Sculpting vowel space, an exploration*

Rob J. J. H. van Son
Netherlands Cancer Institute, ACLC, University of Amsterdam

This study investigates a method for generating synthetic speech that mimics the distorted vowel articulation found in speakers treated for oral cancer. Speech technology often performs poorly for these speakers, in part because little is known about how the articulatory consequences of oral cancer treatment impact speech intelligibility and quality, and because disordered speech data are scarce. To address this gap, we introduce a data augmentation method that directly manipulates vowel formant tracks, specifically F_1 and F_2, to recreate the characteristic flattening and reduction of vowel space associated with restricted tongue mobility after surgery. Using LPC based analysis-resynthesis, F_2 values are capped, F_2 dynamics are limited using low pass filtering, and F_1 is scaled to model compensatory jaw and lip movement. Manipulated vowel-like segments are then spliced into the original recordings while preserving non-vowel material to reduce synthesis artifacts. The resulting speech reproduces key acoustic properties of oral cancer vowel spaces, including severely reduced front-back distinction. Although artifacts remain and audio quality is imperfect, the method shows potential for creating controlled distortions useful for perceptual experiments and for expanding training data in inclusive speech technology development.

1 Introduction

While speech technology is increasingly used in modern life, crucial technologies are often biased against speakers who differ from the norm (Feng et al. 2021, 2024; Scharenborg 2025). One such group are those who, after recovering from oral cancer, suffer from disordered speech (Balaguer et al. 2019a, 2019b, 2020; Tienkamp et al. 2025b). The speech of those treated for oral cancers often is not compatible with speech technology (Feng et al. 2021, 2024; Halpern et al. 2020).

A basic impediment to create better speech technology for these speakers treated for oral cancer, and those suffering from disordered speech in general, is a lack of speech data for system development (Halpern et al. 2022). For this lack

* Microsoft Copilot has been used to improve some formulations in the text. The Netherlands Cancer Institute receives a research grant from ATOS Medical AB Sweden, which contributes to the existing infrastructure for health-related, quality-of-life research in the Department of Head and Neck Oncology and Surgery.

© R.J.J.H. van Son

of disordered speech data, data augmentation techniques have been developed (Geng et al. 2020; Halpern et al. 2022; Jaitly & Hinton 2013; Jin et al. 2021; Prananta et al. 2022). Still, the particulars of post-oral cancer speech are generally different from other disorders (Tienkamp et al. 2024; Tienkamp, et al. 2025a; Tienkamp et al. 2025b) and current data augmentation techniques cannot reproduce crucial aspects of speech following oral cancer treatment.

Another consequence of the scarcity of post-oral cancer speech data are limits to the understanding of the consequences of cancer treatment. Although it is clear that many patients do experience speech impairments (Balaguer et al. 2019a, 2019b) due to changes in their articulation (Tienkamp et al. 2025a), it is still not entirely clear what articulatory changes are behind, e.g., reduced intelligibility (Tienkamp et al. 2025b). This lack of understanding limits the development of evidence-based speech therapies for these patients (Blyth et al. 2024; Tienkamp 2026: 25). The study of the relation between changes in articulation and the perceptual quality of the speech would be helped if the relevant aspects of the speech could be sampled individually, or better, produced by speech synthesis.

Current best practices in speech data augmentation are based on either simple methods, e.g., speaking rate, pitch and vocal tract length perturbation (VTLP), or statistical methods, i.e., deep neural network voice conversion (Geng et al. 2020; Jiao et al. 2018; Jin et al. 2021; Ko et al. 2015; Park et al. 2019; Ragni et al. 2014; Rebai et al. 2017; Zhang et al. 2024). Speaking rate, pitch and VTLP augmentations can be done with simple deterministic methods, e.g., the Praat *Change Gender* command (Boersma et al. 2026), or in a Praat script (Dubagunta et al. 2022; van Son 2019/2025b). However, changes in the articulation as seen in speech after oral cancer require targeted changes in the spectra of the speech, e.g., in the formant tracks of vowels or spectral moments of fricatives. Without training data, such required changes cannot be readily produced by statistical voice conversion techniques.

The current study explores the possibilities of direct changes in vowel formant values to reproduce aspects of vowel acoustic working space as seen in speakers after oral cancer surgery (Tienkamp et al. 2025c). The aim is to make it possible to take a speech recording of a typical speaker and to reproduce the speech with altered vowels realized with features as seen in post-oral cancer speech.

