are impacted by local reconstruction [43], as well as by a well-adapted palate-lowering prosthesis, which is shown to modify F1, F2 and F3 in patients treated for a subtotal glossectomy [28]. Two studies show a correlation between acoustic measures and perceived intelligibility: F2 of /i/ ($r=0.35$) and the size of the VSA are linked with intelligibility ($r=0.39$, $p<0.05$) and articulation ($r=0.42$, $p<0.05$) ratings [26], and F7 and F12 of /i/ are also highly correlated with perceptual ratings ($r=0.84$, [45]). A single study does not find any significant correlation between acoustics and perceptual assessment on F0, F1 and F2 [29].

The studies comparing subjects and healthy controls find that F2 and F3 are lower in the patient group [31]. For women, significant correlations are found between subjects and controls for F2 and F3, but only for F1 for men [33].

Consonants (5 articles)

Three studies analyze spectral moments on plosives and fricatives: the center of gravity/spectral mean [24, 32, 38] and the spectral skewness [24, 26, 38]. The Klatt Voice Onset Time is also analyzed on both consonant groups in one study [24].

On plosive consonants, the duration of air pressure release is measured twice [26, 32]. The /t/ consonant peak energy frequency and the formant transition in the syllable /ta/ is analyzed in [44].

On fricatives, the friction duration and the band energy are calculated in two studies [32, 38]. In [32], F1, F2 and F3 are measured on liquids /l/ and /ʀ/.
The results show that the duration of the air pressure release in /k/ is linked with intelligibility and articulation estimates [26]. Also, the center of gravity and the skewness correlate with the perceptual evaluation in specific contexts (iCi and ɑCɑ context) [24]. The comparison pre *vs.* post-treatment allows considering the spectral mean and the skewness as good measures for short-term effects, and friction duration on /s, z/ does not seem to be relevant for long-time effects. One year after chemoradiotherapy, the spectral burst peak frequency of /k/ is weakened, a significantly higher F3 with lower intensity is found on /l/, and a significant higher spectral burst frequency on /t/ (higher spectral burst frequency) is noted [32]. Across different contexts, the Klatt V.O.T. seems congruent with the perceptual assessment [24]. Lastly, the formant variance F2-F3 at the transition between plosive and vowel returns to normal after surgery, and the Consonant Peak Energy Frequency is lower pre-surgery for some subjects [44].

Global speech (3 articles)

Two articles study the performance of different acoustic features, computed from existing corpora, in order to classify speech into two categories (intelligible / unintelligible). The investigated features are: MFCC (Mel-Frequency Cepstrum Coefficients) and MSCC (Mel S-transform Cepstrum Coefficients) in [30], and MRSTC (Multi Resolution Sinusoidal Transform Coding) in [34]. These features are fed to different classifying algorithms that output a binary decision on the intelligibility: article [34] uses a regression-based classifier, article [30] a support vector machine. A third article uses an Artificial Neural Network (ANN) to predict articulation quality and nasalance [27].
MSCC yield better results than MFCC in classifying intelligible and unintelligible speech on retrospective corpora, and MRSTC show a better classification when they are fed to an SVM (Support Vector Machine) [34]. ANNs significantly predict perceived articulation quality on /ɑ/, as well as perceptual hypernasality on /i/ and /u/ [27].

DISCUSSION

The main goal of this study was to review the scientific literature studying the effects on speech intelligibility of oral or oropharyngeal cancer, using acoustic parameters.
Two main lines of thought emerge from the analysis of the 22 selected articles, regarding the choice of the acoustic parameters, and the unit of analysis chosen to assess intelligibility.

Acoustic parameters according to participants' characteristics

If we look at the most investigated acoustic analyzes used in the studies retained for this review, two main fields can be determined: the nasality measures and the vowel acoustics.

The location of the tumor plays a role in the choice of the acoustic parameter. Most of the studies including oropharyngeal cancer patients use nasalance measures as one of the criteria impacting intelligibility. Among these studies, five include only oropharyngeal cancer [25, 27, 36, 41, 42], and two include patients undergoing surgery for the oropharynx or the cavity oral [32, 40]. The oropharyngeal pathology, because of its location, has an impact on the dynamics of the anatomical structures that account for speech nasality, particularly by its effect on the soft palate or the tonsil.

The majority of the studies including oral cavity cancer patients analyze acoustics on vowels and consonants. If nasalance is mainly assessed at a sentence or text level, most of the other analyzes, however, focus on the acoustic characteristics of isolated vowels, produced singly or more rarely extracted from syllables or continuous speech. The analyzes are mainly carried out on the first formants, which are known to be directly impacted by the oral pathology: the opening of the jaw modifies F1 and the position of the tongue modifies F2. The studies making the link between these formant measures and perceived intelligibility (perceptual comparison criteria are used in 3 articles out of 9 addressing formant measures) put forward the interest of 3 main parameters: the size of the VSA [26], F2 in the vowel /i/ and ANN-based nasalance scores on /i/ [27].

Regarding the analyzes on consonants, their type induces the use of different acoustic parameters. On the plosives [p t k], the spectral analysis of the burst and the air pressure release seem relevant [26, 32]. The center of gravity, spectral slope and band energy are more commonly used for fricatives [26, 32, 38, 44]. Thereby, the acoustic parameters analyzed depend on the location of the tumor: the analyzes on vowels and consonants relate mainly to oral cavity patients, whereas nasalance concerns mainly patients treated for oropharynx cancer. This is congruent with the expected functional impact of the morphological and dynamic changes consecutive to the treatment. It therefore seems appropriate to adapt the choice of acoustic parameters to the pathology presented by the patient in the clinical assessment.

Regarding the size of the tumor, the intelligibility in the context of small tumors is mainly analyzed on vowels (mostly formant analysis), and on consonants (spectral moments). Nasality is only investigated in one study, using an ANN on vowels [27].

The three studies including larger tumors [30, 34, 42] mainly use cepstral coefficients (MFCC, MSCC, MRSTC) [30, 34].

The use of feature extraction and of neural networks is fairly recent in the field of intelligibility assessment and shows promising performances in terms of intelligible / unintelligible binary classification, with the perceptual judgment as the external validation criterion.

The size of the tumor, in accordance with the impact on the anatomical structures involved in speech production, seems to determine the acoustic criteria. Phoneme-specific acoustic parameters are thus mainly used in tumors of small volumes, having a lesser impact on speech dynamics. Regarding tumors of larger volumes, studies look for more general speech-quality parameters to categorize speech as intelligible or unintelligible.

Subcategory analyzes by treatment and by language did not reveal any trend, particularly because of the small numbers of studies and patients in each category. Only two languages are found in more than one article: English and Dutch in four studies. Among them, two analyses the same cardinal vowels and the first two formants [26], [29]: they show different results regarding the correlation between these scores and the perceptive assessment of intelligibility. More studies are thus required to specifically study the effect of the phonemic constitution of a language on patients' intelligibility after treatment.

