ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Associations Between Acoustic, Kinematic, Self-Reported, and Perceptual Measures of Speech in Individuals Surgically Treated for Oral Cancer

Thomas B. Tienkamp,[a,b,c] Teja Rebernik,[d,e] Bence M. Halpern,[f,g] Rob J. J. H. van Son,[f,h] Martijn Wieling,[a,c] Max J. H. Witjes,[b] Sebastiaan A. H. J. de Visscher,[b] and Defne Abur[a,c]

[a]Center for Language and Cognition Groningen, University of Groningen, the Netherlands [b]Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, the Netherlands [c]Research School of Behavioral and Cognitive Neurosciences, University of Groningen, the Netherlands [d]Brussels Centre for Language Studies, Vrije Universiteit Brussel, Belgium [e]Laboratoire de Phonétique et Phonologie, CNRS/Sorbonne Nouvelle, Paris, France [f]Netherlands Cancer Institute, Amsterdam [g]Nagoya University, Japan [h]Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands

ARTICLE INFO

Article History:
Received July 1, 2024
Revision received December 4, 2024
Accepted March 16, 2025

Editor-in-Chief: Rita R. Patel
Editor: Jessica E. Huber

https://doi.org/10.1044/2025_JSLHR-24-00464

ABSTRACT

Purpose: The purpose of this study was to assess differences between individuals treated for oral squamous cell carcinoma (ITOC) and control speakers on acoustic, kinematic, and perceptual measures of speech. Furthermore, this study aimed to assess the interrelatedness of these speech domains alongside self-reported speech outcomes in order to inform clinically relevant measures of speech in ITOC.
Method: Simultaneous acoustic and kinematic data (via electromagnetic articulography sensors on the tongue) were collected from nine ITOC, who received surgical treatment for a tumor located on either the tongue or jaw and eight age- and sex-matched control speakers. All participants were native speakers of Dutch and read the North Wind and the Sun Passage. We calculated the articulatory–acoustic vowel space (AAVS) from the acoustic data and the articulatory–kinematic vowel space (AKVS) from the tongue tip and tongue back sensor data. Inexperienced listeners (n = 35) provided intelligibility and listening effort ratings using a visual analogue scale rating procedure. Self-reported speech problems were assessed using the Speech Handicap Index.
Results: Compared to an age- and sex-matched control group, ITOC demonstrated a significantly smaller AAVS and AKVS of the tongue tip and back, as well as lower intelligibility ratings. A correlation analysis of all speech outcome measures within the ITOC group showed that group-wise, the acoustic, perceptual, and self-reported measures were most strongly associated with each other. While acoustic and kinematic measures were not strongly associated with each other on the group level, within-speaker correlations showed stronger acoustic–kinematic associations.
Conclusions: This study demonstrates that acoustic, perceptual, and self-reported measures are related and quantify speech problem severity between ITOC, while kinematic measures showed no between-speaker relationships in a systematic way. Acoustic and kinematic measures showed greater within-speaker than between-speaker associations, reflecting speaker-specific compensatory behaviors. Our results underscore the importance of assessing the speech outcomes of ITOC across the acoustic, kinematic, perceptual, and self-reported domains to inform rehabilitation strategies.
Supplemental Material: https://doi.org/10.23641/asha.29231144

Correspondence to Thomas B. Tienkamp: t.b.tienkamp@rug.nl. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.* Teja Rebernik is now with the Laboratoire de Phonétique et Phonologie, CNRS/Sorbonne Nouvelle, and the Brussels Centre for Language Studies, Vrije Universiteit Brussel.

Tumors in the oral cavity, which for 90% consist of oral squamous cell carcinoma (OSCC), affect approximately 378,000 people worldwide each year (Bagan et al., 2010; Ferlay et al., 2021). The staging of the tumor is in part determined according to its size, with sizes varying from T1 to T4, with T1 being the smallest and T4 being the largest tumors (Lydiatt et al., 2017). For the majority of OSCC cases, surgical resection is preferred over primary radiation therapy, especially if the tumor is accessible to the surgeon. However, surgery in the oral cavity is associated with problems with swallowing and/or speech articulation (Kreeft et al., 2009; Lam & Samman, 2013). While primary closure of the wound may result in speech problems due to reduced tongue mobility or tissue loss (i.e., missing tissue to form or approximate constrictions or more air escaping through less narrow constrictions), reconstruction of the tongue by means of a free flap may result in movement asymmetry as the flap can only move passively considering it is not functionally integrated (Bressmann, Uy, & Irish, 2005; Bressmann et al., 2007; Kappert et al., 2019). Surgical treatment for mandibular tumors can limit tongue mobility as well, as tumors are rarely restricted to the mandibular bone, and the surgical safety margin may comprise tissue from the floor of the mouth (FOM) or tongue (Bak et al., 2010; de Groot et al., 2020; Tienkamp, Rebernik, Halpern, et al., 2024). For larger tumors, postoperative radiation therapy may be necessary in order to minimize the risk of local recurrence. However, primary or adjuvant radiation therapy can result in additional fibrosis and muscle atrophy, which may further limit tongue mobility and hence speech quality (Lam & Samman, 2013).

As the incidence rates of OSCC are increasing, an understanding of the functional outcomes is key in improving posttreatment quality of life and designing new speech rehabilitation strategies (Constantinescu & Rieger, 2019; Dwivedi et al., 2009). It is especially important to better understand the posttreatment speech outcomes of individuals treated for OSCC (ITOC) as they rank speech as one of their most important priorities following treatment (Arslan et al., 2016; Ringash et al., 2018; Tschiesner et al., 2013). Posttreatment speech outcomes for ITOC can be quantified through various domains. In some cases, objective methods are used to quantify speech changes in ITOC, such as characterizing the speech signal using spectrotemporal measures (Acher et al., 2014; Guo et al., 2023; Tienkamp et al., 2023) or tracking the movement of the articulators directly using kinematic methods (Bressmann, Thind, et al., 2005; Hagedorn et al., 2021; Imai & Michi, 1992; Stone et al., 2014; Tienkamp, Rebernik, Halpern, et al., 2024; Zhou et al., 2013). Subjective methods are used to quantify speech changes in ITOC as well, such as perceptual assessments with inexperienced listeners or clinicians (Bressmann et al., 2004, 2009) or self-reported outcomes provided by the individuals themselves (Fichaux-Bourin et al., 2009; Park et al., 2016; Rinkel et al., 2008). While some studies have examined speech function in ITOC using acoustic and kinematic methods (Zhou et al., 2013); acoustic, kinematic, and perceptual methods (Acher et al., 2014; Wakumoto et al., 1996); self-reported and acoustic methods (Guo et al., 2023); self-reported and perceptual methods (Matsui et al., 2007; Park et al., 2016); and kinematic and perceptual methods together (Imai & Michi, 1992), the literature remains scarce regarding the simultaneous assessment of multiple speech domains in the same group of ITOC, which leaves some critical questions relatively unexplored. For example, assessment of multiple speech domains allows for the appraisal of whether deficits in one speech domain are associated with deficits in another speech domain.

Objective Speech Outcomes

Acoustic and kinematic methods allow for an objective quantitative assessment of posttreatment speech outcomes in ITOC that do not depend on human judgments, which are inherently subjective. Acoustic methods, in which spectrotemporal features are quantified from microphone recordings, do not require expensive equipment; can be made relatively quickly; and, as an accompanying benefit, allow for subsequent perceptual assessment. Acoustic studies often make use of vowel formant frequencies, which are reflective of resonances along the vocal tract. The first two vowel formant frequencies (F_1–F_2) are impacted by changes in tongue positioning and provide some insight into the underlying articulatory–kinematics of the tongue (Mefferd & Green, 2010; Whitfield & Goberman, 2014). Roughly, the F_1 is influenced by tongue and jaw height, whereas the F_2 is influenced by tongue frontedness. In turn, the size of the acoustic triangular vowel space area (VSA), which is determined on the basis of the F_1 and F_2 values of isolated corner vowels (/i, a, u/), provides some indication of the magnitude of the tongue movements. Changes in F_1 and F_2 may be found in speakers treated for tumors located on the tongue or the jaw, as tongue mobility is impacted in speech and nonspeech tasks, regardless of tumor location, while the jaw mobility needed for speech might be preserved in those treated for jaw tumors (de Groot et al., 2020; Tienkamp, Rebernik, Halpern, et al., 2024). For ITOC, the mobility of the residual tongue seems to be an important factor in posttreatment quality of life and speech outcomes, which makes the acoustic VSA a meaningful acoustic metric to quantify speech outcomes (Bressmann et al., 2004; Chepeha et al., 2016; Lam & Samman, 2013; Matsui et al., 2007; van Dijk et al., 2016).

In general, prior work on ITOC has indicated that the acoustic VSA is reduced following surgical treatment. For example, Guo et al. (2023) recorded sustained vowels of 92 individuals with tongue cancer (60 with T1–T2 tumors; 32 with T3–T4 tumors) and found that the size of the VSA was reduced at a 12-month follow-up compared to the preoperative recordings. The results further showed that individuals with T3–T4 tumors had larger VSA reductions compared to individuals with T1–T2 tumors. The negative impact of surgical treatment for OSCC on the acoustic VSA was also observed in 62 individuals treated for T1–T4 tumors using sustained vowels (Takatsu et al., 2017) and 10 individuals treated for T1–T4 tumors using consonant–vowel syllables (Whitehill et al., 2006). While these studies consistently show a reduction of the size of the acoustic VSA in ITOC, most of the studies employed sustained vowels, which may not be fully reflective of our daily communication. To assess speech outcomes in ITOC in a more ecologically valid way, formant frequencies can be analyzed in running speech, which only a limited number of studies have done. Laaksonen et al. (2010) recorded target sentences from 18 individuals treated for T2–T3 tumors with radial forearm free flap reconstructions and found no significant difference in terms of the size of the acoustic VSA between the pre-op and 12-month follow-up. Additionally, no statistical difference in VSA was found between those with and without additional floor of mouth resection or radiation therapy. While no significant differences were found for the overall size of the acoustic VSA pre- and postsurgery, a significant increase of F_1 was found across vowels postsurgery compared to presurgery, which signals a lower tongue position overall (Laaksonen et al., 2010). If high vowels as /i/ and /u/ are produced with a higher F_1, a push-shift might have been initiated as a compensatory response to preserve maximum acoustic distinctiveness from other vowels by lowering the tongue or the jaw. De Bruijn et al. (2009) recorded a reading passage from 51 patients with T2–T4 tumors in the oral or oropharyngeal region and documented a reduced VSA at a 6-month follow-up compared to control speakers. Lastly, Tienkamp et al. (2023) analyzed the spontaneous speech of five ITOC treated for tumors on the tongue or jaw and found no significant differences between the F_1 and F_2 of the corner vowels of ITOC and control speakers.

Even though relationships between acoustic vowel formants and tongue kinematics have been documented in typical speakers (i.e., the size of the acoustic VSA is positively correlated with the size of the kinematic VSA), acoustic measures of speech provide only an indirect measure of kinematic movements and may not fully capture kinematic speech information, especially in individuals with disordered speech such as ITOC (Dromey et al., 2013; Kuo & Berry, 2023; Lee et al., 2017; Mefferd & Green, 2010; Thompson & Kim, 2019, 2024; Whitfield et al., 2018; Wieling et al., 2016). Kinematic speech information may be obscured due to the many-to-one mapping between articulatory strategies and the resulting acoustic output, meaning that various articulatory configurations can be used to achieve the same acoustic goal (Perrier & Fuchs, 2015). For example, Perkell et al. (1993) showed that speakers may either raise the tongue body or increase the amount of lip rounding in order to lower the F_2 of /u/. These varying articulatory configurations, known as motor equivalence strategies, may be especially valuable to ITOC as they could leverage an unaffected articulator (e.g., the lips or the jaw) to compensate for reduced tongue movements (Hagedorn et al., 2022). Considering the variety in tumor size and location between ITOC, motor equivalence strategies may vary considerably as a result. Varying motor equivalence strategies may complicate the established relationship between articulatory–acoustics and articulatory–kinematics, especially when examining group data that use single flesh point locations for kinematic measures as is the case with electromagnetic articulography (EMA). Given the positive correlations between the size of the VSA and speech intelligibility, in both ITOC and individuals with dysarthria, a better understanding to what extent the acoustic VSA is related to the kinematic VSA could help to design speech therapy interventions for ITOC to maximize speech intelligibility (de Bruijn et al., 2009; Thompson et al., 2023; Turner et al., 1995; Whitehill et al., 2006). For example, if the acoustic VSA, kinematic VSA, and intelligibility are positively interrelated, an intervention surrounding maximizing tongue movement to increase the acoustic VSA and, in turn, speech intelligibility could be developed.