The current work focuses on simulating changes of the oral cancer vowel space. Examples of typical and post-oral cancer vowel spaces are presented in Figure 1. The post-oral cancer speech vowel spaces show the characteristic "flattening" in the F_2 (front-back) dimension due tongue mobility limitations after tumor surgery and reconstruction, and also compensatory jaw and lip movements (Tienkamp 2026: 171).

In this explorative study, the aim is to implement a simple model of limitations in tongue mobility (see Table 1), an upper and lower cap on F_2 and a limit on the speed of change of the F_2 using a low pass filter. Compensation using the F_1 (high-low) dimension, i.e., increased openness, was implemented as an expansion of the F_1 range between /i/-/a/ to a higher /a/. Changes were only applied to vowel-like sounds in the F_1-F_2 range of frequencies (0.2-3kHz).

As the synthetic speech should be useful for speech data augmentation and perceptual experiments, audio quality is an important consideration in the design. Audio artifacts are almost inevitable when re-synthesizing manipulated formant tracks. A number of different formant algorithms and source signals for synthesis are available for selection to minimize artifacts depending on speaker and recording quality.

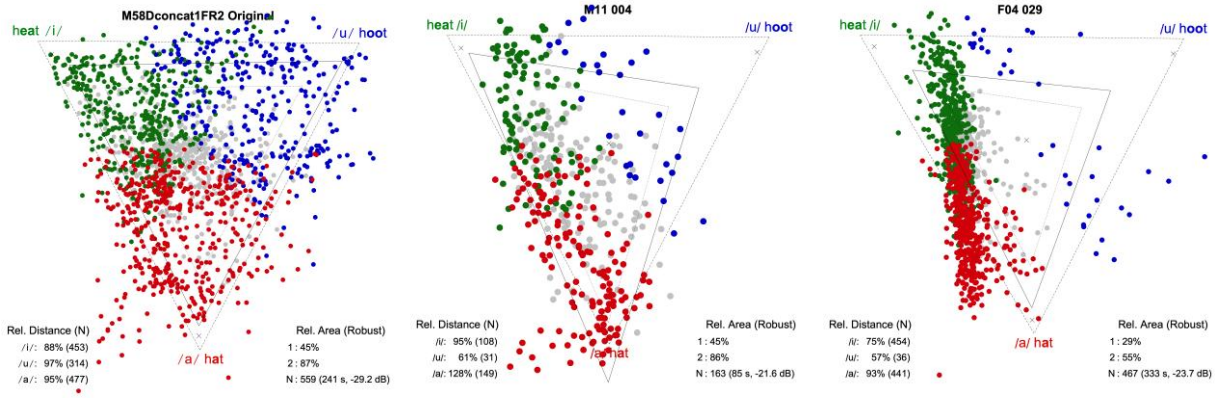


Figure 1: Example vowel spaces (van Son et al. 2018; van Son 2017/2025a). Left, a typical male speaker, M58D, from (van Son 2001); Center, a male speaker after a hemiglossectomy (Robert B. Haase 2017); Right, a female speaker after a total glossectomy (Wyanne Thompson 2017).

2 Methods

Vowel space is remodeled by modifying vowel F_1 and F_2 formant tracks. The operation is performed in two steps.

1. Formant analysis, adaptation, and resynthesis of the whole recording
2. Replacement of the original vowel segments, in the window 0.2-3 kHz, in the recording by newly synthesized manipulated segments

All signal processing, analysis, modifications, and manipulations are performed with Praat (Boersma et al. 2026). The *VowelBender* script[1] used is written in the *Praat* scripting language (van Son 2026/2026). Vowel spaces are calculated and

[1] https://github.com/robvanson/VowelBender/blob/main/VowelBender.praat

plotted using *VowelTriangle*[2] (van Son et al. 2018; van Son 2017/2025a), a *Praat* script that can process speech recordings without manual intervention.

In this paper, manipulations are summarized in (numbered) Praat pseudo-scripts that only present the relevant *Praat* commands in a monospaced (typewriter) font.

2.1 Formant modeling

Formant analysis-synthesis is done with simple 10 pole LPC analysis using 11 kHz or 10 kHz downsampling for, respectively, female and male speakers. Formants are calculated with (1).

