ISSP 2024 - 13th International Seminar on Speech Production
13-17 May 2024, Autrans, France

The Effect of Speaking Style on the Articulatory-Acoustic Vowel Space in Individuals with Tongue Cancer before and after Surgical Treatment

Thomas B. Tienkamp[1], Teja Rebernik[1], Raoul Buurke[1], Katharina Polsterer[1], Rob J.J.H. van Son[2,3], Martijn Wieling[1], Max J.H. Witjes[1], Sebastiaan A.H.J. de Visscher[1], Defne Abur[1]

[1]*University of Groningen, the Netherlands*
[2]*The Netherlands Cancer Institute, the Netherlands*
[3]*University of Amsterdam, the Netherlands*
t.b.tienkamp, t.rebernik, raoul.buurke, k.m.polsterer, m.b.wieling, d.abur[@rug.nl], r.v.son@nki.nl, m.j.h.witjes, s.a.h.j.de.visscher[@umcg.nl]

Abstract

The impact of surgical treatment for tongue cancer is traditionally assessed with vowel formant metrics from read speech or sustained vowels. However, isolated speech might not fully reflect a speaker's typical speech. Here, we assessed the effect of speaking style (read vs. semi-spontaneous) on vowel acoustics of individuals pre- and post-surgery for tongue cancer. Eight individuals (3 females and 5 males) were recorded pre- and approximately six months post-surgery. We calculated the articulatory-acoustic vowel space (AAVS) during read speech (sentences) and semi-spontaneous speech (picture description). Results showed that the AAVS did not differ significantly pre- and post-surgery. Picture descriptions yielded a significantly smaller AAVS compared to the reading task, which was consistent pre- and post-surgery. Our findings suggest that both read and semi-spontaneous speech styles would be suitable to quantify the impact of surgical intervention for tongue cancer on vowel acoustics.

Keywords: speech production, vowel acoustics, tongue cancer

1. Introduction

Surgical intervention for tongue cancer often reduces tongue mobility (Lazarus et al. 2014; Tienkamp et al. 2024). Reduced tongue mobility may lead to more centralised speech where the acoustic distance between phonemes becomes smaller. Studies that assess the effect of surgery for tongue cancer on speech acoustics often use sustained vowels or isolated words and/or sentences over (semi-)spontaneous speech for their clinical feasibility and increased experimental control (Takatsu et al. 2017; Guo et al. 2023). Studies using isolated utterances have indicated that the vowel space area (VSA) is generally reduced in individuals with tongue cancer following surgical treatment (Balaguer et al. 2020; Takatsu et al. 2017; Guo et al. 2023). However, a recent study that used spontaneous speech did not find significant differences between the vowel formants of individuals treated for tongue cancer and control speakers (Tienkamp, van Son, and Halpern 2023). This raises the question to what extent the conflicting findings for VSA metrics in speakers treated for tongue cancer might result from differences in speaking style.

The choice of speech prompt (vowels/syllables or words) or speaking style (slow, read, or semi-spontaneous) has a considerable effect on the resulting speech output. For choice of speech prompt, individual syllables result in larger vowel spaces compared to words or sentences (van Son, Middag, and Demuynck 2018). For speaking style, larger vowel spaces are found when speakers are asked to read aloud a passage more slowly compared to their habitual speech rate (Turner, Tjaden, and Weismer 1995). In contrast, (semi-)spontaneous speech, which is primarily characterised by a faster speech rate, has resulted in the acoustic reduction of both vowels and consonants compared to read speech (Nakamura, Iwano, and Furui 2008; van Son and Pols 1999). Thus, while sustained vowels or read speech might allow for the recording of best-effort attempts as it elicits larger vowel spaces, more spontaneous speech better reflects daily conversational speech.