To measure underlying articulatory movements in ITOC directly, articulatory–kinematic methods need to be employed. The majority of studies targeting the articulatory–kinematics of speech outcomes in ITOC have focused on sentence- or phrase-level movement patterns using methods such as magnetic resonance imaging and EMA. This line of work has indicated that ITOC have reduced tongue movement, especially on the side where the tumor was located, and show less complex vocal tract shaping compared to control speakers (Hagedorn et al., 2021; Stone et al., 2014; Tienkamp, Rebernik, Halpern, et al., 2024). Moreover, ultrasound tongue imaging data showed that, compared to presurgery, tongue movements may become asymmetrical postsurgery as the reconstructed side only moves passively (Bressmann, Thind, et al., 2005). Studies looking at specific phonemes, which most often look at consonants as they are most easily recognized in the kinematic signal, have indicated that ITOC have less palatal contact during alveolar plosives as measured by electropalatography (/t, d/; Imai & Michi, 1992; Suzuki, 1989; Wakumoto et al., 1996). Alveolar fricatives (/s, z/) may be produced with a more posterior constriction location compared to presurgery or control speakers as shown by magnetic resonance imaging (Mady & Beer, 2007; Zhou et al., 2013). The results by Zhou et al. (2013) further showed that the constriction location and the acoustic spectra of /s/ and /ʃ/ were similar, leading to a possible sound merger, which could negatively affect speech intelligibility or acceptability. Collectively, kinematic studies suggest that ITOC have trouble raising the tongue high enough to form anterior constrictions.

Subjective Speech Outcomes

While acoustic and kinematic measures of speech provide objective accounts of the speaker's side of the communicative process, perceptual measures characterize the listener's perspective. Characterizing the listener's perspective provides additional information that may complement objective measures. Perceptual measures of speech often involve Likert scale ratings of a specific percept, such as the intelligibility, acceptability, or naturalness of speech. Typically, perceptual ratings are provided by either inexperienced (e.g., people with no experience with rating atypical speech) or experienced (e.g., speech-language pathologists [SLPs]) listeners. Perceptual studies have indicated that ITOC may have reduced word or sentence intelligibility following surgical treatment compared to pretreatment (Constantinescu et al., 2017) and controls (Loewen et al., 2010). This line of work has also shown interactions between speech outcomes on the one hand and clinical variables (e.g., resection size) or other functional outcomes (e.g., mobility of the residual tongue) on the other hand. That is, ITOC with larger resections received lower speech acceptability ratings compared to those with smaller resections (Bressmann et al., 2009; Matsui et al., 2007; Nicoletti et al., 2004), and ITOC with higher tongue mobility received higher speech intelligibility ratings as compared to those with lower mobility postsurgery (Bressmann et al., 2004; Matsui et al., 2007).

Studies employing both perceptual and kinematic methods have indicated that a tongue mobile enough to make contact with the palate results in better speech intelligibility. For example, Imai and Michi (1992) reported a negative correlation between the number of contacted sensors and the perceived distortion of /s/ by 17 ITOC with varying degrees of resection of the tongue, floor of mouth, and mandible. The authors also found that ITOC who made more linguo-palatal contact produced a less distorted /s/ compared to those with less linguo-palatal contact. Wakumoto et al. (1996) showed that the /t/ produced by ITOC who were not able to make much linguo-palatal contact was often perceived as the bilabial /p/. In contrast, speakers with (more) linguo-palatal contact had a well-perceived /t/. The perception of /t/ did not seem to be modulated by the place of the constriction as constrictions along the dental arch (i.e., the typical constriction location for /t/), as well as more posterior constrictions along the midsagittal plane of the palate led to a well-perceived /t/.

Perceptual studies have further shown the importance of assessing speech in ecologically more valid ways (i.e., ways that characterize daily communication as opposed to optimal quiet conditions in the lab). For example, Eadie et al. (2021) investigated the intelligibility and effort needed to follow the speech of 10 ITOC with either precise or imprecise speech articulation in a quiet versus noisy environment. The precision of speech articulation was assessed by three SLPs. The results of Eadie et al. (2021) showed similar intelligibility and listening effort levels in quiet conditions, but those with imprecise speech scored significantly lower in noisy conditions compared to those with precise articulation (i.e., lower intelligibility and higher effort). Thus, while speakers may be intelligible in optimal listening conditions, those with imprecise articulation are particularly vulnerable to more adverse listening conditions, which most of our daily communication comprise.

Finally, self-reported outcome measures provide important measures of daily communication from the speaker's perspective, which are important to measure in ITOC. For example, while acoustic, kinematic, or perceptual measures may indicate no significant speech impairment, ITOC may still experience speech difficulties, which should warrant further assessment. Self-reported speech outcomes by means of a questionnaire are the most commonly used method to quantify speech function in ITOC as it is quick to administer (Dwivedi et al., 2009). Self-reported outcomes are often quantified with questionnaires that target quality of life in relation to functional outcomes in general (e.g., Hassan & Weymuller, 1993; Laraway & Rogers, 2012; Malone et al., 2004) or questionnaires that target speech in particular. For example, the Speech Handicap Index (SHI) has been developed to track speech-related quality of life issues in ITOC, and it discriminates well between ITOC and individuals with typical speech or between ITOC before and after surgery or radiation-based treatment (e.g., Fichaux-Bourin et al., 2009; Park et al., 2016; Rinkel et al., 2008). Higher scores on the SHI indicate more pronounced speech-related quality of life issues. According to a recent review, the SHI has already been validated and translated into eight languages: Dutch, (U.K.) English, French, Korean, Mandarin Chinese, Lithuanian, Italian, and European Portuguese, showing the shift toward acknowledging and monitoring the patient's viewpoint more closely (Chan et al., 2021). While the primary target of the SHI is to document posttreatment speech-related quality of life, prior studies have established that there is a relationship between the SHI and speaker intelligibility (i.e., better speech-related quality of life related to better intelligibility; Park et al., 2016) and between the SHI and speech acoustics (i.e., better speech-related quality of life was related to a larger VSA; Guo et al., 2023).

The Present Study

Taken together, prior work highlights that acoustic, kinematic, perceptual, and self-reported measures each provide valuable and complementary insights into the speech outcomes of ITOC. To gain a more comprehensive understanding of posttreatment speech outcomes, combining subjective and objective methods may be beneficial. Considering there is no standardized speech assessment protocol to assess the speech outcomes of ITOC in research, the purpose of this study was to assess the speech of ITOC holistically. Such a holistic assessment allows for refinement of our understanding of how different speech domains are linked and affected following surgery. A better understanding of speech outcomes following surgical OSCC treatment and their interrelatedness may be beneficial for SLPs and surgeons, too, as it may help to (a) inform future speech assessment protocols in order to improve generalizability across studies, (b) design better speech treatments for ITOC as articulatory acoustics and kinematics can show the cause of reduced speech function, and (c) allow for better prediction as to which type of surgery will result in which speech deficits. Therefore, in this study, we report a comprehensive investigation of speech outcomes following surgical treatment for OSCC. Specifically, we investigated the acoustic, kinematic, perceptual, and self-reported speech outcomes of ITOC as compared to control speakers.

Our study had two aims. The first aim was to assess multiple speech domains in ITOC and control speakers: the acoustic domain (using the articulatory–acoustic vowel space [AAVS]), the kinematic domain (using the articulatory–kinematic vowel space [AKVS]), and the perceptual domain (using intelligibility and listening effort as rated by inexperienced listeners). A better understanding of how surgical treatment for OSCC impacts different domains of speech is beneficial for both informing the patient about postoperative speech functioning and developing a more nuanced understanding of the concurrence of speech difficulties across various domains. We predicted that ITOC, regardless of treatment site (i.e., both tongue and jaw), would show a smaller AAVS compared to typical speakers based on studies that found a reduced VSA for individuals treated for tongue cancer compared to either presurgery recordings or a control group and the fact that jaw cancer treatment can limit the mobility of the tongue (de Bruijn et al., 2009; de Groot et al., 2020; Guo et al., 2023; Takatsu et al., 2017; Whitehill et al., 2006). While, to the best of our knowledge, no study has assessed the kinematic vowel space in ITOC, we predicted that ITOC would show a smaller AKVS compared to control speakers based on the relationship between F_1–F_2 and tongue positioning and surgery-induced reductions in tongue mobility for both tongue and jaw cancer (de Groot et al., 2020; Kappert et al., 2019; Mefferd & Green, 2010; Tienkamp, Rebernik, Halpern, et al., 2024). Given that surgical treatment for oral cancer might result in atypical tongue movement patterns due to changes to both tongue mobility and musculature, we included both measurement from the tongue tip (TT) and the tongue back (TB) sensors in our study. Lastly, we predicted that ITOC would show lower intelligibility ratings and higher listening effort ratings compared to control speakers based on the literature that documented reductions in intelligibility following treatment (Bressmann et al., 2004, 2009; Constantinescu et al., 2017; Eadie et al., 2021; Loewen et al., 2010; Matsui et al., 2007; Nicoletti et al., 2004).

The second aim was to provide a preliminary assessment of the interrelatedness of the three assessed domains alongside self-reported speech outcomes postsurgery between and within ITOC to inform clinically relevant measures of speech in ITOC, which could help to inform speech rehabilitation strategies postsurgery. Specifically, we investigated to what extent the AAVS, AKVS, listener-provided intelligibility and listening effort ratings, and self-reported speech outcomes as measured by the SHI (Rinkel et al., 2008; Van den Steen et al., 2011) are related to each other for ITOC. We predicted negative relationships between the acoustic and self-reported domains based on Guo et al. (2023) as well as between self-reported and perceptual domains based on Park et al. (2016), meaning that higher levels of self-reported speech problems were associated with reduced speech intelligibility and a smaller AAVS in ITOC. Based on other work with speech disorders and typical speakers (Whitfield & Goberman, 2014, 2017), we expected a positive relationship between the acoustic and perceptual domains, meaning that a higher AAVS would be associated with better higher speech intelligibility levels in ITOC. While the AAVS and AKVS have not been evaluated in tandem in ITOC, we predicted a positive relationship between the acoustic and kinematic vowel space in ITOC based on the established relationship between formant frequencies and tongue position in typical speakers (Lee et al., 2016; Mefferd & Green, 2010; Whitfield et al., 2018), but also in speakers with dysarthria (Lee et al., 2017; Mefferd, 2015).

Method

Participants

Speakers

All speakers were selected from the corpus reported in Halpern et al. (2022) and Tienkamp, Rebernik,

Table 1. Demographics and clinical details of the individuals who received treatment for oral cancer.

Speaker ID	Sex	Age	T-stage	SHI	Side	Location	Procedure type	Reconstruction	FOM	PORT	CRT
NKI02	M	68	T2	24	Both	Mandible	Continuity resection	FOFL	X	-	X
NKI04	M	47	T4	29	Right	Tongue	Hemiglossectomy	ALTF	X	-	X
NKI05	F	54	T4	6	Right	Mandible	Hemimandibulectomy	FOFL	X	X	-
NKI06	F	57	T3	13	Left	Tongue	Hemiglossectomy	RFFF	X	-	-
NKI11	F	56	T4	31	Both	Mandible	Continuity resection	FOFL and PMMF	X	X	-
NKI15	M	69	—	18	Both	Mandible	Total mandibulectomy	FOFL	X	-	-
NKI16	F	71	T4	9	Right	Mandible: RTA	Hemimandibulectomy	FOFL	X	-	X
NKI17	M	57	T1–T2[a]	26	Both[b]	Tongue, maxilla, cheek, oropharynx, mandible	Hemiglossectomy	RFFF	X	-	X
NKI18	M	75	T2	6	Left	Anterior tongue	Partial glossectomy	SSG	X	X	-

Note. SHI = speech handicap index; FOM = floor of mouth resection; PORT = postoperative radiation therapy; CRT = chemoradiation therapy; M = male; FOFL = fibular osteocutaneous free flap; ALTF = anterior lateral thigh flap; F = female; RFFF = radial forearm free flap; PMMF = pectoralis major myocutaneous flap; RTA = retromolar trigone area; SSG = split skin graft.