(1)
```
To Formant (burg): timeStep, 5, sampleFreq/2, 0.025, preEmphasisFreq
```
In which:
```
timeStep = 0.005
sampleFreq = 11kHz for female, and 10kHz for male speakers
preEmphasisFreq = 80 Hz for female, and 50Hz for male speakers
The Burg algorithm can be replaced by another Praat option
```

Formant values are modified using Formant Grids and applying formulas to cap the F_2 values or expand the F_1 movements. Formant movements result from movements of the articulators changing the lengths of the resonating cavities in the vocal tract. Formant frequencies are inversely related to the length of the resonating sections of the vocal tract (Lammert & Narayanan 2015). Changes in movements of articulators and cavity lengths are, thus, more naturally applied on the inverse of the formant values, $1/F_2$. Therefore, smoothing of the articulation movements was performed by converting the full F_2 tracks in the FormantGrids to $1/F_2$ sounds and applying a filter to the sound (2) and replacing the F_2 tracks with the inverse of the filtered values (1/self).

(2)
```
To Matrix: 2
Formula: "1/self"
To Sound
Filter (pass Hann band): 0, 1/smootheningTime, 1
Down to Matrix
Formula: "1/self"
```
In which: `smootheningTime = Time window`

[2] https://github.com/robvanson/VowelTriangle/blob/master/VowelTriangle.praat

To reduce the effects of sudden discontinuous formant tracks, formant transitions into a vowel segment are smoothed using a cosine interpolation over maximally 50 ms.

2.2 Sound source creation

A PointProcess is created from the original recording (3) for identifying vowel segments and creating the *Pulse Train* and *Phonation*.

(3) `To PointProcess (periodic, cc): 75, 600`

For source signal choices *Pulse Train, De-emphasized Pulse Train,* and *Phonation*, the pointprocess is used for the synthesis (4)

(4) `To Sound (pulse train): origSampleFreq, 1, 0.05, 2000`
`To Sound (phonation): origSampleFreq, 1, 0.05, 0.7, 0.03, 3, 4`
In which: `origSampleFreq` is the sample frequency of the original recording

For the choice *LPCerror*, a "cleaned" LPC error signal is produced (5).

(5)
```
To LPC (autocorrelation): lpcOrder, 0.025, timeStep, preEmphasisFreq
plusObject: resampledOriginal
Filter (inverse)
Resample: origSampleFreq, 50
Filter (pre-emphasis): preEmphasisFreq
Filter (pass Hann band): preEmphasisFreq, 4500, 100
```
In which:
`lpcOrder ≈ 60`
`resampledOriginal` is the original recording resampled to `sampleFreq (1)`
`timeStep, preEmphasisFreq` are from (1)
`origSampleFreq` is from (4)

The extremely high LPC order (≈60) is used to whiten the source signal and attenuate the vowel residues in the source signal. A *Pulse Train* source has a white (flat) spectrum. The *De-emphasized Pulse train*, the *LPC error signal* after pre-emphasis, and the *Phonation* source have a spectrum with a spectral slope of -6dB/octave[3].

[3] See Praat manual page: https://www.fon.hum.uva.nl/praat/manual/Source-filter_synthesis_4__Using_existing_sounds.html

2.3 Synthesis

The remodeled formants are synthesized by filtering the new sound source with the altered formant grid. This sound contains altered vowels and all other parts of the recording.

2.4 Vowel segment replacement

Speech resynthesis from manipulated formant tracks and newly constructed sources generally inserts large audible artifacts into the sound. One cause of artifacts is the mismatch between the LPC model behind the formant analysis and the actual acoustics of the vocal tract for most speech sounds. The formant model works best for vowels and then best for the lower formants. This study only concerns the vowels in speech. Therefore, a procedure has been implemented to replace the lower frequency part of the vowel segments in the original recordings with the corresponding parts of the resynthesized formant tracks.

Vowel segments are identified using the code from (van Son et al. 2018; van Son 2017/2025a). Vowel segments are replaced in four steps:

1. The original recording is split on 3kHz in a low and high frequency part[4] (the region 0–200Hz is kept with the high frequency band)

   ```
   Filter (stop Hann band): 200, 3000, 100
   Filter (pass Hann band): 200, 3000, 100
   originalLowPassInt = Get intensity (dB)
   ```

   Note that the two filter lines are *both* applied on the original recording. The *originalLowPassInt* is calculated on the low pass (0.2–3kHz) filtered signal
2. The synthesized speech is low pass filtered and scaled to the original intensity

   ```
   Filter (pass Hann band): 200, 3000, 100
   Scale intensity: originalLowPassInt
   ```
3. The vowel segments in the original low frequency sound are set to zero. The non-vowel segments in the low frequency synthesized speech are set to zero. Both sounds are combined (summed) to create a complete low frequency recording.
4. The new low frequency sound is combined (summed) with the high frequency part of the original recording.

