The Impact of Electromagnetic Articulography Sensors on the Articulatory-Acoustic Vowel Space in Speakers with and without Parkinson's Disease

Thomas B. Tienkamp[1], Teja Rebernik[1], Jidde Jacobi[1], Martijn Wieling[1], Defne Abur[1]

[1] *University of Groningen*
`t.b.tienkamp, t.rebernik, j.jacobi, m.b.wieling, d.abur[@rug.nl]`

Abstract

The somatosensory effect of electromagnetic articulography (EMA) sensors on speech remains relatively unexplored. Moreover, EMA sensors may be more disruptive to speech in individuals with somatosensory deficits (e.g., persons with Parkinson's Disease; PwPD). Thus, we investigated the effect of EMA sensors on the articulatory-acoustic vowel space (AAVS) in both typical speakers (n=23) and PwPD (n=23). The AAVS was calculated before EMA sensor placement, directly after, and after approximately one hour to assess habituation. The AAVS significantly decreased following sensor placement and did not change with habituation, regardless of speaker group. PwPD had a smaller AAVS compared to typical speakers, but were not differentially impacted by EMA sensors. EMA sensor placement led to average reductions of the AAVS of 13.5% for PwPD and 14.2% for typical speakers, which suggests that articulatory-acoustics from studies with and without the use of EMA sensors may not be fully comparable.

Keywords: Electromagnetic articulography, speech acoustics, Parkinson's Disease

1. Introduction

Electromagnetic articulography (EMA) provides fine-grained spatial and temporal information on articulatory movements during speech. While the primary outcome measures of speech studies using EMA are kinematic trajectories, it is not uncommon to also collect parallel acoustic data (Mefferd and Green 2010; Lee, Littlejohn, and Simmons 2017; Thompson and Kim 2019). However, when using EMA, the sensors that are attached to the tongue, jaw and lips may alter the speaker's articulatory-acoustic output as they might interfere with one's articulation. This raises the question to what extent the articulatory-acoustic output with EMA sensors on represents the typical output of a speaker. Given that the sensor coils also change the somatosensory feedback a speaker receives, it further raises the question as to whether the presence of EMA sensors impacts the articulatory-acoustic output of those with sensory deficits, such as persons with Parkinson's Disease (PwPD), to a greater extent than typical speakers (Conte et al. 2013). Parkinson's disease (PD) is a progressive neurodegenerative disease that affects various aspects of motor and sensory functioning, including the speech subsystems (Opara et al. 2017; Broadfoot et al. 2019; Chen and Watson 2017).

Previous studies assessing the impact of EMA sensors on articulatory-acoustics yielded mixed findings across various speaker populations, including typical speakers (Dromey, Hunter, and Nissen 2018; Bartholomew 2020), individuals with apraxia of speech (AOS; Katz, Bharadwaj, and Stettler 2006), and PwPD (Hirsch, Thompson, and Kim 2024). Katz, Bharadwaj, and Stettler (2006) showed that EMA sensors did not cause consistent group-level articulatory-acoustic effects on the production of vowels and fricatives in target words produced by individuals with and without AOS. In contrast, Dromey, Hunter, and Nissen (2018) showed that following sensor placement, the centre of gravity of sibilants embedded in target words was significantly reduced and did not increase over the course of habituation (20 minutes) in typical speakers. Moreover, Bartholomew (2020) observed a decrease in the first formant frequency (F_1) in target words four minutes after sensor placement compared to directly after EMA sensor placement for typical speakers, but comparisons to a pre-placement baseline were not conducted. Lastly, Hirsch, Thompson, and Kim (2024) reported a lower centre of gravity in sibilants directly after EMA sensor placement in speakers with and without PD compared to before sensor placement using a reading passage. In the same study and passage, the authors reported no significant differences in the quadrilateral vowel space area (q-VSA) for speakers with and without PD. However, Hirsch, Thompson, and Kim (2024) did not assess habituation to the sensors over a longer period of time between individuals with and without PD. Thus, the question remains as to what extent the presence of sensor coils across a longer time period may differentially affect PwPD, also in terms of habituation to the sensors themselves.

Therefore, the objective of this study was to determine the effect of EMA sensors on a sentence-level articulatory-acoustic measure of speech for both typical speakers and PwPD. We also assessed whether speakers habituated over time (approximately 60 minutes) and whether habituation varied by speaker group. If speakers adapt to the somatosensory changes introduced by the sensor coils, we would expect the AAVS after a long period of habituation to be significantly larger than the AAVS directly after sensor placement and comparable to the AAVS prior to the sensor placement. If speakers do not adapt to the EMA sensors, we would expect no significant differences in AAVS as a function of time since sensor placement.