To summarize, a tight link seems to exist between the acoustic parameters and the tumor location, as well as between these parameters and the tumor size. Moreover, there is a great variability in acoustic parameters used in the different studies, mainly at the segmental level. The use of cepstral parameters and machine learning tools allows continuous speech analysis, but these techniques are still very recent and research needs to be developed. Currently, acoustic parameters seem to be relevant to complete the perceptual assessment of speech, carried out in current practice. It would therefore seem appropriate to investigate more comprehensive analysis models that not only classify patients' speech according to their functional intelligibility performance, but also study the fine acoustic impact of a tumor to enable targeted management of analytic deficits.

Speech samples

The analysis of the speech samples on which the acoustic parameters are measured shows a predominance of the study of isolated phonemes (vowels or consonants). Sentences or texts are rather used for the measurement of cepstral coefficients (such as MFCC or MSCC) or nasalance.

However, in a functional point of view, the analysis of semi-spontaneous or spontaneous speech would be the closest way to predict the intelligibility in the patient's daily life. From our review, we notice that there are no studies on such tasks, such as an image description or spontaneous speech analysis.

Study limitations

This systematic review surveyed two databases (PubMed and Embase). The Web of Science was also surveyed, but no entry was found that was not also present in the first two databases (i.e., all articles found in the WoS were duplicates of the PubMed and Embase entries). However, it is not excluded that

other studies exist outside the scope of this search.
In the 22 articles that were selected, two studies were carried out on identical or very similar corpora: [26, 27], and [38, 39]. However, both were retained because the main objectives were different and complementary: [26] focused on formant analysis while [27] used ANN; [38] investigated the analysis of the spectral moments on consonants, while [39] studied formants in vowels.
The great variability of the included studies underlines the need for the development of standardized tools of acoustic evaluations in patients treated for head and neck cancer. Standardization can enable to carry out more precise and reliable assessments in the diagnosis of speech disorder and its severity, but also in intra-individual comparisons in patient follow-up.

Future directions for research

Numerous acoustic parameters allow differentiating subjects suffering from cancer of the oral cavity or of the oropharynx, from healthy controls. This is the case for formant analysis mainly in cancers of the oral cavity [31, 33, 39, 43, 44], but also for nasality scores in two studies [40, 42]. The clinical validity of these measures has thereby been underlined. Other parameters allow the measurement of a change before / after treatment, such as spectral burst frequencies on /t/ and /k/ [32] or nasalance scores [36] for oropharyngeal cancer patients, and F1 and F2 for oral cavity cancer patients [35]. These parameters therefore show a good responsiveness.
However, one important question still needs to be addressed: Which golden standard can be used to evaluate the criterion validity of these different parameters? Six studies choose the perceptual evaluation as a golden standard, which is currently the standard in clinical practice. The discussion on the choice of this golden standard remains open.

When conducting our initial database search, the inclusion term "intelligibility" has led to many articles not addressing speech per se, but the quality of voice. It seems that no consensus is reached in the literature regarding the definition of intelligibility.
Moreover, most of the studies focused on the quality of acoustic-phonetic decoding on phonemes (vowels and consonants), to account for the speech intelligibility. However, there are several additional factors that can affect the quality of speech. The inclusion of other elements of the speech signal in addition to the acoustico-phonetic decoding [21] – such as nasality, speech rate [46] and other temporal and / or prosodic parameters related to perceived impairment [47] – defines the more complex notion of speech disorder severity.

The differentiation between the notions of intelligibility and severity of a speech disorder can also be applied to the question of the impact of these disorder levels at a functional (i.e., communication) and at a psychosocial level.

The automatic speech analysis is mainly performed at the segmental level, which is a context allowing a better control of the speech production of the patient. Speech assessment on a read text, which is a semi-spontaneous speech, allows controlling the context of speech production. Although the majority of the speech units from the selected studies are isolated phonemes, and more rarely sentences or texts, none investigated semi-spontaneous or spontaneous speech. True spontaneous speech is based on non-constrained productions, such as conversational speech. But the automatic analysis of this spontaneous speech is more complex to perform because it does not allow any reference to which comparing the performance of the patient, and that it includes many associated linguistic dimensions (phonemic, lexical, syntactic, prosodic). However, the functional impact of the speech disorder lies in the decrease of the patient's ability to transmit a message. Despite these challenges, acoustic measurements on spontaneous speech need to be developed. This context of production is the closest to communication situations experienced by patients on a daily basis, in communication with peers. Thus, the development of automatic tools objectively measuring speech on picture-description task or spontaneous speech (such as talking about the last holidays), using specific parameters (e.g., acoustics on phonemes, coarticulation, prosody, speech rate...) seems to be an interesting lead for future research, facilitated by the recent evolution of technology [48]. Within a perspective of speech evaluation closely reflecting the patient's daily production, the functional impact of the speech disorder must be taken into consideration.

Thus, an overall assessment of speech seems relevant. It would include an objective assessment using specific acoustic measures – specifically according to tumor location –, a perceptual evaluation (which is more global because it involves the complexity of speech disorder perception), and new tools for measuring the functional speech impairment (such as self -questionnaires). On the one hand, this would allow a more reliable and accurate assessment of deficits caused by the tumor or its treatment. Relevant linguistic units are to be searched and studied in speech signal to improve the intelligibility measurement of speech production disorders. On the other hand, this overall assessment could better take into account the functional consequences on daily life communication, by the assessment of associated deficits or communication needs. Indeed, the correlation between severity of speech impairment perceptively assessed and quality of life is only moderate [4]. A multidimensional assessment of speech disorders will allow customizing the therapeutic protocols in rehabilitation by capturing new information in speech signal and targeting more objectively deficits and, but also anticipating the functional and psychosocial impact by adapting therapeutic strategies.

Moreover, the automatic acoustic analysis tools, in addition to categorizing speech into intelligible/unintelligible, could also be used to determine finer cut-off points for speech disorder severity levels, depending on the functional impact.

Conclusion

Speech assessment in patients with cancer of the oral cavity or of the oropharynx by objective acoustic measures is in development. While many studies focus on the acoustic analysis of isolated phonetic features, the link with functional consequences and psychosocial repercussions must be studied.
More studies are needed to develop new automatic tools and to study which information they allow eliciting about the self-perceived impairment and the speech-related quality of life.

ACKNOWLEDGMENTS

This research was funded by the Hospitals of Toulouse (France), and Grant ANR-18-CE45-0008 from The French National Research Agency in 2018 RUGBI project “Improving the measurement of intelligibility of pathological production disorders impaired speech” led by Jérôme Farinas at IRIT.

The authors thank Dr. Guillaume Cabanac from the Information Retrieval and Information Synthesis team (IRIT, Toulouse) for his help with the online tool Netscity, allowing us to gather the bibliometric data of the retained articles.

The authors declare that they have no conflict of interest.