The procedure is illustrated in Figure 2.

[4] The Hann band filters in Praat are a-causal and do not introduce a phase shift.

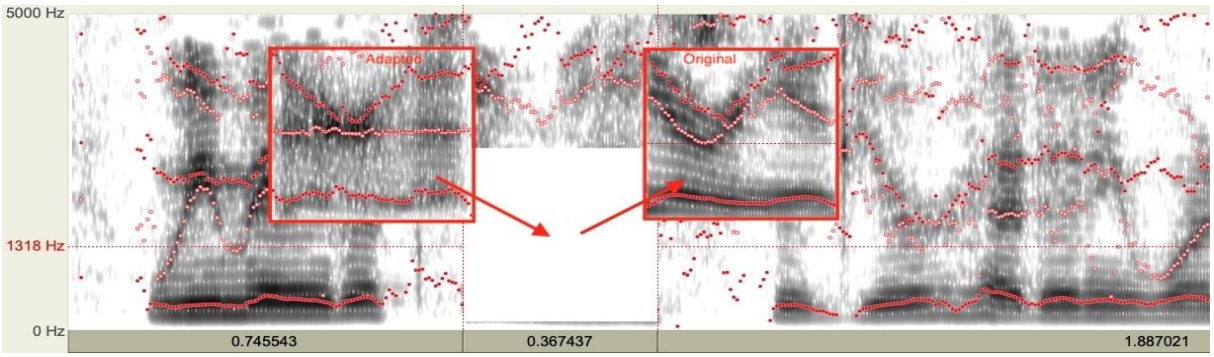


Figure 2: Graphical illustration of the replacement of a vowel-like segment in the original recording with a manipulated synthesized segment. The original spectro-temporal segment is replaced by the corresponding adapted synthetic segment. Only a single replacement is shown for clarity. In this study, all vowel-like segments in a recording are replaced by synthetic segments in this manner.

Table 1: Aspects of vowel formants changed by *VowelBender* (van Son 2026/2026). F_2 tracks are cut-off between a maximum and minimum value, entered as a percentage of the F_2 range between /ə/ and /i/ (front) and /ə/ and /u/ (back) in semitones. To model limitations in speed of tongue movements, F_2 tracks are smoothened with a low-pass filter with a cut-off value of 1/window size in seconds. Smoothing happens on $1/F_2$ frequency (see text). The F_1 /i/-/a/ range is expanded/contracted by a percentage (in semitones) to model articulatory compensation. The example values in the table are based on a male speaker using the Burg algorithm.

Feature	Formant	Change	Value	Example	Limit
Front	F_2	/i/-/ə/	% semitones	75	$F_2 < 1898$Hz
Back	F_2	/ə/-/u/	% semitones	25	$F_2 > 1061$ Hz
Open	F_1	/i/-/a/	% semitones	110	F_1 /a/ → 940
Speed	F_2	window	seconds	0.5	2 Hz low pass filter

3 Results

3.1 Vowel formant space

Figure 1 presents vowel spaces from three speakers, one typical (Dutch) speaker and two (English) speakers who had glossectomies with tongue reconstruction for oral tumors. Figure 3 presents vowel spaces for the same recording of the typical speaker in Figure 1, but now either resynthesized without change (left), and with changes leading to a vowel space comparable to the two speakers with glossectomy (mid and right).