[a]Patient had over seven procedures for T1 or T2 malignancies. [b]Hemiglossectomy and maxillary resection were performed on the right side. Oropharyngeal tumors were removed on both sides.

Halpern, et al. (2024). The corpus contains kinematic and acoustic data collected simultaneously using EMA and a microphone from 20 native Dutch individuals. Twelve underwent surgical treatment for OSCC, and eight were control speakers. ITOC received surgical treatment at least 1 year before participation for tumors that were staged between T1 (smallest) and T4 (largest). ITOC were notified about the project between October 2021 and April 2022 by their treating clinician during regular check-ups at the University Medical Centre Groningen. All individuals provided written informed consent at the time of participation in accordance with the approval granted for this study by the Medical Ethical Review Board of the University of Groningen (NL76137.042.20).

The present study analyzed the kinematic and acoustic data from nine ITOC (five male, four female) as well as eight age- and sex-matched controls (five male, three female). The data of three of the 12 ITOC in the corpus were excluded in the present study as no relevant kinematic data were available for these speakers. The nine ITOC included in this study received surgical treatment for a tumor located either on the jaw (n = 5) or on the tongue (n = 4). The average time post-surgery was 5.1 years with a range of 1.2–11.4 years. The age range of the ITOC (47–75 years, M = 61.1 years) was comparable to that of the control speakers (56–77 years, M = 60.9 years). All included speakers were native speakers of Dutch, denied a history of self-reported neurological or speech disorders (e.g., a stroke or a stutter), did not have self-reported depression-related symptoms, and did not have any nonremovable metal other than medical grade titanium in or around the head. Demographic and treatment details of the ITOC are summarized in Table 1, taken from Tienkamp, Rebernik, Halpern, et al. (2024).

Listeners

A total of 40 inexperienced listeners were recruited to participate in the perceptual evaluation. All participants completed an age-appropriate bilateral pure-tone hearing screening at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz (protocol based on the American Speech-Language-Hearing Association, 2005). For participants under 50 years of age, the cutoff was 25 dB HL at all tested frequencies, whereas for participants above 50 years of age, the cutoff was 25 dB HL under 1000 Hz and 40 dB HL for 2000, 4000, and 8000 Hz (Schow, 1991). Five listeners were excluded from the analysis as they either did not finish the experiment (n = 1) or did not pass the hearing screening (n = 4), so a total of 35 listeners completed the experiment (13 males, 22 females; age = 21–65 years, M = 36.1 years, SD = 13.7). Inexperienced listeners were chosen as their perceptual judgments reflect social communicative settings to a greater extent than SLPs. Moreover, inexperienced listeners' ratings correlate well with those provided by SLPs, including in studies using speech from ITOC (Halpern et al., 2023; Hirsch et al., 2022). All listeners denied a history of neurological or speech disorders. All but one participant indicated that they did not have any people in their direct surroundings with speech or voice disorders.[1] All listeners gave written informed consent, and the study was issued a letter of no objection by the Research Ethics Review Committee of the Faculty of Arts at the University of Groningen (ID95072353).

[1]This participant indicated to have an individual who stutters in their direct surroundings.

Data Collection

Acoustic and Kinematic Speech Recordings

Speakers were recorded in a sound-attenuated room in the mobile laboratory SPRAAKLAB parked either at the University of Groningen or the participant's home (Wieling et al., 2023). Acoustic data were recorded with a Sennheiser ME66 shotgun microphone placed approximately 20 cm from the participant's mouth with a sampling frequency of 22050 Hz and digitized using a Tascam US 4 × 4 sound card. Time-aligned kinematic data were acquired in parallel at 400 Hz with the NDI-Vox articulograph (see Rebernik, Jacobi, Tiede, & Wielding, 2021) using a MATLAB interface linked to the VOX-VRI recording software (Northern Digital Inc., 2019). Sensors were placed following the procedures as specified in Rebernik, Jacobi, Jonkers, et al. (2021). In total, five measurement sensors were placed, but only the two tongue sensors were used for the present study as these are the most relevant in vowel production. The TT sensor was placed 1 cm behind the anatomical tip of the tongue, and the TB sensor was placed at the speaker's /k/ constriction. The /k/ constriction was determined by having the speaker produce /k/ after a palate trace was made using a color transfer applicator stick (Dr. Thompson's; GUNZdental). The color of the applicator stick would mark the tongue at the place where the tongue touched the palate to form the constriction needed for /k/. In addition to the measurement sensors, four reference sensors (on the left and right mastoids, the nasal bridge, and the gingiva above the left upper incisor) were placed to filter out head movements. A bite plate recording was made to head-correct and rotate the data to the occlusal plane in MATLAB Version 2020a (MathWorks, 2023).

All speakers read the Dutch North Wind and the Sun Passage (see Appendix A) while wearing the EMA sensors and having their speech recorded via the microphone. The North Wind and the Sun Passage was selected from the corpus for two reasons. First, the passage is widely used in clinical linguistics, which ensures compatibility across studies. Second, the passage was recorded early during the data collection session, which increased the likelihood of sensors remaining securely attached and decreased the likelihood of any fatigue effects. The passage consisted of eight separate sentences that ranged from eight to 26 words (M = 14.6 words, SD = 5.5 words).

Perceptual Evaluation

To assess perceptual speech outcomes, each individual sentence of the North Wind and the Sun Passage was used in a subsequent perceptual evaluation task. A visual analogue scale (VAS)–based perceptual task was built using a graphical user interface in MATLAB (Abur et al., 2019; MathWorks, 2023). VAS scores correlate well with orthographic transcriptions and allow for the assessment of both intelligibility and listening effort using the same method (Abur et al., 2019; Stipancic et al., 2016; Thompson et al., 2023). To normalize the amplitude of the acoustic signal and ensure that each sentence had the same average sound pressure level, we calculated the average root-mean-square and scaled the amplitude of each sentence's acoustic signal accordingly. To avoid speaker and content familiarization effects, listeners did not rate the same sentence or speaker twice. This resulted in 17 lists with eight sentences each. With 35 listeners, each unique sentence was rated by two listeners, and the sentences of one list were rated by three listeners. Previous work indicated that in paradigms with only minimal exposure, such as the paradigm in the present study, two listeners already provide a strong predictor of intelligibility (Abur et al., 2019). We included multispeaker babble of four male and four female speakers who did not participate in the current study in order to imitate more naturalistic communication settings and reduce ceiling effects for typical speakers (Abur et al., 2021; Bunton, 2006; Tjaden, Sussman, & Wilding, 2014). The babble was mixed into the sentences at a signal-to-noise ratio of +2 dB, which was determined through pilot testing with a similar VAS rating procedure.

Listeners were tested in a quiet room or in the sound booth within the mobile laboratory SPRAAKLAB (Wieling et al., 2023). Experimental stimuli were presented at 65–70 dB SPL through Sennheiser 280 Pro headphones and could be listened to twice. After listening to a stimulus, the listener rated the sample on two 100-mm VAS scales, one for intelligibility and the other for listening effort. Intelligibility was described to the listener as the "degree to which speech is understood" (Abur et al., 2019; Kent et al., 1989). The intelligibility scale ranged from 0 (completely unintelligible) to 100 (completely intelligible). Listening effort was described as the "amount of work, attention or concentration it takes to understand a speech sample" (Eadie et al., 2021). The listening effort scale ranged from 0 (no effort at all) to 100 (extremely effortful). To facilitate the rating, the 0 and 100 anchors with their descriptions as well as a sheet with the definition of the two percepts were provided to the listeners (see Appendix B for the Dutch definitions of the percepts provided to the listeners). Additionally, three stimuli were randomly selected to be repeated at the end to assess within-rater reliability.

Self-Reported Speech Outcomes

Self-reported speech outcomes in ITOC were collected using the 15-item Dutch SHI, which is a validated translation of the French 15-item SHI (Fichaux-Bourin et al., 2009; Van den Steen et al., 2011). It comprises questions of the original 30-item questionnaire specifically designed for patients with oral and oropharyngeal cancer by Rinkel et al. (2008). The 15 items were answered using a 5-point Likert scale that ranged from *never* (0 points) to *always* (4 points). The scores on the SHI can range from 0 to 60, with higher scores indicating higher levels of self-reported speech problems. In its validation, Van den Steen et al. (2011) found a mean SHI score of 5 for control speakers. Mendoza Ramos et al. (2021) proposed the following cutoff scores for the degree of self-reported speech impairments: < 14 no impact, 14–22 light, 23–31 moderate, and > 31 severe impact. The total score of the SHI was used for statistical analysis.

Analysis

Acoustic Analysis: AAVS

The speakers' vowel space was quantified using the AAVS metric (Whitfield & Goberman, 2014). There were two reasons for this methodological choice. First, the AAVS considers all vowels rather than just the corner vowels, providing a more complete picture of a speaker's vowel space. Second, the AAVS can be quantified over running speech, which alleviates the need for manual segmentation of the corner vowels (but see Sandoval et al., 2013; van Son et al., 2018). Instead of using point-based measures, the AAVS provides a range-of-movement measure based on the entire F_1–F_2 trajectory of all voiced segments of an utterance. The AAVS has been used in both typical speakers, as well as individuals with dysarthria to assess articulatory–acoustic differences between speaker groups (Dragicevic et al., 2024; Houle et al., 2024; Whitfield et al., 2018; Whitfield & Goberman, 2014, 2017; Whitfield & Mehta, 2019).

In the case where a participant made a mistake and read the sentence again, only the correctly uttered version was used. To extract the utterance-level formant trajectories, we used the semi-automatic pipeline developed in Tienkamp, Rebernik, Buurke, et al. (2024). First, all voiceless segments from each sentence were removed using a custom script in Praat 6.3.1 (Boersma & Weenink, 2023). Formant frequency measurements for all voiced segments over time were extracted using a custom Praat script based on Carignan (2022). As the optimal formant ceiling may differ per speaker and vowel (see, e.g., Escudero et al., 2009), the script by Carignan (2022) collects formant values by iterating through an utterance with time steps of 5 ms, a window length of 25 ms, and 50-Hz steps within a 3500- to 6000-Hz formant ceiling range. This results in 51 possible formant frequency values for each 5-ms time step, one associated with each formant ceiling. From these collected formant frequencies, outliers 2 *SD*s away from the mean were removed. The median of the remaining formant values was taken as the optimal formant frequency of a specific 5-ms time frame, representing a single data point. Resulting formant tracks were manually checked, and remaining mistrackings were removed (45 data points, 0.06%).

The AAVS was then calculated for each utterance in R Version 4.3.2 (R Core Team, 2024) using the procedure described by Whitfield and colleagues (Whitfield et al., 2018; Whitfield & Goberman, 2014). First, we converted the formant data from Hertz to mels and computed the squared standard deviation of the F_1–F_2 tracks. Next, the unshared variance between the F_1–F_2 tracks was computed by subtracting the R^2 from 1 when fitting a linear model with F_1 predicting F_2. The AAVS in mels^2 was calculated as the square root of the product of the squared variance of the formant tracks and the unshared variance between them.

Kinematic Analysis: AKVS

To calculate the AKVS, the kinematic equivalent of the AAVS, we removed pre- and postutterance silences, as well as sections with audible swallowing from the correctly uttered versions in *mview* (developed by Mark Tiede, Haskins Laboratories). To calculate the AKVS, we used a similar procedure as the one for the AAVS, but with kinematic trajectories rather than acoustic ones. Instead of using the F_1–F_2 tracks, the coordinates representing anteroposterior and vertical movement were extracted from the TT and TB sensors per trimmed utterance in MATLAB and entered into the R script. The data from the TT and TB sensors were not decoupled from the jaw considering the complexity surrounding decoupling tongue and jaw data and because the entire vocal tract configuration impacts the acoustic signal. Following methods in prior work, both voiced and voiceless segments were used in the calculation of the AKVS (Abur et al., 2022; Whitfield et al., 2018). In a validation study, Whitfield et al. (2018) found correlations between the AAVS and AKVS that were comparable to those between the acoustic and kinematic VSA (e.g., Lee et al., 2016). The AKVS is reported in square millimeters. For the AAVS, AKVS of the TT (AKVS-TT), and AKVS of the TB (AKVS-TB), a total of 136 data points per measure were obtained (17 speakers × 8 sentences).