References

[1] Dwivedi, R. C., St. Rose, S., Roe, J. W. G., et al. (2010). First report on the reliability and validity speech handicap index in native english-speaking patients with head and neck cancer. Head and Neck, 36(10), 1391. https://doi.org/10.1002/HED

[2] Mlynarek, A., Rieger, J., Harris, J., et al. (2008). Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. J Otolaryngol Head Neck Surg, 37(1), 2–10.

[3] Reich, M. (2009). Cancer et image du corps : identité, représentation et symbolique : Le corps retrouvé (French). Cancer and Body Image : Identity, Representation (English), 85(3), 247–254. https://doi.org/10.3917/inpsy.8503.0247

[4] Borggreven, P. A., Verdonck-De Leeuw, I. M., Muller, M. J., et al. (2007). Quality of life and functional status in patients with cancer of the oral cavity and oropharynx: Pretreatment values of a prospective study. European Archives of Oto-Rhino-Laryngology, 264(6), 651–657. https://doi.org/10.1007/s00405-007-0249-5

[5] DeNittis, A. S., Machtay, M., Rosenthal, D. I., et al. (2001). Advanced oropharyngeal carcinoma treated with surgery and radiotherapy: Oncologic outcome and functional assessment. American Journal of Otolaryngology - Head and Neck Medicine and Surgery, 22(5), 329–335. https://doi.org/10.1053/ajot.2001.26492

[6] Stelzle, F., Knipfer, C., Schuster, M., et al (2013). Factors influencing relative speech intelligibility in patients with oral squamous cell carcinoma: A prospective study using automatic, computer-based speech analysis. International Journal of Oral and Maxillofacial Surgery, 42(11), 1377–1384. https://doi.org/10.1016/j.ijom.2013.05.021

[7] Colangelo, L. A., Logemann, J. A., Rademaker, A. W. (2000). Tumor size and pretreatment speech and swallowing in patients with resectable tumors. Otolaryngology-Head and Neck Surgery, 122(5), 653–661. https://doi.org/10.1016/S0194-5998(00)70191-4

[8] Dwivedi, R. C., Kazi, R. A., Agrawal, N., et al. (2009). Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers. Cancer Treatment Reviews, 35(5), 417–424. https://doi.org/10.1016/j.ctrv.2009.04.013

[9] Reuschenbach, M., Tinhofer, I., Wittekindt, C., Wagner, S., Klussmann, J. P. (2019). A systematic review of the HPV-attributable fraction of oropharyngeal squamous cell carcinomas in Germany. Cancer Medicine, 8(4), 1908–1918. https://doi.org/10.1002/cam4.2039

[10] Wittekindt, C., Wagner, S., Bushnak, A., et al. (2019). Increasing incidence rates of oropharyngeal squamous cell carcinoma in Germany and significance of disease burden attributed to human papillomavirus. Cancer Prevention Research, 2561, canprevres.0098.2019. https://doi.org/10.1158/1940-6207.capr-19-0098

[11] Ghio, A., Giusti, L., Blanc et al. (2016). Quels tests d'intelligibilité pour évaluer les troubles de production de la parole ?, *Journées d'Etude sur la Parole*, 2016, Paris, France. pp. 589-596. https://hal.archives-ouvertes.fr/hal-01372037

[12] Plisson, L., Pillot-Loiseau, C., Crevier-Buchman, L. C. (2017). 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. *7èmes Journées de phonétique clinique (JPC7)*, Laboratoire de Phonétique et Phonologie, hôpital Européen G. Pompidou, June 2017, Paris, France. https://hal.archives-ouvertes.fr/hal-01566225

[13] Middag, C. (2013). Automatische analyse van pathologische spraak Automatic Analysis of Pathological Speech.

[14] Fex, S. (1992). Perceptual evaluation. *Journal of Voice*, *6*(2), 155–158.

[15] Berisha, V., Utianski, R., Liss, J. (2013). Towards a clinical tool for automatic intelligibility assessment. Training, 2825–2828.

[16] Liberati A, Altman DG, Tetzlaff J, et al. (2009) The PRISMA statement for reporting systematic reviews and metaanalyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009; 151: 1–30.

[17] Moher D, Liberati A, Tetzlaff J, Altman DG. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009; 151:264–269.

[18] Kent, R. D., Weismer, G., Kent, J. F., Rosenbek, J. C. (1989). Toward Phonetic Intelligibility Testing in Dysarthria. Journal of Speech and Hearing Disorders, 54(4), 482. https://doi.org/10.1044/jshd.5404.482

[19] Keintz, C. K., Bunton, K., Hoit, J. D. (2007). Influence of visual information on the intelligibility of dysarthric speech. American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association, 16(3), 222–34. https://doi.org/10.1044/1058-0360(2007/027)

[20] Hustad, K. C. (2008). The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria, 51(June 2008), 562–573.

[21] Lindblom, B. (1990). On the Communication Process: Speaker-Listener Interaction and the Development of Speech. Augmentative and Alternative Communication, 6(4), 220–230. https://doi.org/10.1080/07434619012331275504

[22] National Health and Medical Research Council (NHMRC). (1995). Guidelines for the development and implementation of clinical guidelines. 1st edn. Canberra: Australian Government Publishing Service. Lower

[23] Kmet, L.M., Lee, R.C., Cook, L.S. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Edmonton, Alberta: Heritage Foundation for Medical Research (AHFMR).

[24] Acher, A., Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. Clinical Linguistics & Phonetics, 28(April), 241–256. https://doi.org/10.3109/02699206.2013.802015

[25] Chung, E., Lee, D., Kang, H., Park, M., Chung, C., Rho, Y. (2011). Prospective speech outcome study in patients with soft palate reconstruction in tonsillar cancer. Oral Oncology, 47(10), 988–992. https://doi.org/10.1016/j.oraloncology.2011.07.022

[26] Bruijn, M. J. De, Kuik, J., Quené, H. (2009). Objective Acoustic-Phonetic Speech Analysis in Patients Treated for Oral or Oropharyngeal Cancer. Folia Phoniatrica et Logopaedica, 61, 180–187. https://doi.org/10.1159/000219953

[27] Bruijn, M. J. De, Bosch, L. T. E. N., Kuik, D. J., Langendijk, J. A., Leemans, C. R., Verdonck-de Leeuw, I. (2011). Artificial neural network analysis to assess hypernasality in patients treated for oral or oropharyngeal cancer. Logopedics Phoniatrics Vocology, 36(June), 168–174. https://doi.org/10.3109/14015439.2011.606227

[28] de Carvalho-Teles, V., Ubijara Sennes, L., & Gielow, I. (2008). Speech Evaluation After Palatal Augmentation in Patients Undergoing Glossectomy. Arch Otolaryngol Head Neck Surg, 134(10), 1066–1070.