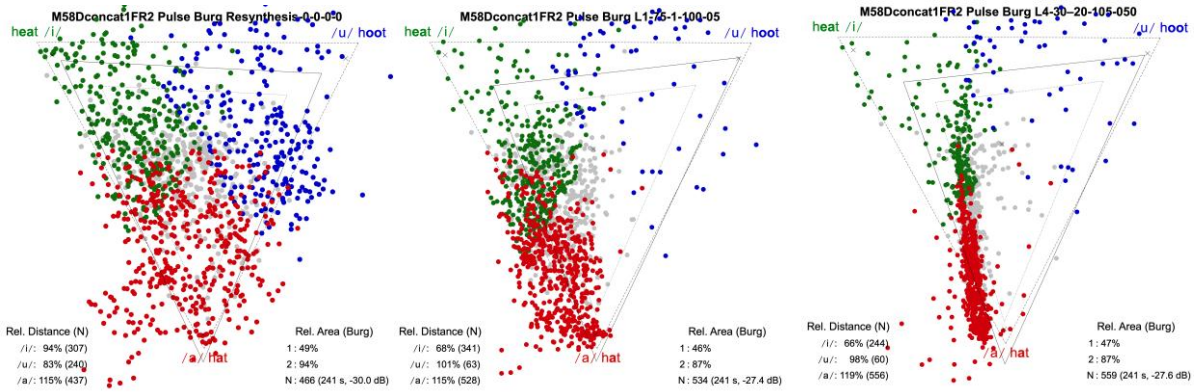


Figure 3: Synthetic vowel spaces based on a recording of a typical male speaker (M58D1FR1, see Fig. 1). Left, plain resynthesis. Center, midway restrictions simulating a hemiglossectomy. Right, severe restrictions simulating a total glossectomy. Conversions use the Burg formant algorithm and a pulse train source.

The vowel spaces in Figure 1 of the two speakers after (hemi-)glossectomy show characteristic flattening in the back-front, /u/-/i/ articulation (F_2 dimension). In the extreme case of the female speaker on the right, all articulatory movement is confined to the jaw (open/close) and lips.

Limiting F_2 values on both sides and reducing the movement speed create vowel spaces (Figure 3) from a typical male speaker with the same characteristics as those in Figure 1. A number of examples based on English and Dutch reading of "The North Wind and the Sun" are available[5] at (van Son 2026/2026).

3.2 Formant track shapes

The effects of the above manipulations on formant track shapes can be seen in Figure 4, where spectrograms and formants of a short sentence fragment clearly show the results of smoothing and limiting the values of F_2. Audio samples for a range of speakers from this corpus (Ribeiro 2018) are available at the *VowelBender* Github Page[6].

[5]https://robvanson.github.io/VowelBender/Examples/North%20Wind%20and%20the%20Sun/
[6]https://robvanson.github.io/VowelBender/Examples/Parallel%20Audiobook%20Corpus%20Treasure/