At present, no direct comparisons have been made between read and more spontaneous speech in speakers surgically-treated for tongue cancer. Yet, a better understanding of how speaking style affects vowel acoustics before and after surgery for tongue cancer can aid in the development of a standardised speech assessment protocol, which does not exist at present. Specifically, it is not clear which speaking style best captures changes in vowel acoustics following surgical treatment for tongue cancer.

The objective of this study was therefore to assess the effect of speaking style (read vs. semi-spontaneous) on the comprehensive acoustic vowel space in individuals undergoing surgical treatment for tongue cancer. To this end, we measured the articulatory-acoustic vowel space (AAVS) across sentence reading and across more spontaneously elicited speech (i.e., a picture description task) in individuals before and after surgery for tongue cancer. We predicted that the picture description task would yield a smaller AAVS (i.e., more reduced speech) compared to the sentence reading task. Moreover, we predicted an overall reduction of the AAVS following treatment as compared to pre-treatment due to a surgery-induced reduction in tongue mobility. Due to a lack of prior studies on the topic, we did not formulate any specific predictions regarding the interaction between speaking style and treatment.

2. Methods

2.1. Participants

The present study is part of a larger project approved by the institution's Medical Ethics Review Board (NL79242.042.21). All participants provided written informed consent before their participation. Eight native speakers of Dutch (five males, three

 10.21437/issp.2024-25

females) with a mean age of 62.1 years (range: 41-77) completed data collection both pre- and post-surgery and were included in this study. Participants were tested a few days before and approximately six months after surgical treatment. Speakers were treated for T1 (n=5), T2 (n=2) or T3 (n=1) tongue tumours located either on the mid-line of the tongue (S07) or the lateral side of the tongue (all other speakers). T-stages can range from T1 (smallest) to T4 (largest). For six speakers, the tumour was localised on the anterior $2/3$ of the tongue, whereas for two speakers (S02 and S04), the tumour was localised on the posterior $1/3$ of the tongue. The tongue was reconstructed using a radial forearm free flap for two speakers (S01 and S02), whereas the wound was locally closed for other speakers. One speaker received (chemo)radiation post-surgery (S02) and was recorded six months after the last radiation session to ensure a comparable time post-treatment. Table 1 shows the demographic and clinical information of all speakers.

Table 1: *Speaker demographics and clinical information. F = female, M = male, Anterior = anterior $2/3$ of the tongue. Posterior = posterior $1/3$ of the tongue.*

Speaker	Sex	Age	T-stage	Location
S01	F	75	T3	Anterior
S02	M	41	T2	Posterior
S03	M	54	T1	Anterior
S04	F	77	T1	Posterior
S05	M	55	T1	Anterior
S06	M	68	T2	Anterior
S07	F	61	T1	Anterior
S08	M	62	T1	Anterior

2.2. Procedures

All speakers were recorded in the mobile sound booth SPRAAKLAB (Wieling, Rebernik, and Jacobi 2023) and were fitted with an omni-directional microphone (Shure MX-153T) angled 45° from the mouth with a seven centimetre mic-to-mouth distance. Their speech was recorded at a 22,050 Hz sampling rate. To elicit semi-spontaneous speech, participants were asked to describe two pictures in detail using their habitual speaking style: the Cookie Theft picture (Goodglass, Kaplan, and Weintraub 2001) and the Cat Rescue picture (Nicholas and Brookshire 1993). To elicit read speech, participants were asked to read aloud 15 phonemically-balanced sentences with the phonemes of Dutch at the frequency the phonemes typically occur (Luts et al. 2014). In the case of a misreading, speakers were asked to repeat the sentence and only the correctly read instance was used for analysis.