Perceptual Analysis: Intelligibility and Listening Effort

For each unique sentence, we calculated the mean rating of the two or three ratings and used this for our subsequent statistical analysis, which resulted in 136 data points (17 speakers × 8 sentences). To assess within-rater reliability, we calculated the intraclass correlation coefficient (ICC) together with its 95% confidence interval (CI) in R Version 4.4.2 (R Core Team, 2024) using the *ICC*() function of the *psych* package Version 2.4.6.26 (Revelle, 2024). We used a two-way mixed effects model (ICC3,1) to calculate the agreement between a rater's first and second ratings of the three randomly selected reliability trials. The within-rater reliability was moderate for both the intelligibility (ICC = .72, CI [.61, .80], $p < .001$) and listening effort ratings (ICC = .62, CI [.49, .73], $p < .001$).

To assess between-raters reliability, we computed the ICC together with its 95% CI with a two-way random effects agreement model ICCs (2,*k*). Considering that not all stimuli received the same number of ratings, we assessed between-raters reliability for lists that received two and three ratings separately. Based on Koo and Li (2016), we interpret ICC values below .5 as poor, between .5 and .75 as moderate, between .75 and .9 as strong, and above .9 as excellent. To be included in the subsequent analysis, at least moderate agreement was required (ICC > .5).

For the lists that received two ratings, the average-measure ICC indicated that between-raters reliability for the intelligibility ratings was moderate (ICC = .7, CI [.58, .79], $p < .001$). For the listening effort ratings, the average-measure ICC indicated poor between-raters reliability (ICC = .48, CI [.26, .63], $p < .001$). For the list that received three ratings, the average-measure ICC indicated that between-raters reliability for the intelligibility ratings was excellent (ICC = .92, CI [.76, .98], $p < .001$). For the listening effort ratings, the average-measure ICC indicated moderate agreement (ICC = .71, CI [.14, .93], $p < .001$). Due to the poor reliability and the strong correlation between intelligibility and listening effort ratings ($r = -.7$, $p < .001$), we did not include the listening effort ratings in our subsequent statistical analyses.

Statistical Analysis

After extracting all measures, data were processed and analyzed using R Version 4.3.2 (R Core Team, 2024). Data cleaning was done using the *tidyverse* package Version 2.0.0 (Wickham et al., 2019), and visualizations were made using the *ggplot2* package Version 3.5.1 (Wickham, 2016).

The first aim was to assess whether the acoustic measure (AAVS), the kinematic measures (AKVS-TT and AKVS-TB), and the perceptual measure (intelligibility) differed between typical speakers and ITOC. To this end, we fitted a linear mixed-effects regression model using the *lmer()* function of the *lme4* package Version 1.1.35.3 (Bates et al., 2015). We *z*-transformed the data of each measure to get the measures on the same scale. Our hypothesis model had the *z* score as the outcome variable and group (ITOC, control) as the predictor variable. To account for the inherent variability between speakers and the possibility that the measures might be differentially affected in individual speakers, we added a by-speaker random intercept and a by-speaker slope for measure. To account for the variability between sentences (e.g., in terms of their phonetic make-up and length), we included a by-sentence random intercept. A by-sentence slope for group or measure led to singular fits and was therefore not included. We modeled individuals treated for tumors located on the tongue and jaw as a single group as an initial analysis indicated that the two subgroups did not differ significantly from each other.

In our subsequent exploratory analysis, we assessed whether adding measure (AAVS, AKVS-TT, AKVS-TB, intelligibility), and the interaction between measure and group, improved the fit of the model to assess whether specific speech domains were impacted more than others. We also assessed the effect of sex (male, female) and age as these variables are known to affect acoustic and kinematic measures of speech (Kent & Vorperian, 2018; Thies et al., 2022). Due to the small sample size and the relative confound between tumor location and sex (e.g., more female speakers were treated for jaw tumors than male speakers and vice versa), we refrained from fitting any complex interactions that included sex. The best model was determined by using the *anova()* function, where a *p* value below .05 indicated that the more complex model was preferred.

Model criticism was employed by refitting the final model to the original data excluding outliers. That is, data points for which the absolute value of the residuals of the original model exceeded 2 *SD*s were removed. This retained approximately 95% of the data. In order to not report results dependent on outliers, the trimmed data set was used if and only if outliers drove the absence or presence of significant effects (Baayen, 2008, Chapter 6.2.3). The 95% CIs were computed for the final model by means of a bootstrap analysis with 1,000 simulations using the *boot* package Version 1.3.31 (Canty & Ripley, 2022). Model checks were performed to confirm that the model adhered to the assumptions surrounding multicollinearity, autocorrelation, normality of the residuals, and homoscedasticity using the *car* package Version 3.1.2 (Fox & Weisberg, 2019). The α level was set a priori at .05.

The second aim was to assess the interrelationships between acoustic (AAVS), kinematic (AKVS-TT, AKVS-TB), perceptual (intelligibility), and self-reported (SHI) measures of speech between and within ITOC. To assess between speaker interrelationships, we ran a Spearman's rank correlation test on the mean values of each domain per speaker (e.g., the mean AAVS over the eight sentences). To assess within-speaker interrelationships, we ran a Spearman's rank correlation test separately for each speaker. Each correlation test was based on the data points of the eight sentences. All correlation tests were run using the *corr.test()* function of the *psych* package Version 2.4.6.26

Table 2. Mean and standard deviation for each outcome measure for each individual who received surgical treatment for oral squamous cell carcinoma and the control and individuals treated for oral squamous cell carcinoma (ITOC) groups as a whole.

Participant	Subgroup	SHI	AAVS in mels^2	AKVS-TT in mm^2	AKVS-TB in mm^2	Intelligibility
Females						
Controls	—	—	20,547 (6,268)	6.8 (2.7)	5.0 (1.6)	83.1 (15.7)
ITOC	—	14.8 (9.9)	16,006 (2,810)	6.92 (2.4)	4.63 (2.1)	50.7 (29.4)
NKI05	Jaw	6	17,108 (3,433)	8.10 (1.5)	3.85 (1.1)	63.2 (32.5)
NKI11	Jaw	31	15,688 (2,764)	6.31 (1.8)	6.39 (1.9)	34.7 (21.6)
NKI16	Jaw	9	16,849 (2,556)	4.57 (0.9)	2.77 (0.6)	44.7 (28.9)
NKI06	Tongue	13	14,380 (1,919)	8.71 (2.4)	5.53 (2.1)	60.0 (29.4)
Males						
Controls	—	—	19,941 (6,292)	13.1 (3.7)	9.65 (4.8)	73.9 (24.5)
ITOC	—	20.5 (8.3)	12,019 (3,951)	6.36 (3.0)	5.82 (2.8)	34.9 (31.3)
NKI02	Jaw	24	11,980 (2,827)	9.85 (3.7)	9.05 (3.1)	11.6 (8.30)
NKI15	Jaw	18	12,590 (3,280)	8.34 (1.0)	5.92 (0.9)	55.1 (28.7)
NKI04	Tongue	29	9,315 (1,014)	3.62 (1.1)	2.68 (0.6)	14.8 (8.2)
NKI17	Tongue	26	8,499 (1,672)	5.09 (1.5)	4.28 (1.3)	16.6 (15.6)
NKI18	Tongue	6	17,709 (2,126)	4.91 (1.1)	7.15 (1.8)	76.5 (18.8)

Note. SHI = speech handicap index (ranges from 0 to 60); AAVS = articulatory–acoustic vowel space; AKVS = articulatory–kinematic vowel space; TT = tongue tip; TB = tongue back.

(Revelle, 2024). Given the exploratory nature and small clinical sample size of our study ($n = 9$), we used nonparametric tests and focused on effect size rather than significance. That is, only correlations of moderate strength ($|r_s| > .4$) were interpreted. Our analysis code can be found online at https://osf.io/3sprj/.

Results

Group Differences Among Speech Domains

Descriptive statistics for all speech measures are provided in Table 2 and visualized in Figure 1. Our final model used the entire data set, as model criticism indicated that the results were not affected by outliers. Across speech measures, ITOC had significantly lower scores compared to control speakers ($\beta = -.99$ SD, $t = -5.94$, CI [−1.39, −0.63], $p < .001$). In our exploratory analysis, we found that the addition of age significantly improved the fit of our model, $\chi^2 = 6.6(1)$, $p = .01$. The significant effect of age (age range: 47–77 years) indicated that scores increased significantly with age ($\beta = .03$ SD, $t = 2.75$, CI [0.006, 0.05], $p = .02$).[2] The addition of measure, $\chi^2(3) = 0$, $p = 1$; sex, $\chi^2(1) = 0.01$, $p = .92$; or the interaction between group and measure, $\chi^2(6) = 0.9$, $p = .99$, did not improve the fit of the model and were therefore not included. Thus, our final model indicates that, compared to control speakers, ITOC had a significantly smaller acoustic (AAVS) and kinematic (AKVS-TT and AKVS-TB) vowel space. Moreover, the speech of ITOC received lower intelligibility ratings as compared to control speakers. The magnitude of these differences did not differ significantly between the different measures.

Between-Speakers Correlation Results

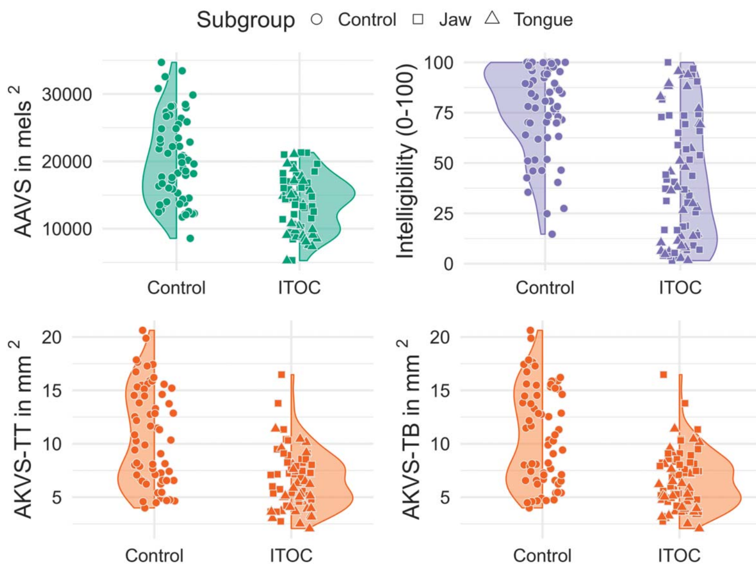
The Spearman's rank correlation tests revealed strong statistically significant negative correlations between perceptual and self-reported measures ($r_s = -.79$, $p = .01$) and self-reported and acoustic measures ($r_s = -.73$, $p = .03$). There was also a strong statistically significant positive correlation between acoustic and perceptual measures ($r_s = .8$, $p < .01$). Lastly, there was a moderate positive correlation between the two kinematic measures that was not statistically significant ($r_s = .6$, $p = .09$). Figure 2 visualizes all group-level, between-speaker correlation results. Scatter plots for all between-speaker variable combinations are available in Supplemental Material S1.

Within-Speaker Correlation Results

The within-speaker correlations revealed no strong statistically significant correlations between the acoustic and perceptual domain, with correlations ranging between −.26 and .62, with only a moderate-to-strong (.62) positive correlation for one speaker (all ps > .05). In contrast,

[2]This effect seems to be primarily driven by the fact that the youngest speaker, NKI04 (ITOC), received the lowest scores and the oldest speaker, NKI07 (control), received the highest scores. The age effect was not significant when only NKI04 was removed ($p = .098$), not significant when only NKI07 was removed ($p = .08$), and not significant when both NKI04 and NKI07 were removed ($p = .51$).