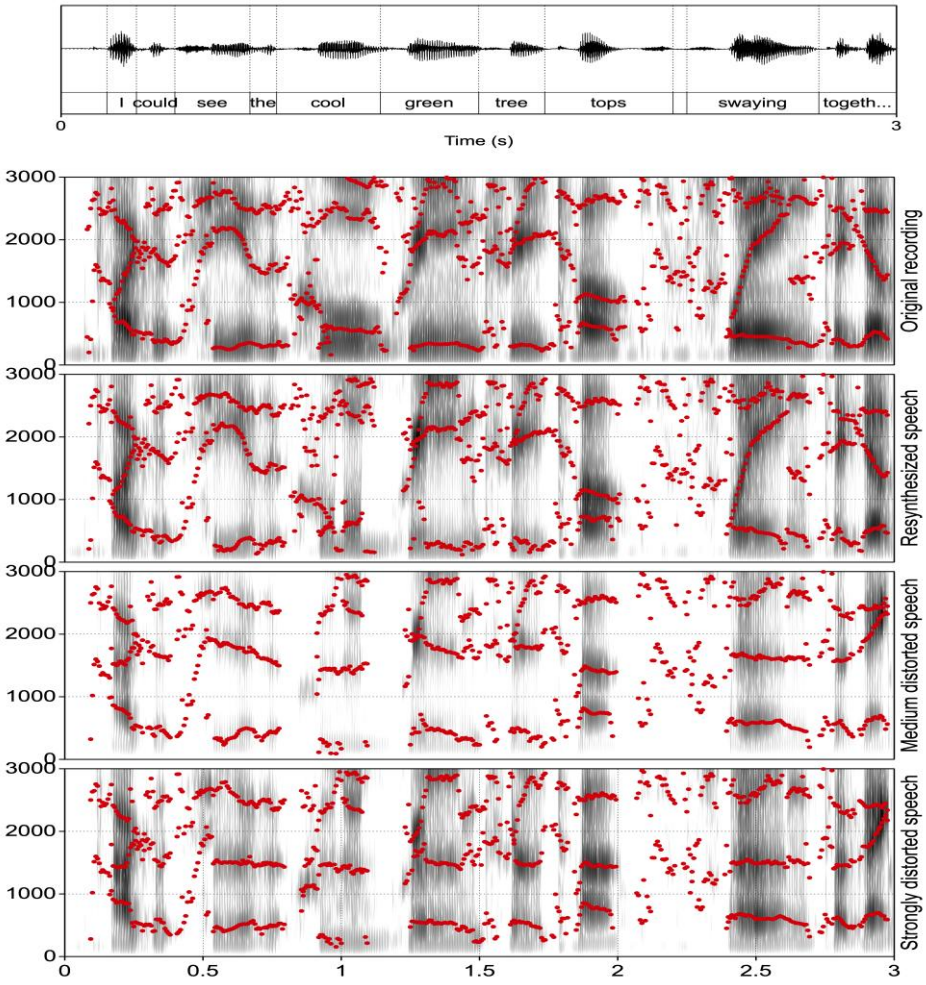


Figure 4: Spectrograms and formants of original, re-synthesized unchanged, and manipulated formants. The sentence fragment reads *"I could see the cool green tree tops swaying togeth[er]"* from a reading of "*Treasure Island*" by Robert Louis Stevenson (Ribeiro 2018; van Son 2026/2026).The parameters for the conversions are the same as in Figure 3, but for this figure, the *LPCerror* signal was used as a source. Note the flattening of the F_2 vowel tracks in the distorted speech.

3.3 Speech quality

A formal evaluation of the quality of the synthesized speech is out of the scope of this paper. An informal listening assessment showed the clear presence of audible artifacts and a reduced speech quality compared to the original recording. Audio examples are available at the *VowelBender* Github Page[7] (van Son 2026/2026).

A first impression suggests that recording quality, e.g., the presence of noise, affects the quality of the estimation of the pitch periods needed for the voice source synthesis. A reduced quality voice source affects all aspects of speech synthesis quality. Re-synthesis imposes an intensity and spectral slope that often

[7] https://robvanson.github.io/VowelBender/#audio-examples

differs from the original. Although the intensity is scaled to approach the original, this is not perfect. At the moment, nothing is done to adapt the spectral slope of the synthesized vowels.

Manipulations of formant frequencies, bandwidths, and voice source varying over time introduce spurious sounds when the LPC model used does not really fit the underlying and intended sounds. Limiting the re-synthesized sounds to specific phoneme segments can reduce these artifacts, but also introduces mismatches between the selected phones and their neighbors[8].

Overall, synthesized speech quality depends on speaker, recording quality, manipulations, and the chosen voice source and formant algorithms. Although the differences do not seem to be large, they are audible[9].

4 Discussion

The aim of this study was to explore ways to generate synthetic examples of speech with distorted, or deviant vowel articulation. The intention is to use this synthetic distorted speech to study the perceptual effects of the distortions and to generate synthetic data for developing speech technology applications useful for post-oral cancer speech. The current study was limited to simulating vowels as the relations between vowel articulation, vowel synthesis, and vowel sounds are relatively well understood.

The simple formant manipulations explored in this study were already able to reproduce the shape of vowel acoustic working space of oral cancer speakers. Formal perceptual experiments and detailed studies of vowel formant track behavior in post-oral cancer speech are needed to refine the simulation of real post-oral cancer speech, especially comparing speech from speakers before and after treatment for oral cancers (e.g., Tienkamp et al. 2025c). Formal evaluation of the synthetic distorted speech is needed to determine how well the intelligibility and perception of the synthetic speech equal that of real post-oral cancer speech.

The proposed method of splicing the relevant spectro-temporal fragments of manipulated vowels into the original recording opens up a targeted approach where the effects of specific manipulations of selected phonemes or phones are produced in otherwise unchanged speech recordings. The current work uses a simple algorithm to automatically segment vowel-like sounds. But other automatically or hand-made segmentations can be supplied. In the current study, formant analysis-synthesis was done with fixed parameters, LPC order (10-pole), and sample frequencies (11 and 10 kHz). This is not optimal (e.g., Escudero et al.

[8] https://robvanson.github.io/VowelBender/Examples/Labeled%20Speech/
[9] https://robvanson.github.io/VowelBender/Examples/Source%20Formant%20examples/

2009), and better analysis parameters can be applied on a per speaker and phone basis.