[29] Dwivedi, R. C., St. Rose, S., Chisholm, E. J., et al. (2016). 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 and Neck, 36(10), 1391. https://doi.org/10.1002/HED

[30] Fang, C., Li, H., Ma, L., Zhang, M. (2017). Intelligibility Evaluation of Pathological Speech through Multigranularity Feature Extraction and Optimization. Computational and Mathematical Methods in Medicine, 2017.

[31] Ha, J., Sung, I., Son, J., Stone, M., Ord, R., Cho, Y. (2016). Analysis of speech and tongue motion in normal and post-glossectomy speaker using cine MRI. J Appl Oral Sci, 24(5), 472–480.

[32] Jacobi, I., Rossum, M. A. Van, Molen, L. Van Der, Hilgers, F. J. M., Brekel, M. W. M. Van Den. (2013). Acoustic Analysis of Changes in Articulation Proficiency in Patients With Advanced Head and Neck Cancer Treated With Chemoradiotherapy. Annals of Otology, Rhinology & Laryngology, 122(12), 754–762.

[33] Kazi, R., Prasad, V. M. N., Kanagalingam, J., Georgalas, C., Venkitaraman, R., Harrington, K. J. (2007). Analysis of formant frequencies in patients with oral or oropharyngeal cancers treated

by glossectomy. INT J LANG COMMUN DISORD, 42(5), 521–532. https://doi.org/10.1080/13682820601056566

[34] Kim, J. C., Rao, H., Clements, M. A. (2014). Speech intelligibility estimation using multi-resolution spectral features for speakers undergoing cancer treatment. J. Acoust. Soc. Am, 136(September), 315–321. https://doi.org/10.1121/1.4896410

[35] Knuuttila, H., Pukander, J., Ma, T. (1999). Speech Articulation after Subtotal Glossectomy and Reconstruction with a Myocutaneous Flap. Acta Otolaryngol, 119, 621–626.

[36] Kraaijenga, S. A. C., Molen, L. Van Der. (2014). 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. European Archives of Oto-Rhino-Laryngology. https://doi.org/10.1007/s00405-014-3379-6

[37] Kumar, P., Jain, V., Thakar, A., & Aggarwal, V. (2013). Effect of varying bulb height on articulation and nasalance in maxillectomy patients with hollow bulb obturator. Journal of Prosthodontic Research, 57(3), 200–205. https://doi.org/10.1016/j.jpor.2013.02.002

[38] Laaksonen, J.-P., Rieger, J., Harris, J., Seikaly, H. (2011). A longitudinal acoustic study of the effects of the radial forearm free flap reconstruction on sibilants produced by tongue cancer patients. Clinical Linguistics & Phonetics, 25(April), 253–264. https://doi.org/10.3109/02699206.2010.525681

[39] Laaksonen, J., Rieger, J., Happonen, R., Harris, J., Seikaly, H. (2010). Speech after radial forearm free flap reconstruction of the tongue : A longitudinal acoustic study of vowel and diphthong sounds. Clinical Linguistics & Phonetics, 24(January), 41–54. https://doi.org/10.3109/02699200903340758

[40] Markkanen-leppa, M., Isotalo, E., Mäkitie, A. A., Suominen, E., Asko-Seljavaara, S., Haapanen, M.-L. (2005). Speech Aerodynamics and Nasalance in Oral Cancer Patients Treated with Microvascular Transfers ¨. Journal of Craniofacial Surgery, 16(6), 990–995.

[41] Moerman, M., Vermeersch, H., Lierde, K. Van, Fahimi, H., Van Cauwenberge, P. (2003). Refinement of the free radial forearm flap reconstructive technique after resection of large oropharyngeal malignancies with excellent functional results. Head and Neck, (September), 772–777. https://doi.org/10.1002/hed.10295

[42] Seikaly, H., Rieger, J., Wolfaardt, J., Moysa, G., Harris, J., Jha, N. (2003). Functional Outcomes After Primary Oropharyngeal Cancer Resection and Reconstruction With the Radial Forearm Free Flap. The Laryngoscope, 113(May), 7–9.

[43] Takatsu, J., Hanai, N., Suzuki, H. (2016). Phonologic and Acoustic Analysis of Speech Following Glossectomy and the Effect of Rehabilitation on Speech Outcomes. Journal of Oral Maxillofacial Surgery, 75(7), 1530–1541. https://doi.org/10.1016/j.joms.2016.12.004

[44] Wakumoto, M., Ohno, K., Imai, S., Yamashita, Y., Akizuki, H., Michi, K.-I. (1996). Analysis of the articulation after glossectomy. Journal of Oral Rehabilitation, 23, 764–770.

[45] Yoshida, H., Furuya, Y., Shimodaira, K., Kanazawa, T., Kataoka, R., Takahashi, K. (2000). Spectral characteristics of hypernasality in maxillectomy. Journal of Oral Rehabilitation, 27, 723–730.

[46] Yorkston, K. M., Strand, E. A., Kennedy, M. R. T. (1996). Comprehensibility of Dysarthric Speech: Implications for Assessment and Treatment Planning. American Journal of Speech-Language Pathology, 5(1), 55–65.

[47] Auzou, P. (2007). Les objectifs du bilan de la dysarthrie. In Les dysarthries (pp. 189–195). Marseille: Solal.

[48] Saon, G., Chien, J. (2012). In modern speech recognition. IEEE Signal Processing Magazine, 29(November), 18–33. https://doi.org/10.1109/MSP.2012.2197156

Appendix A: Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis

Reference	Design (according to NHMRC)[2]	Qualsyst (by Kmet et al.)[3]	Participants[4], diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[24] Acher et al., 2014	IV	15/20 75 % (good)	*Participants:* 2 patients (1 M, 1F) with hemi-glossectomy and bilateral neck dissection, chemo-radiotherapy *T classification (T/N/M):* not reported *Language:* French	Patient 1: M, 28 y.o. Patient 2: F, 62 y.o.	"Spectral moments" (energy in frequency domain): Center of Gravity (COG, 1st moment): average of frequency distribution of spectral energy, Skewness (3rd moment, right-left asymmetry of spectral envelope: the higher the skewness, the more the spectral energy is localized on low frequencies as in posterior phonemes, Kurtosis (4th moment): shape of noise envelope of consonant Klatt Voice Onset Time (VOT, temporal parameter): transition from consonant to vowel, extended from stops to sibilants	*Measure:* 5-point scale on intelligibility (1-5: normal to unintelligible) *Raters:* 5 speech therapists *Task:* VCV sequences presented twice in a random order	24 CVCVC. The median consonant is analyzed when it is surrounded by a symmetrical vowel context (i-i, u-u, ɑ-ɑ) C: /t/, /d/, /k/, /g/, /s/, /z/, /ʃ/ or /ʒ/ V: /i/, /u/ or /ɑ/	COG and Skewness: changes correspond with the perceptual evaluation of a large majority of the analyzed consonants in the i-i and ɑ-ɑ context Kurtosis: not seem to be an efficient parameter in this context Klatt VOT: congruent with the perceptual evaluation when the latter could not be explained by spectral parameters
[25] Chung et al., 2011	III-3	20/22 91 % (strong)	*Participants:* 53 patients (48 M, 5 F) surgically treated for tonsil cancer (squamous cell carcinoma) *T classification (T/N/M):* T1=10, T2=26, T3=13, T4=4 *Language:* Korean	56.7 y.o., range 39-80 y.o.	Nasalance: calculation of the ratio of acoustic energy output of nasal sounds from the nasal and oral cavity	*Measure*: percentage of correct identification of consonants *Raters*: 1 speech-language pathologist *Task*: nasal text versus no nasal text reading aloud	Reading a no nasal passage, and a high nasal passage	Denuded reconstruction technique (p < 0.001), extent of soft palate resection (p = 0.001), and T-stage (p < 0.001) were significantly associated with the nasalance score assessed objectively and perceptively (Denuded reconstruction: p=<0.001, extent of soft palate resection: p<0.001, T stage: p=0.006)