Figure 1. Violin plots depicting the distribution of the data for each speech measure. The acoustic AAVS in green, the perceptual measure intelligibility in purple, and the kinematic AKVS for the tongue tip and tongue back in orange. Dots represent individual data points and are shaped by subgroup (Circles = control speakers. Squares = individuals treated for jaw tumors. Triangles = individuals treated for tongue tumors). AAVS = articulatory–acoustic vowel space; AKVS = articulatory–kinematic vowel space; TOC = individuals treated for oral squamous cell carcinoma; TB = tongue back; TT = tongue tip.



where there was no group-level association between acoustic and kinematic domains, six speakers (66%) showed moderate-to-strong positive correlations for either the AKVS-TT (r_s between .14 and .57) or AKVS-TB (r_s between .17 and .7), though all but one correlation (r_s = .7) did not reach statistical significance. No statistically significant within-speaker association was found between kinematic and perceptual domains with effect sizes ranging from negative moderate to positive moderate (r_s between −.52 and .48, all $ps > .05$). Lastly, both kinematic measures were strongly positively correlated with each other for five speakers (56%; r_s between .83 and .98, all $ps < .01$) and moderately positively for one speaker (r_s = .62, p = .1). Figure 3 visualizes all within-speaker correlation results. Scatter plots for all variable combinations for each speaker are available in Supplemental Materials S2–S10.

Discussion

The goal of this study was to provide a comprehensive assessment of the speech outcomes of ITOC. To this end, we compared a measure of the acoustic vowel space (AAVS), measures of the kinematic vowel space (AKVS-TT and AKVS-TB), and a perceptual measure (intelligibility) between ITOC and control speakers. In addition, we assessed the interrelatedness between and within ITOC between the acoustic, kinematic, perceptual, and self-

Figure 2. Correlation matrix of the between-speaker interrelatedness of speech domains. Correlation coefficients are based on a Spearman's rank correlation test on the mean values of each domain per speaker. Asterisks denote a $p < .05$. AAVS = articulatory–acoustic vowel space; AKVS = articulatory–kinematic vowel space; PRO = Patient Reported Outcome; SHI = speech handicap index; TB = tongue back; TT = tongue tip.

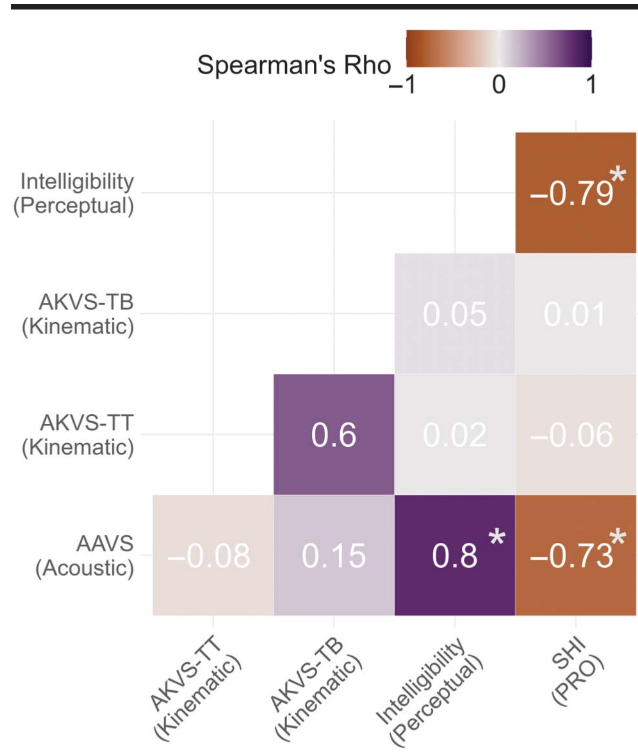


Figure 3. Correlation matrix of the within-speaker interrelatedness of speech domains in ITOC. Correlation coefficients are based on a Spearman's rank correlation test on the speech outcomes across domains on the eight sentences for each speaker. Asterisks denote a $p < .05$. AAVS = articulatory–acoustic vowel space; AKVS = articulatory–kinematic vowel space; TB = tongue back; TT = tongue tip.

reported speech domains in ITOC to inform clinically relevant measures of speech.

Group Comparisons

For our first aim, we compared speech outcome measures between ITOC and control speakers and hypothesized that ITOC, regardless of treatment site, would have a smaller acoustic vowel space, kinematic vowel space, and lower speech intelligibility ratings. In line with our hypothesis and prior work, ITOC had a smaller AAVS (de Bruijn et al., 2009; Guo et al., 2023; Takatsu et al., 2017; Whitehill et al., 2006), AKVS of both tongue sensors (Chepeha et al., 2016; de Groot et al., 2020; Hagedorn et al., 2021; Kappert et al., 2019; Speksnijder et al., 2011; van Dijk et al., 2016), and lower speech intelligibility scores compared to typical speakers (Constantinescu et al., 2017; Loewen et al., 2010). Considering that the addition of measure as a fixed effect or the interaction between measure and group did not improve the fit of the model, our results suggest that the observed difference between ITOC and control speakers remained consistent across the different assessed speech domains.

Earlier studies have already shown that ITOC have reduced tongue mobility compared to controls, either in nonspeech tasks (e.g., Kappert et al., 2019; van Dijk et al., 2016) or in a speech task measuring the absolute movement size (i.e., the articulatory working space in mm^2; Tienkamp et al., 2023). Our work extends these findings by showing that the movement area that ITOC tend to use the most during speech (i.e., the generalized working space) is also reduced and more centralized compared to typical speakers. A reduced working space can result in a less clear distinction between different articulatory gestures.

While the acoustic findings of the current work largely align with prior work (e.g., de Bruijn et al., 2009; Guo et al., 2023; Takatsu et al., 2017; Whitehill et al., 2006), they are in contrast with the findings by Laaksonen et al. (2010), who found no significant VSA reduction postsurgery compared to presurgery, and Tienkamp et al. (2023), who found no significant differences in the F_1 and F_2 of corner vowels between ITOC and controls. This discrepancy is likely due to differences in tumor and resection size. That is, half of our sample included speakers treated for large tumors (T3–T4), whereas 80% of Laaksonen and colleagues' sample had smaller T2 tumors. Additionally, most of our participants underwent a hemiglossectomy, whereas in Tienkamp et al. (2023), two of the five speakers received only a partial glossectomy.

Our finding that ITOC had lower speech intelligibility scores compared to typical speakers also highlights the long-term effects of surgical treatment for more advanced OSCC. The ITOC in our study, who were primarily treated for advanced OSCC at least 1 year ago (M = 5.1 years), scored significantly lower compared to control speakers. Our results contrast with speech intelligibility scores of individuals treated for smaller tumors, as they reach close to preoperative levels a year following surgery or radiation (for systematic reviews, see Jacobi et al., 2010; Lam & Samman, 2013).

Overall, our results showed that ITOC had worse speech outcomes compared to typical speakers across all assessed speech domains. While deficits in acoustic, kinematic, perceptual, and self-reported speech domains have been established across prior work assessing one (e.g., Hagedorn et al., 2021; Laaksonen et al., 2010; Tienkamp et al., 2023), two (e.g., Guo et al., 2023; Park et al., 2016), or three (Acher et al., 2014; Wakumoto et al., 1996) speech domains, no study has assessed all four domains within the same speaker group. Comparing the four speech domains in the same group of ITOC is important because it remains relatively underexplored to what extent speech problems found in one domain are associated with problems in other speech domains. Our results directly show the concurrence and relatedness of speech problems across all four speech domains and further underscore the need for the development of a comprehensive speech assessment protocol for ITOC in a research setting, tailored to address affected speech domains separately. Such a protocol could help to tease apart acoustic and kinematic contributions to reduced functional speech outcomes, which in turn could help to improve speech rehabilitation strategies. Moreover, a standardized protocol could help to provide better generalizability across studies. Reducing variability in speech assessment methods is desirable given the considerable variability in speech outcomes in ITOC.

While the absence of a standardized assessment protocol for ITOC has been problematized by Schuster and Stelzle (2012) and some guidelines have been provided by Clarke et al. (2016), no standardized protocol for assessing speech outcomes in ITOC has been developed; this is in part due to the lack of research on comprehensive speech outcomes in ITOC. Hence, the current work adds to the body of literature on speech outcomes in ITOC and is the first work to demonstrate specific outcomes across the acoustic, kinematic, perceptual, and self-reported domains. Clarke et al. (2016) recommend the inclusion of both oral–motor (e.g., a tongue range of motion task) and articulatory strength and precision examination in ITOC, but they do not recommend any specific tasks. Since the established differences between ITOC and typical speakers in our study were found in all assessed speech domains, our results support the inclusion of acoustic, kinematic, and perceptual measures of speech to determine articulatory strength and precision in ITOC. Moreover, our results support the inclusion of a self-reported outcome measure considering the mean SHI score in our study (17.9) is well above that of the mean of the typical controls (5.0) reported in Van den Steen et al. (2011).

Interrelatedness of Speech Domains

Our second aim was to assess the interrelatedness between the acoustic, kinematic, perceptual, and self-reported speech domains between and within ITOC. Our between-speaker correlation results indicated strong associations between the acoustic (AAVS), perceptual (intelligibility), and self-reported (SHI) domains for ITOC. This is in line with previous research that found correlations between articulatory–acoustics, quantified by the VSA/AAVS, and perceptual measures of speech, quantified through speech intelligibility/clarity ratings (Turner et al., 1995; Weismer et al., 2001; Whitfield & Goberman, 2014, 2017). A larger acoustic working space allows for a greater distinction between individual phonemes, which makes speech more intelligible as the chance of misidentifying a particular phoneme becomes smaller. The association between the SHI, speech intelligibility, and the AAVS further agrees with prior work that found associations between self-reported outcomes and speech intelligibility (Park et al., 2016) and the size of the VSA (Guo et al., 2023). Reduced intelligibility of speech may lead to reduced participation in social interactions, especially those in nonoptimal listening conditions (e.g., loud cafés), which may negatively affect someone's quality of life. While the SHI does not measure social participation directly, the scores of the SHI are known to correlate very strongly (r = .89) with scores from the Communicative Participation Item Bank questionnaire (Baylor et al., 2013, 2021; van Sluis et al., 2023). Given the relationship between the size of the AAVS and speech intelligibility, it is perhaps not surprising that a larger AAVS is also related to better speech-related quality of life.

In contrast to our between-speaker results, the within-speaker correlations did not show similar patterns. While a strong statistically significant positive correlation was found between speech intelligibility and the acoustic AAVS between speakers, only one speaker (NKI02; T2, continuity resection with fibular reconstruction) showed a moderate-to-strong positive correlation between speech intelligibility and the acoustic AAVS, though it did not reach statistical significance. The strong correlations observed between different speech domains across individual speakers likely reflect differences in overall speech severity, as more severe speakers consistently scored on the lower end of the spectrum across multiple domains. In contrast, within-speaker correlations assess speech domains along the same level of speech severity by definition, which might limit the observed variability across domains if unique stimuli are used (i.e., no repetitions of the same material). Although our stimuli captured sufficient variability to detect associations between the articulatory–acoustic and articulatory–kinematic domains for six speakers, acoustic variation may have been too subtle to influence intelligibility ratings. To elicit more robust within-speaker variation to assess interrelatedness, future work could include repetitions of the same material with different speech style modifications (e.g., loudness, speed, or clarity) as these style modifications have been shown to result in associations between speech domains in typical speakers, but also in those with dysarthria (Mefferd, 2015; Mefferd & Green, 2010; Tjaden et al., 2013; Tjaden, Sussman, & Wilding, 2014; Whitfield et al., 2018; Whitfield & Goberman, 2017).

Given that surgical or radiation-based treatments for OSSC can result in reduced speech function compared to presurgery and that our between-speaker correlation results were largely driven by speech severity, longitudinal assessment of speech domains in ITOC may also show more robust within-speaker associations. If the AAVS can capture treatment-induced differences in intelligibility or self-reported speech problems within a single speaker, it might serve as a clinically meaningful acoustic outcome measure. Moreover, as the AAVS can be computed (semi) automatically over running speech using our pipeline, the measure also shows clinical applicability.