2.3. Acoustic analysis

We used the articulatory-acoustic vowel space (AAVS) in this study (Whitfield and Goberman 2014) to quantify vowel articulation in each speaking style. An advantage of the AAVS over point-based metrics, such as the VSA, is that the AAVS can be computed over full trajectories of running speech (e.g., picture descriptions and full sentences). For this reason, the AAVS takes all vowels into account, thus increasing ecological validity. We calculated the AAVS according to methods established in prior work and developed a custom semi-automatic pipeline (Whitfield and Goberman 2014; Abur, Perkell, and Stepp 2022). First, all instances of 'uh' and 'uhm' were manually removed from the picture description recordings. Next, we removed all voiceless segments using a custom script in Praat (version 6.3.1; Boersma and Weenink 2023). Continuous first and second formant frequency (F_1 and F_2) traces were extracted in Praat from the voiced segments using a script based on Carignan (2022). As formant frequency tracking accuracy is considerably influenced by both speaker and vowel characteristics, the Carignan (2022) script extracts the 'optimal' formant frequency by calculating the F_1 and F_2 using formant ceilings ranging from 3500-6000 Hz with 50 Hz intervals (see e.g., Escudero et al. 2009), time steps of 5 ms, and 25 ms time windows. From these 51 possible formant values (one associated with each ceiling), those two standard deviations away from each mean formant value were removed. From the remaining formant frequencies, the median value was taken as the optimal formant frequency for each given 5 ms time step (representing a single data point).

The resulting formant trajectories were filtered using a median absolute deviation filter, removing data points 2.5 times away from the median absolute deviation of the dataset (5,584 rows, 1.8%).[1] We calculated the AAVS on a mel-scale for each speaker at each assessment point, and for each speaking style (four AAVS values per speaker). The formant trajectories of both picture descriptions were combined to calculate one AAVS value. The AAVS was calculated as the square-root of the product of the squared variance of the formant tracks and the unshared variance between them (see equation (1)). The unshared variance was calculated by subtracting the R^2 of a linear model with F_1 predicting F_2 from 1.

$$\text{AAVS} = \sqrt{(\sigma_{F_1})^2 \times (\sigma_{F_2})^2 \times (1 - \text{R}^2)} \qquad (1)$$

2.4. Statistical analysis

The data were analysed using linear mixed-effects regression in R (version 4.3.2; R Core Team 2023; Bates et al. 2015; Kuznetsova, Brockhoff, and Christensen 2017). Our hypothesis-testing model included the AAVS as a function of surgery (pre vs. post surgery) in interaction with style (read speech vs. semi-spontaneous), together with a by-speaker random intercept. We further assessed the influence of speaker sex and articulation rate (number of syllables / phonation time in seconds) in an exploratory modeling procedure, as these variables can impact vowel acoustics. The articulation rate was calculated using a Praat script by De Jong and Wempe (2009). All numerical variables were centered around the mean and the α-level was set at 0.05. We concluded our analysis by verifying the model's assumptions and employing model criticism (Fox and Weisberg 2019). Data points with an absolute residual exceeding 2.5 standard deviations from their fitted value were removed. We only used this trimmed dataset when outliers drove the absence or presence of statistically significant effects (Baayen 2008).

3. Results

Our results are based on the trimmed dataset that removed one data point from the analysis (3%). An overview of the AAVS values per style and time point is provided in Figure 1-A. The AAVS post-treatment was not significantly smaller compared to

[1] Additional manual filtering only removed an extra 600 rows (0.2%). The correlation between the AAVS with and without manual filtering was $r = .99$ and our subsequent results were nearly identical. We therefore use the AAVS values without manual filtering.