[2] NHRMC hierarchy: Level 1 Systematic reviews; Level II Randomized control trials; Level III–1 Pseudo-randomized control trials; Level III–2 Comparative studies with concurrent controls and allocation not randomized (cohort studies), case control studies, or interrupted time series with a control group; Level III–3 Comparative studies with historical control, 2 or more single-arm studies, or interrupted time series without a control group; Level IV Case series.

[3] Methodological quality: strong > 80%; good 60–79%; adequate 50–59%; poor < 50%.

[4] M: Males, F: Females

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[26] De Bruijn et al., 2009	III-2	21/24 88 % (strong)	*Participants:* 51 patients (28 M, 23 F) treated surgically (with reconstruction) for advanced oral or oropharyngeal squamous cell carcinoma (21 oral cavity, 30 oropharynx) Included 6 months after treatment *T classification (T/N/M):* T2=26, T3-4=25 *Language:* Dutch	53.8 y.o., SD 8.7 Range: 23 - 73 y.o.	Vowel formants (/ɑ/, /i/, /u/): F1: first formant frequency associated with "height" (degree of opening of the vocal tract) F2: second formant frequency associated with the anterior-posterior tongue position Size of vowel space (area of the vowel triangle): amount of reduction in the vowel system measured in Hz Analysis of the velar consonants: /k/: duration of air pressure release (short silent period of pressure building + the pressure release) /x/: spectral slope	*Measure*: correlation coefficients between objective parameters and subjective blinded assessment of articulation and nasal resonance (using a 4-point scale) *Raters*: 2 speech pathologists *Task*: read text	Cardinal vowels in Dutch (/ɑ/, /i/, /u/) and velar consonants (/k/, /x/) from a read text with an approximate length of 60 seconds	On vowels: F1 /i/: r = -0.42 with nasal resonance F2 /i/: r = 0.35 with intelligibility and articulation Comparison between subjective assessment and size of vowel area: r=0.39 ($p<0.05$) with intelligibility, r=0.42 ($p<0.05$) with articulation On velar consonants: /k/: r>0.40 (significatively different) for with intelligibility and articulation /x/: r=0.33 ($p<0.05$) with nasalance Parameters involved in prediction of intelligibility: duration of air pressure release on /k/, size of vowel space and F1 /i/ ($p<0.05$)
[27] De Bruijn et al., 2011	III-2	20/22 91 % (strong)	*Participants:* 51 patients (28 M, 23 F) treated for advanced oral or oropharyngeal squamous cell carcinoma (21 oral cavity, 30 oropharynx) *T classification (T/N/M):* T2=26, T3-4=25 *Language:* Dutch	53.8 y.o., SD 8.7 Range: 23 – 73 y.o.	Artificial neural nets (ANNs): feature representation of an input speech signal, contain a number of model parameters (weights of the connections between network nodes) that determine the relation between input (in this paper, input context of 7 input MFCC) and output (specifically nasalance in this study: ANN-nasalance)	*Measure*: multivariate linear (for intelligibility) or logistic (for articulation quality and hypernasality) regression to obtain insight into the role of objective parameters in subjective speech evaluation (blinded assessment of articulation and nasal resonance (using a 4-point scale) *Raters*: 2 speech pathologists *Task*: read text	Two realizations of cardinal vowels in Dutch (/ɑ/, /i/, /u/) in different phonological contexts (stop consonants, liquid consonants and nasal consonants) Vowels extracted from a reading text with an approximate length of 60 seconds	Predictions by the amount on nasalance (ANN-nasalance): Intelligibility (R^2=21.3%): 2nd realizations of /ɑ/ (p=0.03) and /i/ (p=0.023) Articulation quality (R^2=48.7%): 2nd realization of /ɑ/ (p=0.05) Hypernasality (R^2=24.9%): 1st realizations of /i/ (p=0.048) and /u/ (p=0.008) Analyses on /ɑ/ predict articulation quality, on /i/ and /u/ predict hypernasality assessed perceptually

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[28] de Carvalho-Teles et al, 2008	III-3	18/22 82 % (strong)	*Participants:* 36 patients (33 M, 3 F) with total or subtotal glossectomy or hemiglossectomy and using a stable and well-adapted palate-lowering prosthesis for at least 3 months *T classification (T/N/M):* not reported *Language:* Portuguese	Range 30-80 y.o.	Spectrographic assessment of the formants (mean values of F1, F2 and F3 extracted from the most stable part of each vowel held 5 seconds approximatively) of the 7 vowels of Brazilian Portuguese, with and without the prosthesis	*Measure:* formants with and without prosthesis	Repetition of 18 syllables with plosive, fricative voiced and voiceless sounds, which are nasal and liquid sounds together with the vowel /ɑ/ Sustained emission in the usual frequency and intensity of the vowels /ɑ/, /e/, /é/, /i/, /o/, /ó/, /u/	With and without the prosthesis: F1: statistically significant differences for /ɑ/, /e/, /u/ ($p<0.001$), and statistical trend of difference for /o/ ($p=0.09$). F2: significant difference for /o/, /ó/, /u/ ($p<0.001$) and statistical trend for /e/, /i/ ($p=0.06$ and $p=0.08$, respectively). F3: significant difference for /ɑ/, /ó/ (analysis of variance, $p<0.001$).
[29] Dwivedi et al., 2016	III-2	24/24 100 % (strong)	*Participants:* 62 patients (41 M, 21 F) with oral cavity (21: 18 tongue, 3 floor of mouth) and oropharyngeal cancer (41: 13 base of tongue, 26 tonsil, 2 soft palate), surgically treated *T classification (T/N/M):* T1=19, T2=31, T3=6, T4=6 *Language:* English	Mean: 58.9 y.o.	F0, F1 and F2 analyzed on the mid-stable portion of the sustained vowel /i/ (containing the maximum possible periods – not fewer than 200 milliseconds for F0). LPC technique to evaluate the F1 and F2 formant frequencies	*Measure*: comparison of F1, F2, F0 and perceptual evaluation between patients and controls (4-point Likert scale on intelligibility, articulation, nasality, rate and weakness, overall grade) *Raters*: 3 experienced speech and language therapists *Task*: reading text passage ("The story of Arthur the rat")	Sustained vowel /i/ at a comfortable pitch and loudness (at least 5 sec) For perceptual assessment: reading specific words (bead, bed, booed) and reciting a standard passage at a comfortable pitch and loudness	Lack of correlation between F1 and F2 on /i/ and perceptual speech parameters (overall, intelligibility and articulation grade) ($p>0.15$) In patients: F1: affected by the elevation of the tongue, mouth closure and pharyngeal constriction F2: increased with elevation of the anterior tongue or depression of the posterior region of the oral cavity F0: rise in male patients with oral cavity cancer (especially oral tongue)