Our between-speaker results did not reveal a strong association between the acoustic (AAVS) and kinematic (AKVS) vowel space measures in ITOC, which was not in line with our prediction and earlier work that found a relationship between the acoustic and kinematic VSA in typical and dysarthric speakers (Lee et al., 2016, 2017; Mefferd, 2015; Whitfield et al., 2018). Thus, a decrease in the AAVS was not necessarily associated with a decrease in AKVS-TT or AKVS-TB in ITOC. One possible explanation is that the relationship between the AAVS and AKVS has only been verified for typical speakers, and not those with atypical speech, such as ITOC (Whitfield et al., 2018). While relationships between the acoustic and kinematic VSA have been documented in individuals with Parkinson's disease and amyotrophic lateral sclerosis (Lee et al., 2017; Mefferd, 2015), these speakers had mechanically intact speech–motor systems, whereas the ITOC included in this study had anatomical alterations as a result of surgery. Paired with the possibility of using the lips in a compensatory manner and the notion that compensatory mechanisms may differ as a function of tumor size and location, the relationship between articulatory–acoustics and kinematics between-speaker in ITOC might be less straightforward compared to populations studied in prior work.

A second explanation is that the relationship between articulatory–acoustics and kinematics has been strong for within-speaker comparisons but more inconsistent for between-speaker comparisons due to individual differences in vocal tract morphology and the presence of motor equivalence strategies (Lee et al., 2016, 2017; Whitfield et al., 2018). EMA provides considerably sparse information as articulatory movements are only tracked by single sensor coils on the tongue, jaw, and lips, which may not fully capture individual differences in vocal tract morphology or individualized motor equivalence strategies (Kuo & Berry, 2023; Mefferd, 2017). The findings of the current work are in line with this notion, as six (66%) speakers showed at least moderate correlations between the AAVS and the AKVS of the TT or TB, with correlations ranging between .43 and .71. Of the six speakers, four were treated for a jaw tumor, comprising 80% of the jaw group, whereas two were treated for a tongue tumor, representing 50% of the tongue group. While very preliminary, these results suggest that the relationship between articulatory–acoustics and kinematics is compromised more for those treated for tongue tumors. Surgical modification of the tongue might result in more atypical movement compared to treatment of the jaw, where the tongue musculature remains mostly unaltered, though mobility issues might arise due to resection of the FOM or tethering of the tongue to the FOM. However, these results need to be replicated in future work with a larger sample size.

Four speakers (44%; two with a continuity resection, one with a hemiglossectomy, and one speaker with multiple resections on the mandible, tongue, and oropharynx) showed near-perfect positive correlations (r_s > .93) between the movement size of the TT and the TB, with one additional speaker showing a strong positive correlation of .83. This finding may be in line with previous work that found that ITOC have difficulty in controlling different parts of the tongue in tandem (Acher et al., 2014; Hagedorn et al., 2021). While the AKVS is only able to quantify movement size and does not specify qualitative movement patterns, the TT and TB serve different functions in speech production, and one would not predict the movement size to overlap this much. For example, the correlation between TT and TB movement size for typical speakers was only .12 in our sample. The findings of the current work further suggest that near-perfect correlation of TT and TB movement size is associated with low speech intelligibility as the four speakers had the lowest speech intelligibility scores. As Hagedorn et al. (2021) note, natural speech requires the coordination of different parts of the tongue, as well as with other articulators as individual gestures may temporally overlap. If this coordination breaks down, increased overlap may occur, which could negatively affect speech intelligibility. Future work may quantify the rate of independent movement of different parts of the tongue in ITOC to further assess this possibility.

Limitations

This study provided a comprehensive assessment of the speech outcomes of ITOC compared to control speakers. Nevertheless, there were several limitations. Firstly, we did not collect any self-reported measures from the included control speakers, which precluded us from directly testing whether ITOC had worse self-reported speech compared to control speakers. However, considering that Van den Steen et al. (2011) reported a mean SHI score of 5 in 73 typical speakers, it can be assumed that the ITOC included in our study had worse self-reported speech outcomes compared to typical speakers as they had a mean SHI score of 17.9 (*SD* = 9.4). We further only analyzed sentence-level indices of movement size and did not focus on other global measures (e.g., speed or duration) or more fine-grained levels of analysis (e.g., individual phonemes). Future work may investigate whether the association between the acoustic and kinematic domains is stronger using different or more fine-grained levels of analysis.

Secondly, the perceptual ratings of intelligibility and listening effort provided by the inexperienced listeners had moderate or poor between-raters reliability. One potential explanation for the reduced reliability might be that listeners rated a single sentence for each speaker and only eight stimuli in total, which may have been too few to reach strong between-raters agreement. As listeners were not provided with examples or training, they might need additional stimuli to fully calibrate their ratings consistently, which may have led to additional variability across ratings as all stimuli were presented in a random order across participants. However, we opted for this methodological approach to avoid learning and familiarization effects. A second reason for the poor between-raters agreement for listening effort might be that additional factors affect perceived listening effort ratings. For example, differences in interpretation of the instruction by individual speakers might cause additional disagreement (Alhanbali et al., 2019). Moreover, ratings of listening effort are also influenced by speaker characteristics, such as intelligibility, speech rate, and voice quality (Borrie et al., 2012; Nagle, 2015; Whitehill & Wong, 2006). These speaker characteristics might affect some listeners more than others, resulting in lower agreement considering that most lists were rated by two listeners only. Still, our reliability metrics for speech intelligibility were slightly lower but comparable to that of earlier work (Kim & Kuo, 2011; Kuruvilla-Dugdale et al., 2019; Stipancic et al., 2016, 2023; Tjaden, Kain, & Lam, 2014; Tjaden, Sussman, & Wilding, 2014).

Thirdly, the ITOC included in our study formed a relatively small and heterogeneous cross-sectional cohort in terms of tumor size, location, and reconstruction technique. This resulted in additional variability in our data as outcomes following treatment for OSCC are highly variable (Bressmann, 2021), making it difficult to draw strong conclusions, especially in terms of the relationships between articulatory–acoustic and kinematic measures and self-reported and kinematic measures. Future research should investigate the effect of tumor location in a more controlled way and include a preoperative session to allow for a comprehensive within-speaker analysis rather than a comparison to a control group.

Conclusions

This study provided the first comprehensive assessment of the speech outcomes of individuals surgically treated for OSCC (ITOC) across multiple speech domains (acoustic, kinematic, and perceptual) compared to control speakers. Compared to age- and sex-matched control speakers, ITOC had smaller acoustic and kinematic vowel spaces and lower speech intelligibility ratings, demonstrating the multifaceted nature of the speech problems experienced by ITOC. In addition, the interrelatedness of speech domains was assessed between and within ITOC. Between-speaker correlations revealed strong relationships between the acoustic, perceptual, and self-reported domains but no relationship between acoustic and kinematic speech measures. Within-speaker correlations revealed no strong statistically significant relationship between acoustic and perceptual measures, but stronger associations between acoustic and kinematic measures were found. Collectively, our findings suggest that to provide the most comprehensive account of postoperative speech function in ITOC, speech assessment protocols should include both articulatory–kinematic and articulatory–acoustic measures of speech alongside perceptual and self-reported measures, as measures seem to provide complementary information.

Author Contributions

Thomas B. Tienkamp: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing. **Teja Rebernik:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Bence M. Halpern:** Conceptualization, Investigation, Methodology, Funding acquisition, Writing – review & editing. **Rob J. J. H. van Son**: Conceptualization, Methodology, Resources, Funding acquisition, Supervision, Writing – review & editing. **Martijn Wieling:** Conceptualization, Formal analysis, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Max J. H. Witjes**: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Sebastiaan A. H. J. de Visscher**: Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Defne Abur**: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Data Availability Statement

The full data set presented in this study is not available on request, as some participants did not provide consent to have their data shared. However, the acoustic, kinematic, and self-reported data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by a University of Groningen Center for Language and Cognition PhD grant (awarded to Thomas B. Tienkamp) and by the Research School of Behavioral and Cognitive Neurosciences of the University of Groningen. This work was further supported by a research grant from Atos Medical (Hörby, Sweden) awarded to the Department of Head and Neck Oncology and Surgery of the Netherlands Cancer Institute, which contributes to the existing infrastructure for quality of life research (Rob van Son; Bence M. Halpern). We are most grateful to all speakers and listeners who participated in this study and to the Faculty of Arts at the University of Groningen for providing the mobile sound booth (SPRAAKLAB) where the study data were collected.

References

Abur, D., Enos, N. M., & Stepp, C. E. (2019). Visual analog scale ratings and orthographic transcription measures of sentence intelligibility in Parkinson's disease with variable listener exposure. *American Journal of Speech-Language Pathology, 28*(3), 1222–1232. https://doi.org/10.1044/2019_AJSLP-18-0275

Abur, D., Perkell, J. S., & Stepp, C. E. (2022). Impact of vocal effort on respiratory and articulatory kinematics. *Journal of Speech, Language, and Hearing Research, 65*(1), 5–21. https://doi.org/10.1044/2021_JSLHR-21-00323

Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and feedforward auditory-motor processes for voice and articulation in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 64*(12), 4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153

Acher, A., Perrier, P., Savariaux, C., & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics & Phonetics, 28*(4), 241–256. https://doi.org/10.3109/02699206.2013.802015

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing, 40*(5), 1084–1097. https://doi.org/10.1097/AUD.0000000000000697

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry*. https://www.asha.org/policy/gl2005-00014/

Arslan, H. H., Ahmadov, A., Cebeci, S., Binar, M., & Karahatay, S. (2016). Life priorities in head and neck cancer patients between ages of 45 to 65. *Journal of Craniofacial Surgery, 27*(4), e398–e401. https://doi.org/10.1097/SCS.0000000000002671

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Bagan, J., Sarrion, G., & Jimenez, Y. (2010). Oral cancer: Clinical features. *Oral Oncology, 46*(6), 414–417. https://doi.org/10.1016/j.oraloncology.2010.03.009

Bak, M., Jacobson, A. S., Buchbinder, D., & Urken, M. L. (2010). Contemporary reconstruction of the mandible. *Oral Oncology, 46*(2), 71–76. https://doi.org/10.1016/j.oraloncology.2009.11.006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baylor, C., Eadie, T., & Yorkston, K. (2021). The Communicative Participation Item Bank: Evaluating, and reevaluating, its use across communication disorders in adults. *Seminars in Speech and Language, 42*(03), 225–239. https://doi.org/10.1055/s-0041-1729947

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research, 56*(4), 1190–1208. https://doi.org/10.1044/1092-4388(2012/12-0140)

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer program] (Version 6.20.06). https://praat.org

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55*(1), 290–305. https://doi.org/10.1044/1092-4388(2011/10-0349)

Bressmann, T. (2021). Speech disorders related to head and neck cancer: Laryngectomy, glossectomy, and velopharyngeal and maxillofacial defects. In J. S. Damico, N. Müller, & M. J. Ball (Eds.), *The handbook of language and speech disorders* (2nd ed., pp. 495–527). Wiley-Blackwell. https://doi.org/10.1002/9781119606987.ch22

Bressmann, T., Ackloo, E., Heng, C.-L., & Irish, J. C. (2007). Quantitative three-dimensional ultrasound imaging of partially resected tongues. *Otolaryngology—Head and Neck Surgery, 136*(5), 799–805. https://doi.org/10.1016/j.otohns.2006.11.022

Bressmann, T., Jacobs, H., Quintero, J., & Irish, J. C. (2009). Speech outcomes for partial glossectomy surgery: Measures of speech articulation and listener perception. *Canadian Journal of Speech-Language Pathology and Audiology, 33*(4), 204–210.