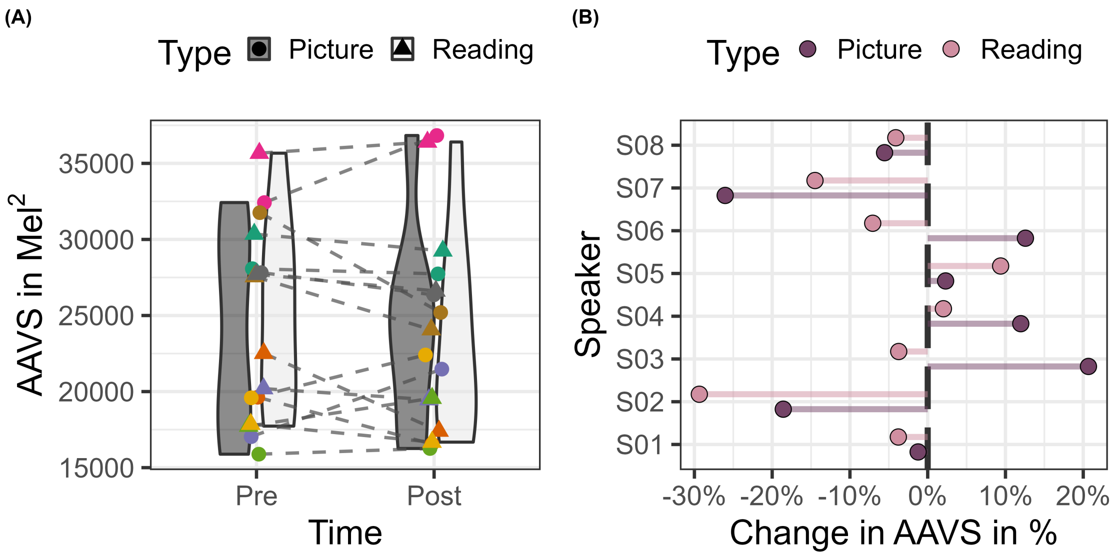

Figure 1: **(A)** Articulatory-acoustic vowel space (AAVS) per time point and speaking style. Different colours represent individual speakers, different shapes represent the speaking styles (circles = picture description, triangles = sentence reading). **(B)** Change in AAVS in percentage compared to pre-treatment between both speaking styles per speaker. A negative value indicates that the AAVS was smaller post-treatment compared to pre-treatment. A positive value indicates an increase in AAVS.

pre-treatment ($p = .66$). On average, semi-spontaneous speech yielded a significantly smaller AAVS compared to read speech (β = -2,621 mel^2, T = -2.7, CI = [-4,471, -529], $p = .02$). There was no significant interaction between time and style ($p = .22$). Figure 1-B shows the change in AAVS in percentage compared to pre-treatment per speaking style and speaker. Decreases in AAVS for both speaking styles were observed for four speakers (S01, S02, S07, S08) post-surgery, with the largest decrease in AAVS for speakers S02 and S07.

Our exploratory analysis revealed a significant effect of sex which indicated that, on average, males had a lower AAVS compared to females (β = -11,026 mel^2, T = -3.9, CI = [-16,793, -5,250], $p < .01$). A significant effect of articulation rate indicated a positive relationship between articulation rate and AAVS (β = 5,184 mel^2, T = 2.3, CI = [870, 9,633], $p < .05$). With the inclusion of the exploratory variables, the fixed effect of speaking style became significant.

4. Discussion and conclusion

We assessed the effect of speaking style on the articulatory-acoustic vowel space (AAVS) of individuals with tongue cancer pre- and post-surgical intervention. The results suggest that the surgical intervention did not impact the overall vowel space for the speakers included in this study. This is not in line with previous work that reported a reduced VSA following treatment compared to pre-treatment for tongue cancer (Guo et al. 2023; Takatsu et al. 2017). One important difference, compared to earlier work, is that the speakers included in our study were mostly treated for smaller tumours (T1) whereas the studies by Guo et al. (2023) and Takatsu et al. (2017) also included individuals with large tumours (T4). To rule out the possibility of pre-treatment speech impairments influencing our findings, we verified that our speakers had typical AAVS values before treatment by comparing them to those of Dutch typical speakers (Hoekzema et al. 2024, current proceedings).