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[30] Fang, et al., 2017	III-2	14/20 70 % (good)	*Corpus 1* NKI-CCRT (Clapham, 2012): 55 patients (45 M, 10 F) with head and neck cancer surgery and chemotherapy *T classification (T/N/M):* T1=8, T2=15, T3=21, T4=11 *Corpus 2* SVD: 2000 persons with 71 subjects (both organic and functional disorders, gender not reported) *Language of the 2 corpuses:* German	Not reported (but reference to Clapham, 2012: mean 58 y.o., range 32-79)	MFCC (Mel-Frequency cepstrum coefficients) MSCC (Mel S-transform cepstrum coefficients): S-transform is a time-frequency analysis method which combines the advantage of wavelet transform with short time Fourier transform (better antinoise, time resolution and time-frequency localization).	*Measure*: Sensitivity (Se), specificity (Sp), Under a curve Area (UA) and Accuracy calculated on MFCC and MSCC in two groups	Corpus 1: reading German neutral text Corpus 2: recordings of vowels /ɑ/, /i/, /u/ produced at normal, high, low and low-high-low pitch, and recordings of a German sentence	Se: 67.15% (MSCC), 56.25% (MFCC) Sp: 62.36% (MSCC), 46.90% (MFCC) UA: 64.75% (MSCC), 51.58% (MFCC) Accuracy: 63.67% (MSCC), 50.54% (MFCC) MSCC parameters improved significantly in the classification rate between intelligible and not intelligible than the MFCC on both corpuses.
[31] Ha et al., 2016	III-2	20/22 91 % (strong)	*Participants:* 13 patients (8 M, 5 F) with post-partial lateral glossectomy patients *T classification (T/N/M):* T1=7, T2=6 *Language:* American English	Mean: 45.3 y.o.	First 3 formants of the middle segment of vowels /i/ and /u/, automatically extracted by Linear Prediction Coefficient (LPC)	*Measures*: calculation of (F2/F1), (F3/F2), (F3/F1), and comparison with a control group of 23 normal controls	Vowels /i/ and /u/ from seven repetitions of "a geese" and "a souk"	Comparatively to controls, patients had significantly lower F2/F1 ratios (F=5.911, p=0.018), and lower F3/F1 ratios that approached significance (F=3.482, p=0.067) In formant analysis, F2 and F3 of patients are lower.

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[32] Jacobi et al., 2013	III-3	19/20 95 % (strong)	*Participants:* 34 patients (27 M, 7 F) with advanced head and neck tumors treated by chemoradio-therapy 3 groups: BT: 8 base of tongue, 1 retromolar trigone NT: 6 tonsil, 1 soft palate, 5 others L: 13 larynx and hypopharynx *T classification (T/N/M):* T1=6, T2=12, T3=12, T4=4 *Language:* Dutch	Median: 58 y.o. Range: 39 – 77	F1 and F2 on vowels /ɑ/, /i/, /u/ Nasality on /ɑ/: relative band energy in the area of the second and third formants Center of the burst frequency and energy on /p/, /t/, /k/ Lower cutoff point of the frication spectrum, center of gravity, and band energy on /s/, /z/, /x/ Presence of the /t/ before the frication noise acoustically indicated by a burst and an energy break on /tj/ F1, F2 and F3 and their amplitudes for /l/, /r/	*Measure*: pairwise comparisons between before and after treatment	Standard Dutch text Diadochokinesis (“pataka” repetition) List of words (DYVA)	Ten weeks after the end of treatment: Nasality: decrease compared to baseline (p = 0.062 and t = 1.935). Significant differences for /t/ (higher spectral burst frequency), /s/ (more diffuse) and /r/ productions (increase of F2 and F3) 1 year after the end of treatment, significant differences (p<0.05) for /r/, /k/ (weaker), /l/ (significantly higher F3 with lower intensity), /x/ (lower frequency measures in initial and final position), /s/, /t/, /tj/
[33] Kazi et al., 2007	III-2	20/20 100 % (strong)	*Participants:* 26 patients (19 M, 7 F) with squamous cell carcinoma who underwent partial glossectomy *T classification (T/N/M):* not reported *Language:* Not reported	Mean: 53.1 y.o., SD 8.7	F1, F2 and F3: averaged LPC coefficients (for an estimate for each speaker)	*Measure*: correlation of formant frequencies between 31 control subjects and patients' group	Sustained vowel /i/ produced at a comfortable pitch and loudness for at least 5 seconds	Significant formant correlations: - between normal and study females on F2 (p=0.04) and F3 (p=0.02) - between normal and study males on F1 (p=0.01) No other significant formant correlations on other comparisons

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[34] Kim et al., 2014	III-3	19/22 86 % (strong)	*Participants:* corpus NKI-CCRT (Clapham, 2012): 55 patients (45 M, 10 F) with head and neck cancer surgery and chemotherapy: 5 oral cavity (2 floor of mouth, 3 tongue), 24 oropharynx (10 base of tongue, 7 tonsil, 2 soft palate, 5 others), 26 others (larynx, nasopharynx) *T classification (T/N/M) (Van der Molen, 2009):* T1=8, T2=15, T3=21, T4=11 *Language:* Dutch	Not reported (but reference to Van der Molen, 2009: mean 58 y.o., range 32-79)	MRSTC: multi resolution sinusoidal transform coding using wavelet-like analysis (lower frequency components calculated over a greater analysis windows length, higher frequency components estimated with a shorter window length)	*Measures*: Accuracy (ACC) defined as: (number of hits)/(number of instances) Unweighted accuracy (UWA): $UWA = \frac{1}{C}\sum_{c=1}^{C}\frac{Nb\ of\ hits\ in\ class\ c}{Nb\ of\ instances\ in\ class\ c}$ Comparison between SVM (Support Vector Machine) baseline and SVM baseline + MRSTC	Reading a 189-word passage from a Dutch fairy tale	On binary intelligibility classifier: ACC: 68.0 (SVM), 72.7 (SVM + MRSTC) UWA: 66.2 (SVM), 71.2 (SVM + MRSTC) Improvement in accuracy and unweighted accuracy with adding MRSTC to SVM
[35] Knuuttila et al., 1999	III-3	16/20 80 % (strong)	*Participants:* 9 patients (8 M, 1 F) operated for lingual cancer *T classification (T/N/M):* T1=1, T2=6, T3=2 *Language:* Finnish	Range: 43 – 75 y.o.	First (F1) and second (F2) formant estimated on a stable part of the sustained vowel (Fast Fourier Transform Analysis was used for spectral analyses)	*Measure*: differences in acoustic measurements of vowels after and before tongue resection (F1 and F2)	16 sustained Finnish vowels produced twice	Significant difference only for F1 of /i/ (p=0.01) and F2 of /ɑ/ (p=0.001). Higher F1 of /i/ (mean: 32, p=0.05) and lower F2 of /ɑ/ (mean: -97, p=0.01) after surgery