Bressmann, T., Sader, R., Whitehill, T. L., & Samman, N. (2004). Consonant intelligibility and tongue motility in patients with partial glossectomy. *Journal of Oral and Maxillofacial Surgery, 62*(3), 298–303. https://doi.org/10.1016/j.joms.2003.04.017

Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics & Phonetics, 19*(6–7), 573–588. https://doi.org/10.1080/02699200500113947

Bressmann, T., Uy, C., & Irish, J. C. (2005). Analysing normal and partial glossectomee tongues using ultrasound. *Clinical Linguistics & Phonetics, 19*(1), 35–52. https://doi.org/10.1080/02699200410001669834

Bunton, K. (2006). Fundamental frequency as a perceptual cue for vowel identification in speakers with Parkinson's disease. *Folia Phoniatrica et Logopaedica, 58*(5), 323–339. https://doi.org/10.1159/000094567

Canty, A., & Ripley, B. D. (2022). *boot: Bootstrap R (S-Plus) functions*. https://CRAN.R-project.org/package=boot

Carignan, C. (2022). *Formant optimization*. https://github.com/ChristopherCarignan/formant-optimization

Chan, H. F., Ng, M. L., Rosen, C. A., & Schneider, S. L. (2021). Cultural adaptation and validation of speech handicap index: A scoping review. *American Journal of Speech-Language Pathology, 30*(2), 748–760. https://doi.org/10.1044/2020_AJSLP-20-00236

Chepeha, D. B., Spector, M. E., Chinn, S. B., Casper, K. A., Chanowski, E. J. P., Moyer, J. S., Morrison, R., Carvill, E., & Lyden, T. H. (2016). Hemiglossectomy tongue reconstruction: Modeling of elevation, protrusion, and functional outcome using receiver operator characteristic curve. *Head & Neck, 38*(7), 1066–1073. https://doi.org/10.1002/hed.24417

Clarke, P., Radford, K., Coffey, M., & Stewart, M. (2016). Speech and swallow rehabilitation in head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology & Otology, 130*(S2), S176–S180. https://doi.org/10.1017/S0022215116000608

Constantinescu, G., & Rieger, J. M. (2019). Speech deficits associated with oral and oropharyngeal carcinomas. In P. C. Doyle (Ed.), *Clinical care and rehabilitation in head and neck cancer* (pp. 265–279). Springer International. https://doi.org/10.1007/978-3-030-04702-3_16

Constantinescu, G., Rieger, J. M., Winget, M., Paulsen, C., & Seikaly, H. (2017). Patient perception of speech outcomes: The relationship between clinical measures and self-perception of speech function following surgical treatment for oral cancer. *American Journal of Speech-Language Pathology, 26*(2), 241–247. https://doi.org/10.1044/2016_AJSLP-15-0170

de Bruijn, M. J., ten Bosch, L., Kuik, D. J., Quené, H., Langendijk, J. A., Leemans, C. R., & Verdonck-de Leeuw, I. M. (2009). Objective acoustic–phonetic speech analysis in patients treated for oral or oropharyngeal cancer. *Folia Phoniatrica et Logopaedica, 61*(3), 180–187. https://doi.org/10.1159/000219953

de Groot, R. J., Merkx, M. A. W., Hamann, M. N. S., Brand, H. S., de Haan, A. F. J., Rosenberg, A. J. W. P., & Speksnijder, C. M. (2020). Tongue function and its influence on masticatory performance in patients treated for oral cancer: A five-year prospective study. *Supportive Care in Cancer, 28*(3), 1491–1501. https://doi.org/10.1007/s00520-019-04913-y

Dragicevic, D. A., Dahl, K. L., Perkins, Z., Abur, D., & Stepp, C. E. (2024). Effects of a concurrent working memory task on speech acoustics in Parkinson's disease. *American Journal of Speech-Language Pathology, 33*(1), 418–434. https://doi.org/10.1044/2023_AJSLP-23-00214

Dromey, C., Jang, G.-O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55*(2), 315–328. https://doi.org/10.1016/j.specom.2012.09.001

Dwivedi, R. C., Kazi, R. A., Agrawal, N., Nutting, C. M., Clarke, P. M., Kerawala, C. J., Rhys-Evans, P. H., & Harrington, K. J. (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

Eadie, T. L., Durr, H., Sauder, C., Nagle, K., Kapsner-Smith, M., & Spencer, K. A. (2021). Effect of noise on speech intelligibility and perceived listening effort in head and neck cancer. *American Journal of Speech-Language Pathology, 30*(3S), 1329–1342. https://doi.org/10.1044/2020_AJSLP-20-00149

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America, 126*(3), 1379–1393. https://doi.org/10.1121/1.3180321

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer, 149*(4), 778–789. https://doi.org/10.1002/ijc.33588

Fichaux-Bourin, P., Woisard, V., Grand, S., Puech, M., & Bodin, S. (2009). Validation d'un questionnaire d'auto-évaluation de la parole [Parole Handicap Index]. *Revue de Laryngologie, d'otologie et de Rhinologie, 130*(1), 45–51.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Guo, K., Xiao, Y., Deng, W., Zhao, G., Zhang, J., Liang, Y., Yang, L., & Liao, G. (2023). Speech disorders in patients with tongue squamous cell carcinoma: A longitudinal observational study based on a questionnaire and acoustic analysis. *BMC Oral Health, 23*(1), Article 192. https://doi.org/10.1186/s12903-023-02888-1

Hagedorn, C., Kim, J., Sinha, U., Goldstein, L., & Narayanan, S. S. (2021). Complexity of vocal tract shaping in glossectomy patients and typical speakers: A principal component analysis. *The Journal of the Acoustical Society of America, 149*(6), 4437–4449. https://doi.org/10.1121/10.0004789

Hagedorn, C., Lu, Y., Toutios, A., Sinha, U., Goldstein, L., & Narayanan, S. (2022). Variation in compensatory strategies as a function of target constriction degree in post-glossectomy speech. *JASA Express Letters, 2*(4), Article 045205. https://doi.org/10.1121/10.0009897

Halpern, B. M., Feng, S., van Son, R., van den Brekel, M., & Scharenborg, O. (2023). Automatic evaluation of spontaneous oral cancer speech using ratings from naive listeners. *Speech Communication, 149,* 84–97. https://doi.org/10.1016/j.specom.2023.03.008

Halpern, B. M., Rebernik, T., Tienkamp, T. B., van Son, R., van den Brekel, M., Wieling, M., Witjes, M., & Scharenborg, O. (2022). *Manipulation of oral cancer speech using neural articulatory synthesis.* arXiv. https://doi.org/10.48550/arXiv.2203.17072

Hassan, S. J., & Weymuller, E. A. (1993). Assessment of quality of life in head and neck cancer patients. *Head & Neck, 15*(6), 485–496. https://doi.org/10.1002/hed.2880150603

Hirsch, M. E., Thompson, A., Kim, Y., & Lansford, K. L. (2022). The reliability and validity of speech-language pathologists' estimations of intelligibility in dysarthria. *Brain Sciences, 12*(8), Article 1011. https://doi.org/10.3390/brainsci12081011

Houle, N., Feaster, T., Mira, A., Meeks, K., & Stepp, C. E. (2024). Sex differences in the speech of persons with and without Parkinson's disease. *American Journal of Speech-Language Pathology, 33*(1), 96–116. https://doi.org/10.1044/2023_AJSLP-22-00350

Imai, S., & Michi, K.-I. (1992). Articulatory function after resection of the tongue and floor of the mouth: Palatometric and perceptual evaluation. *Journal of Speech and Hearing Research, 35*(1), 68–78. https://doi.org/10.1044/jshr.3501.68

Jacobi, I., van der Molen, L., Huiskens, H., van Rossum, M. A., & Hilgers, F. J. M. (2010). Voice and speech outcomes of chemoradiation for advanced head and neck cancer: A systematic review. *European Archives of Oto-Rhino-Laryngology, 267*(10), 1495–1505. https://doi.org/10.1007/s00405-010-1316-x

Kappert, K., van Alphen, M., Smeele, L., Balm, A., & van der Heijden, F. (2019). Quantification of tongue mobility impairment using optical tracking in patients after receiving primary surgery or chemoradiation. *PLOS ONE, 14*(8), Article e0221593. https://doi.org/10.1371/journal.pone.0221593

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders, 74,* 74–97. https://doi.org/10.1016/j.jcomdis.2018.05.004

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), 482–499. https://doi.org/10.1044/jshd.5404.482

Kim, Y., & Kuo, C. (2011). Effect of level of presentation to listeners on scaled speech intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica, 64*(1), 26–33. https://doi.org/10.1159/000328642

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kreeft, A. M., van der Molen, L., Hilgers, F. J., & Balm, A. J. (2009). Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: A systematic review of the literature. *European Archives of Oto-Rhino-Laryngology, 266*(11), 1687–1698. https://doi.org/10.1007/s00405-009-1089-2

Kuo, C., & Berry, J. (2023). The relationship between acoustic and kinematic vowel space areas with and without normalization for speakers with and without dysarthria. *American Journal of Speech-Language Pathology, 32*(4S), 1923–1937. https://doi.org/10.1044/2023_AJSLP-22-00158

Kuruvilla-Dugdale, M., Threlkeld, K., Salazar, M., Nolan, G., & Heidrick, L. (2019). A comparative study of auditory-perceptual speech measures for the early detection of mild speech impairments. *Seminars in Speech and Language, 40*(05), 394–406. https://doi.org/10.1055/s-0039-1694997

Laaksonen, J.-P., Rieger, J., Happonen, R.-P., 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*(1), 41–54. https://doi.org/10.3109/02699200903340758

Lam, L., & Samman, N. (2013). Speech and swallowing following tongue cancer surgery and free flap reconstruction—A systematic review. *Oral Oncology, 49*(6), 507–524. https://doi.org/10.1016/j.oraloncology.2013.03.001

Laraway, D. C., & Rogers, S. N. (2012). A structured review of journal articles reporting outcomes using the University of Washington Quality of Life Scale. *British Journal of Oral and Maxillofacial Surgery, 50*(2), 122–131. https://doi.org/10.1016/j.bjoms.2010.12.005

Lee, J., Littlejohn, M. A., & Simmons, Z. (2017). Acoustic and tongue kinematic vowel space in speakers with and without dysarthria. *International Journal of Speech-Language Pathology, 19*(2), 195–204. https://doi.org/10.1080/17549507.2016.1193899

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America, 139*(1), 426–440. https://doi.org/10.1121/1.4939894

Loewen, I. J., Boliek, C. A., Harris, J., Seikaly, H., & Rieger, J. M. (2010). Oral sensation and function: A comparison of patients with innervated radial forearm free flap reconstruction to healthy matched controls. *Head & Neck, 32*(1), 85–95. https://doi.org/10.1002/hed.21155

Lydiatt, W. M., Patel, S. G., O'Sullivan, B., Brandwein, M. S., Ridge, J. A., Migliacci, J. C., Loomis, A. M., & Shah, J. P. (2017). Head and neck cancers—Major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA: A Cancer Journal for Clinicians, 67*(2), 122–137. https://doi.org/10.3322/caac.21389

Mady, K., & Beer, A. (2007). Articulatory parameters in consonant production after tumour surgery: A real-time MRI investigation. *Archives of Acoustics, 32*(1), 135–145.