The absence of a reduced AAVS could stem from varying post-treatment changes among speakers, as some had an increase in AAVS following treatment (e.g., S03, S04, S05) whereas others showed a decrease (e.g., S02 and S07). The largest increases post-surgery were seen for the semi-spontaneous speech style in speakers treated for anterior tumours. The surgery may have relieved tumor-related discomfort without significantly affecting articulatory function, potentially resulting in increased range of motion during speech post-treatment. In contrast, the two speakers with the largest decrease in AAVS post-surgery (S02 and S07) were treated for either a posterior tumour or a tumour located on the mid-line of the tongue, which seem to have a more pronounced effect on vowel articulation.

On average, speakers with faster articulation rates had a larger AAVS which might seem contradictory at first, as a faster articulation rate typically results in a smaller VSA (Turner, Tjaden, and Weismer 1995). However, speakers whose speech was more affected by surgery might have slowed their speech rate as a compensatory strategy, whereas speakers whose speech was less affected may have remained at their habitual articulation rate and preserved the size of their acoustic working space.

The results of our study further suggest that the AAVS can capture differences induced by speaking style. Previous work showed that the AAVS yielded larger AAVS values in clear speech compared to typical speech during a reading passage (Whitfield and Goberman 2014). We extend these findings by showing that spontaneously elicited speech from picture descriptions resulted in a smaller AAVS compared to a reading task with individual sentences. While it is still possible that speakers produced ‘clear’ speech during the picture description, ‘clear’ semi-spontaneous speech still elicits smaller formant ranges compared to ‘clear’ read speech (Hazan and Baker 2010). The effect of speaking style on AAVS did not change as a result of surgery for the speakers in our study.

It should be noted that our results are based on a small group-level assessment, which is a limitation of our study. A second limitation is that the phonemic content of both speaking styles was not identical. However, we tried to control for this by including sentences that included a distribution of Dutch phonemes at the frequency the phonemes typically occur.

To conclude, to quantify the effect of surgical treatment for tongue cancer on the acoustic vowel space, our results suggest that both reading and semi-spontaneous speech styles would be suitable prompts to use.

5. Acknowledgements

The authors would like to express their gratitude to all speakers who were willing to participate in this study. This work was supported by a University of Groningen CLCG PhD grant awarded to the first author and by the university's Research School of Behavioral and Cognitive Neurosciences. This work was also 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.

6. References

Abur, Defne, Joseph S. Perkell, and Cara E. Stepp (2022). "Impact of Vocal Effort on Respiratory and Articulatory Kinematics". In: *Journal of Speech, Language, and Hearing Research* 65.1, pp. 5–21.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.

Balaguer, Mathieu, Timothy Pommée, Jérôme Farinas, Julien Pinquier, Virginie Woisard, and Renée Speyer (2020). "Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review". In: *Head & Neck* 42.1, pp. 111–130.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48.

Boersma, Paul and David Weenink (2023). *Praat: doing Phonetics by computer [computer programme]*. Version Number: 6.3.1. URL: `https://praat.org`.

Carignan, Christopher (2022). *Formant Optimization*. URL: `https://github.com/ChristopherCarignan/formant-optimization`.

De Jong, Nivja H. and Ton Wempe (2009). "Praat script to detect syllable nuclei and measure speech rate automatically". In: *Behavior Research Methods* 41.2, pp. 385–390.

Escudero, Paola, Paul Boersma, Andréia Schurt Rauber, and Ricardo A. H. Bion (2009). "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese". In: *The Journal of the Acoustical Society of America* 126.3, pp. 1379–1393.

Fox, John and Sanford Weisberg (2019). *An R Companion to Applied Regression*. Third edition. Thousand Oaks CA: Sage.

Goodglass, Harold, Edith Kaplan, and Sandra Weintraub (2001). *BDAE: The Boston diagnostic aphasia examination*. Lippincott Williams & Wilkins Philadelphia, PA.