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[36] Kraaijenga et al., 2014	III-3	19/22 86 % (strong)	*Participants:* 18 patients treated by concurrent chemo-radiotherapy in 2 groups: NT group: 6 nasopharyngeal and tonsil tumors patients LHBT group: 12 with laryngeal, hypopharyngeal and base of tongue tumors *T classification (T/N/M):* not reported *Language:* Not reported	Mean age: 63 y.o., range 45 – 79	Nasality measured with the software Praat	*Measure*: comparison of antiformants between a baseline (post-surgery) and 2 years (Y+2) and 6 years (Y+6) post-treatment *Task*: sustained /ɑ/	Sustained /ɑ/	Improvements at Y+2 and deterioration at Y+6 in the "NT group". Deterioration compared to baseline also in "LHBT group" (paired t test p = 0.087).
[37] Kumar et al., 2013	III-3	16/22 73 % (good)	*Participants:* 10 patients (9 M, 1 F) treated by maxillectomy *T classification (T/N/M):* not reported *Language:* Hindi	Mean: 43 y.o., range 15 – 75	Nasalance measured by the nasal view of Dr. Speech software	*Measure*: automatic percentage of nasalance at different times (pre-operative, at compete healing without obturator, and 24h and 6 weeks after placement of the obturator)	Phonation of a sustained vowel	Nasalance increases abruptly from 20.16 +/- 5.52 to 52.04 +/- 19.25 after surgery (p = 0.005)

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[38] Laaksonen et al., 2011	III-3	21/22 95 % (strong)	*Participants:* 17 patients (11 M, 6 F) with tongue cancer (anterior 2/3 of the tongue was reconstructed with a radial forearm free flap) *T classification (T/N/M):* T2=14, T3=3 *Language:* Canadian English	M: 53 y.o. (mean), range 27 – 67 F: 59 y.o. (mean), range 17 – 72	Spectral moments (from spectrum computed from the waveform / spectrogram display using the fast Fourier transform) of the long time average (LTA) spectrum: mean (first moment), and skewness (third moment) Frication duration of sibilant sounds	*Measure*: Comparison of the measurements produced before tongue resection, 1, 6 and 12 months after reconstruction	Reading of six stimulus sentences and the Zoo passage: analyses of 892 tokens for /s,z/, and 88 tokens for /ʃ/	Spectral mean: Reduction of acoustic distinctiveness between /s, z/ and /ʃ/ Decreased first (pre-operative vs. 1-month post-operative, p< 0.0001, mean difference: 1094Hz) and then increased (1-month vs. 6-month post operative, p < 0.0001, mean difference: 752 Hz) without returning to pre-operative levels (pre-operative vs. 12-month post-operative, p < 0.0001, mean difference: decreasing of 654 Hz). Females got back to the pre-operative level gradually within 1-year period (pre-operative vs. 12-month post-operative, p < 0.0001, mean difference: 1142 Hz) Skewness: short term for both groups. Females achieved pre-operative level at 1-year post-operation (skewness: p=0.01, mean difference: 0.47). Frication duration on /s, z/: long-term effects not found (p>0.05)
[39] Laaksonen et al., 2010	III-3	17/20 85 % (strong)	*Participants:* 18 (12 M, 6 F) patients with tongue cancer (reconstruction of the anterior 2/3 of the tongue) *T classification (T/N/M):* T2=15, T3=3 *Language:* Canadian English	Range 27 – 72 y.o.	F1, F2, F0 and duration were analyzed Formant frequencies were obtained using linear predictive coding analysis (LPC)	*Measure*: Comparison of the parameters before the tongue resection and 1, 6 and 12 months after tongue reconstruction	Vowels /i/, /ɪ/, /ʌ/ and /u/ (chosen because of their articulatory, acoustical and perceptual distinctiveness) and diphthongs /aɪ/, /eɪ/ from a set of 6 sentences including [hVd] sequences (e.g. "heed" or "who'd")	For the male patient group, long-term effects were observed in F2 and in vowel duration. F2 decreased (interaction: F=3.262, p=.002; pre-op. vs 12 months post-op. [pairwise comparison] p=.003, mean difference 68 Hz) For the female patient group, no statistically significant changes were observed (all p-values > 0.05) for any of the outcome measures.

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[40] Markkanen-leppa et al., 2005	III-3	22/22 100 % (strong)	*Participants:* 44 patients (29 M, 15 F) treated surgically for oral cavity (OC), oro-pharyngeal (OP) or hypo-pharyngeal cancer *T classification (T/N/M):* T2=20, T3=5, T4=11, tumor recurrence=8 *Language:* Not reported	Mean: 56.2 y.o., range 38 – 80	Speech resonance is evaluated by nasalance (i.e. objective and specific acoustic substitution of perceived nasality) with a Nasometer	*Measure*: Comparison of values before and after surgery (post 6 weeks, 3 months, 6 months, 12 months)	Sentences loaded with voiceless plosive consonants or voiced consonants. Each sentence is repeated 3 times	Normal nasalance before and after operation in the OC patient group. In OP patients, however, nasalance increased after operation, differing significantly from OC patients at 6 weeks and 6 months after surgery ($p<0.05$)
[41] Moerman et al., 2003	IV	14/22 64 % (good)	*Participants:* 4 patients with oropharyngeal lesions with extension toward the soft palate (3 M, 1 F) *T classification (T/N/M):* T2=2, T3=2 *Language:* Not reported	Not reported	Nasalance measurement by nasometry	*Measure*: Comparison between mean nasalance scores	Reading of an oronasal, an oral and a nasal text	Normal scores for the nasal, oronasal, and oral text.