Malone, J. P., Stephens, J. A., Grecula, J. C., Rhoades, C. A., Ghaheri, B. A., & Schuller, D. E. (2004). Disease control, survival, and functional outcome after multimodal treatment for advanced-stage tongue base cancer. *Head & Neck, 26*(7), 561–572. https://doi.org/10.1002/hed.20012

MathWorks. (2023). *MATLAB 2023* (Version A). The MathWorks Inc. https://mathworks.com/?s_tid=gn_logo

Matsui, Y., Ohno, K., Yamashita, Y., & Takahashi, K. (2007). Factors influencing postoperative speech function of tongue cancer patients following reconstruction with fasciocutaneous/myocutaneous flaps—A multicenter study. *International Journal of Oral and Maxillofacial Surgery, 36*(7), 601–609. https://doi.org/10.1016/j.ijom.2007.01.014

Mefferd, A. S. (2015). Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis. *Journal of Speech, Language,* *and Hearing Research, 58*(3), 576–589. https://doi.org/10.1044/2015_JSLHR-S-14-0188

Mefferd, A. S. (2017). Tongue- and jaw-specific contributions to acoustic vowel contrast changes in the diphthong /ai/ in response to slow, loud, and clear speech. *Journal of Speech, Language, and Hearing Research, 60*(11), 3144–3158. https://doi.org/10.1044/2017_JSLHR-S-17-0114

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53*(5), 1206–1219. https://doi.org/10.1044/1092-4388(2010/09-0083)

Mendoza Ramos, V., Paulyn, C., Van den Steen, L., Hernandez-Diaz Huici, M. E., De Bodt, M., & Van Nuffelen, G. (2021). Effect of boost articulation therapy (BArT) on intelligibility in adults with dysarthria. *International Journal of Language & Communication Disorders, 56*(2), 271–282. https://doi.org/10.1111/1460-6984.12595

Nagle, K. (2015). Effect of intelligibility and speech rate on perceived listener effort. *The Journal of the Acoustical Society of America, 137*(Suppl. 4), 2433–2433. https://doi.org/10.1121/1.4920878

Nicoletti, G., Soutar, D. S., Jackson, M. S., Wrench, A. A., Robertson, G., & Robertson, C. (2004). Objective assessment of speech after surgical treatment for oral cancer: Experience from 196 selected cases. *Plastic and Reconstructive Surgery, 113*(1), 114–125. https://doi.org/10.1097/01.PRS.0000095937.45812.84

Northern Digital Inc. (2019). *Vox-EMA system user guide*. http://support.ndigital.com

Park, S. S., Choi, S. H., Hong, J. A., Hong, Y. H., Jeong, N. G., Lee, S. Y., Sung, M.-W., & Hah, J. H. (2016). Validity and reliability of the Korean version of the Speech Handicap Index in patients with oral cavity cancer. *International Journal of Oral and Maxillofacial Surgery, 45*(4), 433–439. https://doi.org/10.1016/j.ijom.2015.11.006

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *The Journal of the Acoustical Society of America, 93*(5), 2948–2961. https://doi.org/10.1121/1.405814

Perrier, P., & Fuchs, S. (2015). Motor equivalence in speech production. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 225–247). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118584156

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 12*(1), Article 6. https://doi.org/10.5334/labphon.237

Rebernik, T., Jacobi, J., Tiede, M., & Wieling, M. (2021). Accuracy assessment of two electromagnetic articulographs: Northern Digital Inc. WAVE and Northern Digital Inc. VOX. *Journal of Speech, Language, and Hearing Research, 64*(7), 2637–2667. https://doi.org/10.1044/2021_JSLHR-20-00394

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Ringash, J., Bernstein, L. J., Devins, G., Dunphy, C., Giuliani, M., Martino, R., & McEwen, S. (2018). Head and neck cancer survivorship: Learning the needs, meeting the needs. *Seminars in Radiation Oncology, 28*(1), 64–74. https://doi.org/10.1016/j.semradonc.2017.08.008

Rinkel, R. N., Leeuw, I. M. V., van Reij, E. J., Aaronson, N. K., & Leemans, C. R. (2008). Speech Handicap Index in patients with oral and pharyngeal cancer: Better understanding of patients' complaints. *Head & Neck, 30*(7), 868–874. https://doi.org/10.1002/hed.20795

Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America, 134*(5), EL477–EL483. https://doi.org/10.1121/1.4826150

Schow, R. L. (1991). Considerations in selecting and validating an adult/elderly hearing screening protocol. *Ear and Hearing, 12*(5), 337–348. https://doi.org/10.1097/00003446-199110000-00006

Schuster, M., & Stelzle, F. (2012). Outcome measurements after oral cancer treatment: Speech and speech-related aspects—An overview. *Oral and Maxillofacial Surgery, 16*(3), 291–298. https://doi.org/10.1007/s10006-012-0340-y

Speksnijder, C. M., van der Bilt, A., van der Glas, H. W., Koole, R., & Merkx, M. A. W. (2011). Tongue function in patients treated for malignancies in tongue and/or floor of mouth; A one year prospective study. *International Journal of Oral and Maxillofacial Surgery, 40*(12), 1388–1394. https://doi.org/10.1016/j.ijom.2011.09.003

Stipancic, K. L., Golzy, M., Zhao, Y., Pinkerton, L., Rohl, A., & Kuruvilla-Dugdale, M. (2023). Improving perceptual speech ratings: The effects of auditory training on judgments of dysarthric speech. *Journal of Speech, Language, and Hearing Research, 66*(11), 4236–4258. https://doi.org/10.1044/2023_JSLHR-23-00322

Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research, 59*(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271

Stone, M., Langguth, J. M., Woo, J., Chen, H., & Prince, J. L. (2014). Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research, 57*(3), 707–717. https://doi.org/10.1044/1092-4388(2013/13-0085)

Suzuki, N. (1989). Clinical applications of EPG to Japanese cleft palate and glossectomy patients. *Clinical Linguistics and Phonetics, 3*(1), 127–136.

Takatsu, J., Hanai, N., Suzuki, H., Yoshida, M., Tanaka, Y., Tanaka, S., Hasegawa, Y., & Yamamoto, M. (2017). Phonologic and acoustic analysis of speech following glossectomy and the effect of rehabilitation on speech outcomes. *Journal of Oral and Maxillofacial Surgery, 75*(7), 1530–1541. https://doi.org/10.1016/j.joms.2016.12.004

Thies, T., Hermes, A., & Mücke, D. (2022). Compensation in time and space: Prominence marking in aging and disease. *Language, 7*(1), Article 21. https://doi.org/10.3390/languages7010021

Thompson, A., Hirsch, M. E., Lansford, K. L., & Kim, Y. (2023). Vowel acoustics as predictors of speech intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research, 66*(8S), 3100–3114. https://doi.org/10.1044/2022_JSLHR-22-00287

Thompson, A., & Kim, Y. (2019). Relation of second formant trajectories to tongue kinematics. *The Journal of the Acoustical Society of America, 145*(4), EL323–EL328. https://doi.org/10.1121/1.5099163

Thompson, A., & Kim, Y. (2024). acoustic and kinematic predictors of intelligibility and articulatory precision in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 67*(10), 3595–3611. https://doi.org/10.1044/2024_JSLHR-24-00153

Tienkamp, T. B., Rebernik, T., Buurke, R., Polsterer, K., van Son, R. J. J. H., Wieling, M. B., Witjes, M. J. H., de

Visscher, S. A. H. J., & Abur, D. (2024). The effect of speaking style on the articulatory–acoustic vowel space in individuals with tongue cancer before and after surgical treatment. *13th International Seminar on Speech Production,* 95–98. https://doi.org/10.21437/issp.2024-25

Tienkamp, T. B., Rebernik, T., Halpern, B. M., van Son, R. J. J. H., Wieling, M., Witjes, M. J. H., de Visscher, S. A. H. J., & Abur, D. (2024). Quantifying articulatory working space in individuals surgically treated for oral cancer with electromagnetic articulography. *Journal of Speech, Language, and Hearing Research, 67*(2), 384–399. https://doi.org/10.1044/2023_JSLHR-23-00111

Tienkamp, T. B., van Son, R. J. J. H., & Halpern, B. M. (2023). Objective speech outcomes after surgical treatment for oral cancer: An acoustic analysis of a spontaneous speech corpus containing 32.850 tokens. *Journal of Communication Disorders, 101,* Article 106292. https://doi.org/10.1016/j.jcomdis.2022.106292

Tjaden, K., Kain, A., & Lam, J. (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 57*(4), 1191–1205. https://doi.org/10.1044/2014_JSLHR-S-13-0086

Tjaden, K., Lam, J., & Wilding, G. (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research, 56*(5), 1485–1502. https://doi.org/10.1044/1092-4388(2013/12-0259)

Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

Tschiesner, U., Sabariego, C., Linseisen, E., Becker, S., Stier-Jarmer, M., Cieza, A., & Harreus, U. (2013). Priorities of head and neck cancer patients: A patient survey based on the brief ICF core set for HNC. *European Archives of Oto-Rhino-Laryngology, 270*(12), 3133–3142. https://doi.org/10.1007/s00405-013-2446-8

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 38*(5), 1001–1013. https://doi.org/10.1044/jshr.3805.1001

Van den Steen, L., Van Nuffelen, G., Guns, C., De Groote, M., Pinson, L., & De Bodt, M. (2011). De spraak handicap index: Een instrument voor zelfevaluatie bij dysartriepatiënten [The Speech Handicap Index: A tool for self-evaluation for individuals with dysarthria]. *Logopedie, 24,* 26–30.

van Dijk, S., van Alphen, M. J. A., Jacobi, I., Smeele, L. E., van der Heijden, F., & Balm, A. J. M. (2016). A new accurate 3D measurement tool to assess the range of motion of the tongue in oral cancer patients: A standardized model. *Dysphagia, 31*(1), 97–103. https://doi.org/10.1007/s00455-015-9665-7

van Sluis, K. E., Passchier, E., van Son, R. J. J. H., van der Molen, L., Stuiver, M., van den Brekel, M. W. M., Van den Steen, L., Kalf, J. G., & van Nuffelen, G. (2023). Dutch translation and validation of the Communicative Participation Item Bank (CPIB)—Short form. *International Journal of Language & Communication Disorders, 58*(1), 124–137. https://doi.org/10.1111/1460-6984.12775

van Son, R., Middag, C., & Demuynck, K. (2018). Vowel space as a tool to evaluate articulation problems. *Interspeech,* 357–361. https://doi.org/10.21437/Interspeech.2018-68

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*(11), 764–770. https://doi.org/10.1046/j.1365-2842.1996.d01-186.x

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica, 53*(1), 1–18. https://doi.org/10.1159/000052649

Whitehill, T. L., Ciocca, V., Chan, J. C., & Samman, N. (2006). Acoustic analysis of vowels following glossectomy. *Clinical Linguistics & Phonetics, 20*(2–3), 135–140. https://doi.org/10.1080/02699200400026694

Whitehill, T. L., & Wong, C. C.-Y. (2006). Contributing factors to listener effort for dysarthric speech. *Journal of Medical Speech-Language Pathology, 14*(4), 335–342.

Whitfield, J. A., Dromey, C., & Palmer, P. (2018). Examining acoustic and kinematic measures of articulatory working space: Effects of speech intensity. *Journal of Speech, Language, and Hearing Research, 61*(5), 1104–1117. https://doi.org/10.1044/2018_JSLHR-S-17-0388

Whitfield, J. A., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders, 51,* 19–28. https://doi.org/10.1016/j.jcomdis.2014.06.005

Whitfield, J. A., & Goberman, A. M. (2017). Articulatory–acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology, 19*(2), 184–194. https://doi.org/10.1080/17549507.2016.1193897

Whitfield, J. A., & Mehta, D. D. (2019). Examination of clear speech in Parkinson disease using measures of working vowel space. *Journal of Speech, Language, and Hearing Research, 62*(7), 2082–2098. https://doi.org/10.1044/2019_JSLHR-S-MSC18-18-0189

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis.* Springer-Verlag New York. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wieling, M., Rebernik, T., & Jacobi, J. (2023). SPRAAKLAB: A mobile laboratory for collecting speech production data. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023),* 2060–2064.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics, 59,* 122–143. https://doi.org/10.1016/j.wocn.2016.09.004

Zhou, X., Woo, J., Stone, M., & Espy-Wilson, C. (2013). A cine MRI-based study of sibilant fricatives production in post-glossectomy speakers. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing,* 7780–7784. https://doi.org/10.1109/ICASSP.2013.6639178

Appendix A

North Wind and the Sun Passage

Dutch text as included in the present study. Square brackets ([]) denote the eight individual stimuli.

[De noordenwind en de zon waren erover aan het redetwisten wie de sterkste was van hen beiden.] [Juist op dat moment kwam er een reiziger aan, die gehuld was in een warme mantel.] [Ze waren het erover eens dat degene die er als eerste in slaagde de reiziger zijn mantel uit te doen, als sterker moest worden beschouwd dan de ander.] [De noordenwind begon toen uit alle macht te blazen.] [Maar hoe harder hij blies, des te dichter trok de reiziger zijn mantel om zich heen.] [Ten lange leste gaf de noordenwind het op.] [Daarna begon de zon krachtig te stralen, en hierop trok de reiziger onmiddellijk zijn mantel uit.] [De noordenwind moest dus wel bekennen dat de zon van hen beiden de sterkste was.]

Appendix B

Dutch Translations of the Percepts of Intelligibility and Listening Effort

Verstaanbaarheid

De mate in hoeverre de spraak te verstaan is en de boodschap/inhoud te volgen is.
Score 0: Ik heb niks van het fragment verstaan
Score 100: Ik heb alles van het fragment verstaan

Luister moeite
De moeite, concentratie, of energie die het kost om het fragment te verstaan.
Score 0: Het kostte totaal geen moeite
Score 100: Het kostte extreem veel moeite