Guo, Kaixin, Yudong Xiao, Wei Deng, Guiyi Zhao, Jie Zhang, Yujie Liang, Le Yang, and Guiqing Liao (2023). "Speech disorders in patients with Tongue squamous cell carcinoma: A longitudinal observational study based on a questionnaire and acoustic analysis". en. In: *BMC Oral Health* 23.1, p. 192.

Hazan, Valerie and Rachel Baker (2010). "Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style?" In: *DiSS-LPSS Joint Workshop 2010*, pp. 7–10.

Hoekzema, Nikki, Teja Rebernik, Thomas B. Tienkamp, Sasha Chaboksavar, Valentina Ciot, Annetje Gleichman, Roel Jonkers, Aude Noiray, Martijn B. Wieling, and Defne Abur (2024). "Assessing differences in articulatory-acoustic vowel space in Parkinson's disease by sex and phenotype". In: *Proceedings of the 13th International Seminar on Speech Production*.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13.

Lazarus, C. L., H. Husaini, A. S. Jacobson, J. K. Mojica, D. Buchbinder, D. Okay, and M. L. Urken (2014). "Development of a New Lingual Range-of-Motion Assessment Scale: Normative Data in Surgically Treated Oral Cancer Patients". In: *Dysphagia* 29.4, pp. 489–499.

Luts, Heleen, Sofie Jansen, Wouter Dreschler, and Jan Wouters (2014). "Development and normative data for the Flemish/Dutch Matrix test". In: *Unpublished Article*.

Nakamura, Masanobu, Koji Iwano, and Sadaoki Furui (2008). "Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance". In: *Computer Speech & Language* 22.2, pp. 171–184.

Nicholas, Linda E and Robert H Brookshire (1993). "A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia". In: *Journal of Speech, Language, and Hearing Research* 36.2, pp. 338–350.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: `https://www.R-project.org/`.

Takatsu, Jun, Nobuhiro Hanai, Hidenori Suzuki, Masahiro Yoshida, Yasuhiro Tanaka, Seiya Tanaka, Yasuhisa Hasegawa, and Masahiko Yamamoto (2017). "Phonologic and Acoustic Analysis of Speech Following Glossectomy and the Effect of Rehabilitation on Speech Outcomes". In: *Journal of Oral and Maxillofacial Surgery* 75.7, pp. 1530–1541.

Tienkamp, Thomas, Teja Rebernik, Bence Halpern, Rob van Son, Martijn Wieling, Max Witjes, Sebastiaan de Visscher, and Defne Abur (2024). "Quantifying Changes in Articulatory Working Space in Individuals Surgically Treated for Oral Cancer with Electromagnetic Articulography". In: *Journal of Speech, Language and Hearing Research* 67.2, pp. 384–399.

Tienkamp, Thomas, Rob van Son, and Bence Halpern (2023). "Objective speech outcomes after surgical treatment for oral cancer: An acoustic analysis of a spontaneous speech corpus containing 32.850 tokens". In: *Journal of Communication Disorders* 101, p. 106292.

Turner, Greg S., Kris Tjaden, and Gary Weismer (1995). "The Influence of Speaking Rate on Vowel Space and Speech Intelligibility for Individuals With Amyotrophic Lateral Sclerosis". In: *Journal of Speech, Language, and Hearing Research* 38.5, pp. 1001–1013.

van Son, Rob, Catherine Middag, and Kris Demuynck (2018). "Vowel Space as a Tool to Evaluate Articulation Problems". In: *Proceedings of Interspeech 2018*. ISCA, pp. 357–361.

van Son, Rob and Louis C. W. Pols (1999). "An acoustic description of consonant reduction". In: *Speech Communication* 28.2, pp. 125–140.

Whitfield, Jason A. and Alexander M. Goberman (2014). "Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease". In: *Journal of Communication Disorders* 51, pp. 19–28.

Wieling, Martijn, Teja Rebernik, and Jidde Jacobi (2023). "SPRAAKLAB: a mobile laboratory for collecting speech production data". In: *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, pp. 2060–2064.