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[42] Seikaly et al., 2003	III-3	17/20 85 % (strong)	*Participants:* 18 patients (12 M, 6 F) treated for oropharyngeal cancer by surgery *T classification (T/N/M):* T1=1, T2=4, T3=10, T4=3 *Language:* Not reported	Mean: 55.1 y.o. (45 – 75)	Nasalance balance by nasometry assessed by Nasometer, PERCI-SARS and the Computerized Assessment of Intelligibility of Dysarthric Speech (CAIDS)	*Measure*: Comparison between mean nasalance scores at three times in point: before surgery, 1 month after surgery and pre-radiotherapy (RT), and 6 to 9 months after surgery and completion of radiotherapy	Not reported	Preoperative nasalance values differed significantly from pre-RT time values (p=0.05) Preoperative word intelligibility scores differed significantly from both pre-RT time (p<0.01) and post-RT (p<0.05) No significant differences for any of theses variables between pre-RT time and post-RT time.
[43] Takatsu et al., 2016	III-3	19/20 95 % (strong)	*Participants:* 62 patients with tongue cancer Group 1 (G1): 40 (29 M, 11 F) with partial glossectomy *T classification (T/N/M):* T1=6, T2=34 Group 2 (G2): 22 (15 M, 7 F) reconstructed *T classification (T/N/M):* T1=1, T2=4, T3=3, T4=14 *Language:* Japanese	G1: 55.8 y.o., range 30 – 77 G2: 47.5 y.o., range 21-69 y.o.	Formant frequency of vowels (center of 30-ms intervals of sustained vowels, excluding the first and last 25 ms): F1, F2 and slopes of formant transitions (formant slopes) between diphthongs /ɑ i/ and /u i/ Vowel area: inside area of a triangle for each patient made by plotting F1 and F2 for 3 vowels (/ɑ/, /i/, /u/)	*Measure*: Comparison of acoustic characteristics data collected during the preoperative and postoperative periods, and after rehabilitation	Sustained vowels: 3 seconds of /ɑ/, /i/, /u/ Word repetition task: /t a i k o/, /t ʃ i s a i/, /k a i k o/, /s u i k a/ were sampled 3 times each.	F2 of /i/ significantly decreased after glossectomy surgery (p<0 .001) G1: F1 slope decreased during the postoperative period (/t a i/: p=0.007; /k a i/: p=0.042). F2 slope significantly decreased for all diphthongs (p<0 .001). G2: F1 values increased for /i/ (p=0.064) and /u/ but significantly decreased for /ɑ/ (p=0.021). F2 values decreased for all vowels, with a significant difference for /i/ (p<0.001) and /u/ (p=0.002).

Reference	Design (according to NHMRC)	Qualsyst (by Kmet et al.)	Participants, diagnosis, language	Patients' age	Acoustic parameters (Definitions)	Comparison criteria	Speech sample	Author(s) main conclusion(s)
[44] Wakumoto et al., 1996	III-3	18/20 90 % (strong)	*Participants:* 10 patients treated by glossectomy Group 1 (G1): 5 directly sutured patients (4 M, 1 F) *T classification (T/N/M):* T1=1, T2=3, T4=1 Group 2 (G2): 5 patients reconstructed with forearm flip (5 M) *T classification (T/N/M):* T2=2, T3=2, T4=1 *Language:* Not reported	G1: median 37 y.o., range 28 – 67 y.o. G2: median 55 y.o., range 49 – 63 y.o.	Frequency characteristics at consonant section with the spectral envelope extracted by FFT from LPC coefficients: calculation of the formant frequency by peak picking method. CPF (Consonant peak energy frequency): physical evaluation score that aims to quantitatively evaluate the frequency characteristics at the consonant section Formant variance at transient portion: F2-F3	*Measure*: Comparison of the scores collected before operation, 1, 6- and 12-month post-surgery	Pronunciation of the target syllable /t a/, selected among speech intelligibility test samples because of its pronunciation with the front side closure of the oral cavity using tongue tip	G1 Directly sutured subjects: CPF: some got lower scores pre-operatively than at the post-operative sessions F2-F3: some showed close to the baseline scores 1-month post-operation G2 Reconstruction: CPF: some disclosed a concentration tendency approximatively at the 3-5 kHz area 1-month post-operation F2-F3: 1-month post-operation, some showed the scores close to the baseline
[45] Yoshida et al., 2000	III-3	19/20 95 % (strong)	*Participants:* 15 patients (4 M, 11 F) treated for various types of palatomaxillary or maxillary sinus cancers *T classification (T/N/M):* not reported *Language:* Not reported	Range 38 – 78 y.o.	Spectral analyses on 1/3-octave spectra obtained from data transformed by FFT (Fast Fourier transform) analysis and 1/3-octave band-pass filtered in the frequency range of 125 Hz to 6.3 kHz	*Measure*: 5-point scale assessment for hypernasality *Raters*: 2 speech pathologists	Sustained vowel /i/ during 0.5 second at an individually preferred pitch and loudness	Correlation: high correlation between the perceptual ratings and the predicted values (r=0.8419, adjusted r2=0.6872, F[2,27]=32,8480, P<0.001) by a stepwise regression (with the perceptive score as outcome, and F7 and F12, amplitudes of the 7th and 12th multiples in the normalized 1/3-octave spectra, as dependent variables)

Appendix B: Acoustic parameters analyzed in the included studies

	Nasalance		Vowels				Consonants				Global speech	
	Dedicated software or naso-meters	Ratio of acoustic energy output of nasal nasal and oral cavity	F1 and F2 (+/- F3)	F1 to F12	VSA	Transition slope	Spectral moments	Duration of air pressure release	Friction duration and the band energy	F1, F2, F3 on /l/ and /ʀ/	Feature extraction: MFCC, MSCC, MRTSC	ANN-nasalance
[24] Acher & Fougeron, 2014							✓					
[25] Chung et al., 2011		✓										
[26] De Bruijn et al., 2009			✓		✓		✓	✓				
[27] De Bruijn et al., 2011												✓
[28] de Carvalho-Teles, Ubijara Sennes, & Gielow, 2008			✓									
[29] Dwivedi et al., 2016			✓									
[30] Fang, Li, Ma, & Zhang, 2017											✓	
[31] Ha et al., 2016			✓									
[32] Jacobi, Rossum, Molen, Hilgers, & Brekel, 2013		✓					✓	✓	✓	✓		
[33] Kazi et al., 2007			✓									
[34] Kim, Rao, & Clements, 2014											✓	
[35] Knuuttila, Pukander, & Ma, 1999			✓									
[36] Kraaijenga & Molen, 2014	✓											
[37] Kumar, Jain, Thakar, & Aggarwal, 2013	✓											
[38] Laaksonen, Rieger, Harris, & Seikaly, 2011							✓		✓			
[39] Laaksonen, Rieger, Happonen, Harris, & Seikaly, 2010			✓									
[40] Markkanen-leppa et al., 2005	✓											
[41] Moerman, Vermeersch, Lierde, Fahimi, & Van Cauwenberge, 2003	✓											
[42] Seikaly et al., 2003	✓											
[43] Takatsu, Hanai, & Suzuki, 2016			✓		✓	✓						
[44] Wakumoto et al., 1996							✓					
[45] Yoshida et al., 2000				✓								