The speech examples created for this work are certainly not perfect or life-like. However, there are several avenues open for improvements. Most importantly, they show a route to study in more detail the effects of treatment for oral cancer on speech quality and intelligibility. Such studies could lead to even better simulations which then can be used as synthetic data for developing speech technology applications that can help speakers and care givers.

5 Conclusions

Replacing vowels in running speech with synthetic vowels manipulated to generate aspects of the acoustic vowel space of oral cancer speakers is possible using manipulated formant analysis-synthesis procedures. The quality of manipulated speech resulting from the first explorations is not perfect, but shows promise of future improvements.

6 References

Balaguer, M., Boisguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019a). Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *136*(5): 355–359. https://doi.org/10.1016/j.anorl.2019.05.012

Balaguer, M., Boisguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019b). Factors influencing intelligibility and severity of chronic speech disorders of patients treated for oral or oropharyngeal cancer. *European Archives of Oto-Rhino-Laryngology*, *276*(6): 1767–1774. https://doi.org/10.1007/s00405-019-05397-6

Balaguer, M., Pommée, T., Farinas, J., Pinquier, J., Woisard, V., & Speyer, R. (2020). Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review. *Head & Neck*, *42*(1): 111–130. https://doi.org/10.1002/hed.25949

Blyth, K., Stainlay, H., & McCabe, P. (2024). Speech and swallowing intervention following oral cancer treatment: A survey of speech-language pathologists in Australia and New Zealand. *International Journal of Speech-Language Pathology*, *26*(5): 652–662. https://doi.org/10.1080/17549507.2023.2240043

Boersma, P., Weenink, D., & Shimorina, A. (2026). *Praat: Doing phonetics by computer (v6.4.61)* [Computer Program]. https://praat.org/

Dubagunta, S. P., van Son, R. J. J. H., & Magimai.-Doss, M. (2022). Adjustable deterministic pseudonymization of speech. *Computer Speech & Language*, *72*: 101284. https://doi.org/10.1016/j.csl.2021.101284

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

Feng, S., Halpern, B. M., Kudina, O., & Scharenborg, O. (2024). Towards inclusive automatic speech recognition. *Computer Speech & Language*, *84*: 101567. https://doi.org/10.1016/j.csl.2023.101567

Feng, S., Kudina, O., Halpern, B. M., & Scharenborg, O. (2021). Quantifying bias in automatic speech recognition. *arXiv:2103.15122 [Cs, Eess]*. http://arxiv.org/abs/2103.15122

Geng, M., Xie, X., Liu, S., Yu, J., Hu, S., Liu, X., & Meng, H. (2020). Investigation of data augmentation techniques for disordered speech recognition. *Proc. Interspeech 2020*: 696–700. https://doi.org/10.21437/Interspeech.2020-1161

Haase, R. B. (2017, December 22). *16-Weeks post-hemiglossectomy update—Healing from tongue cancer surgery* [Video recording]. https://www.youtube.com/watch?v=moZluyaE4S8

Halpern, B. M., Feng, S., van Son, R., van den Brekel, M., & Scharenborg, O. (2022). Low-resource automatic speech recognition and error analyses of oral cancer speech. *Speech Communication*, *141*: 14–27. https://doi.org/10.1016/j.specom.2022.04.006

Halpern, B. M., Son, R. van, Brekel, M. van den, & Scharenborg, O. (2020). Detecting and analysing spontaneous oral cancer speech in the wild. *Proc. Interspeech 2020*: 4826–4830. https://doi.org/10.21437/Interspeech.2020-1598

Jaitly, N., & Hinton, G. E. (2013). Vocal tract length perturbation (VTLP) improves speech recognition. *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*: *117*, 21.

Jiao, Y., Tu, M., Berisha, V., & Liss, J. (2018). Simulating dysarthric speech for training data augmentation in clinical speech applications. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*: 6009–6013. https://doi.org/10.1109/ICASSP.2018.8462290

Jin, Z., Geng, M., Xie, X., Yu, J., Liu, S., Liu, X., & Meng, H. (2021). Adversarial data augmentation for disordered speech recognition. *Proc. Interspeech 2021*: 4803–4807. https://doi.org/10.21437/Interspeech.2021-168

Ko, T., Peddinti, V., Povey, D., & Khudanpur, S. (2015). Audio augmentation for speech recognition. *Proc. Interspeech 2015*: 3586–3589. https://doi.org/10.21437/Interspeech.2015-711

Lammert, A. C., & Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLOS ONE*, *10*(7): e0132193. https://doi.org/10.1371/journal.pone.0132193

Park, D. S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. *Proc. Interspeech 2019*: 2613–2617. https://doi.org/10.21437/Interspeech.2019-2680

Prananta, L., Halpern, B., Feng, S., & Scharenborg, O. (2022). The effectiveness of time stretching for enhancing dysarthric speech for improved dysarthric speech recognition. *Proc. Interspeech 2022*: 36–40. https://doi.org/10.21437/Interspeech.2022-190

Ragni, A., Knill, K. M., Rath, S. P., & Gales, M. J. (2014). Data augmentation for low resource languages. *Proc. Interspeech 2014*: 810–814. https://doi.org/10.21437/Interspeech.2014-207

Rebai, I., BenAyed, Y., Mahdi, W., & Lorré, J.-P. (2017). Improving speech recognition using data augmentation and acoustic model fusion. *Procedia Computer Science*, *112*: 316–322. https://doi.org/10.1016/j.procs.2017.08.003

Ribeiro, M. S. (2018). *Parallel audiobook corpus*. University of Edinburgh. https://doi.org/10.7488/ds/2468

Scharenborg, O. (2025). Inclusive speech technology: Developing automatic speech recognition for everyone. *The Journal of the Acoustical Society of America*, *157*(4_Supplement), A267–A267. https://doi.org/10.1121/10.0038021

Thompson, W. (2017, May 4). *You&Me Making It Bigger 2017 Workshop* [Video recording]. https://www.youtube.com/watch?v=Mrb5M20bSXA

Tienkamp, T. (2026). *Speech motor control following surgical treatment for oral cancer: Acoustic and kinematic insights* [Doctoral Dissertation, University of Groningen]. https://doi.org/10.33612/diss.1466708158

Tienkamp, T. B., Rebernik, T., Buurke, R., Polsterer, K. M., Van Son, R., Wieling, M., Witjes, M. J. H., De Visscher, S., & Abur, D. (2024). The effect of speaking style on the articulatory-acoustic vowel space in individuals with tongue cancer before and after surgical treatment. *ISSP 2024 - 13th International Seminar on Speech Production*: 95–98. https://doi.org/10.21437/issp.2024-25

Tienkamp, T. B., Rebernik, T., D'Cruz, R. A., Son, R. J. J. H. van, Wieling, M., Witjes, M. J. H., Visscher, S. A. H. J. de, & Abur, D. (2025a). Articulatory–kinematic changes in speech following surgical treatment for oral or oropharyngeal cancer: A systematic review. *International Journal of Language & Communication Disorders*, *60*(1): e13148. https://doi.org/10.1111/1460-6984.13148

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. (2025b). Associations between acoustic, kinematic, self-reported, and perceptual measures of speech in individuals surgically treated for oral cancer. *Journal of Speech, Language, and Hearing Research*, *68*(7): 3069–3089. https://doi.org/10.1044/2025_JSLHR-24-00464

Tienkamp, T., Van Ast, F., Van Der Veen, R., Rebernik, T., Buurke, R., Hoekzema, N., Polsterer, K., Sekeres, H., Van Son, R., Wieling, M., Witjes, M., De Visscher, S., & Abur, D. (2025c). Articulatory clarity and variability before and after surgery for tongue cancer. *Proc. Interspeech 2025*: 3558–3562. https://doi.org/10.21437/Interspeech.2025-708

van Son, R. J. J. H. (2001). *The IFA spoken language corpus* (Version 1.0) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.14904089

van Son, R. J. J. H. (2025a). *VowelTriangle* [Praat]. https://github.com/robvanson/VowelTriangle (Original software published 2017)

van Son, R. J. J. H. (2025b). *PseudonymizeSpeech* [Praat]. https://github.com/robvanson/PseudonymizeSpeech (Original software published 2019)

van Son, R. J. J. H. (2026). *VowelBender* [Praat]. https://github.com/robvanson/VowelBender (Original work published 2026)

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

Zhang, Y., Yue, Z., Patel, T., & Scharenborg, O. (2024). Improving child speech recognition with augmented child-like speech. *Proc. Interspeech 2024*: 5183–5187. https://doi.org/10.21437/Interspeech.2024-485

Rob J.J.H. van Son, r.v.son@nki.nl

Department of Head and Neck Oncology and Surgery

Netherlands Cancer Institute

Amsterdam, the Netherlands