2. Methods

2.1. Participants

This study used data from a previous study that received ethical clearance from the institutional Medical Ethics Review Board (NL66063.042.18; Jacobi 2022). We used the data from 46 individuals who gave written permission for their data to be used for follow-up studies. This included 23 typical speakers (18 male, 5 female; mean age = 68.4 years, standard deviation (SD) = 6.2) and 23 PwPD (18 male, 5 female; mean age = 69.1 years, SD = 7.0). Four other speakers participated, but were excluded as they either did not have recordings before sensor placement (n=3) or were not diagnosed with idiopathic PD (n=1). Speakers did not report any hearing, speech, or neurological problems (other than PD) through self-report. All participants were native speakers of Dutch. PwPD participated while ON levodopa and had been diagnosed with idiopathic Parkinson's disease by a neurologist one to 19 years prior to their participation in the study.

2.2. Procedures

All speakers read the Dutch version of the North Wind and the Sun passage before and after EMA sensors (Northern Digital Inc. Wave system) were attached to the tongue, jaw, and lips (Jacobi 2022). Two sensors were placed on the tongue: one approximately one cm from the anatomical tongue tip, and one five mm anterior of the participant's /k/ constriction. Sensors were also placed on the jaw, and the vermilion border of the upper and lower lips. Acoustic data were assessed at three time points: time point 0 (T0), prior to sensor placement; time point 1 (T1), directly after sensor placement; and time point 2 (T2), at the end of the experiment, which lasted approximately one hour and consisted of multiple speaking tasks. The T2 recording was only made for 32 speakers, including 14 PwPD (10 male, 4 female), and 18 typical speakers (14 male, 4 female). Speakers were recorded in a quiet room of their own home with a microphone (Audio Technica AT875R) at a 22,050 Hz sample rate with a mouth-to-mic distance of approximately 20 cm.

2.3. Acoustic analysis

Any speech segments from the researcher giving instructions or any loud background noise (e.g., a clock) were removed from the speech recordings. All voiceless segments were subsequently removed from the speech recordings using a custom script in Praat 6.3.1 (Boersma and Weenink 2023). From these voiced segments, continuous first and second formant (F_1 and F_2) traces were extracted in Praat using a script based on Carignan (2022). As Escudero et al. (2009) showed, formant tracking accuracy is heavily influenced by both speaker and vowel characteristics. The Carignan (2022) script therefore aims to calculate the 'optimal' formant value by extracting the F_1 and F_2 with formant ceilings ranging from 3,500-6,000 Hz with 50 Hz intervals, resulting in 51 measurements (one for each ceiling) per analysis frame. The script uses the Burg algorithm, time steps of 5 ms, and a 25 ms time window. From these 51 possible formant values, those two standard deviations away from the mean formant value were removed. From the remaining formant values, the median value was taken as the optimal formant frequency of a particular 5 ms time step.

The resulting formant traces were filtered using a median absolute deviation filter which removed data points 2.5 times away from the median absolute deviation of the dataset. This removed 16,626 rows (4.1%), where every row corresponds to a 5 ms time step.[1] The AAVS was calculated on a mel-scale based on these filtered trajectories per speaker and time point, resulting in two or three AAVS values per speaker depending on whether the T2 recording was made. To calculate the AAVS, we used the methods established in earlier work (Whitfield and Goberman 2014; Abur, Perkell, and Stepp 2022). First, we computed the squared variance of both the F_1 and F_2 tracks. Next, we calculated the unshared variance by subtracting the R^2 of a linear model with F_1 predicting F_2 from 1. Finally, we take the square-root of the product of the squared variance and unshared variance (see Formula 1).

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

2.4. Statistical analysis

Linear mixed-effects models were used to analyse the data in R 4.3.2 (R Core Team 2023; Bates et al. 2015; Kuznetsova, Brockhoff, and Christensen 2017). Our hypothesis model included the effect of group (PwPD vs. Typical) and time (T0, T1, T2) on the AAVS, and a by-speaker random intercept. All numerical variables were centered around the mean. We assessed whether adding an interaction between group and time improved the fit of the model by using the *anova()* function. A p-value below .05 would indicate that the interaction significantly improves the model.

Following our hypothesis test, we assessed the effects of speaker sex and age in an exploratory manner using model selection procedures, as these variables may impact vowel formants. We compared models using the *anova()* function and kept the more complex level if it significantly improved the fit of the model (i.e., $p < .05$).

To conclude our analysis, we employed model criticism by refitting our model on a trimmed dataset in which we removed data points whose residuals were at least two SDs away from their fitted value (Baayen 2008, Chapter 6). We used this trimmed data set if, and only if, outliers drove the presence or absence of any significant effects. Finally, we verified that the model met the assumptions of normality, homoscedasticity, multicollinearity and autocorrelation (Fox and Weisberg 2019).

3. Results

Our results are based on the dataset with trimmed residuals in which 5 data points (4.03%) were removed. Descriptive results per sex, group, and time are provided in Figure 1A. The AAVS at T0 was significantly larger compared to T1 (β = 3,325 mel^2, T = 8.0, CI = [2,433, 4,096], $p < .001$). On average, the AAVS was 13.5% smaller for PwPD at T1 compared to T0, and 14.2% smaller for typical speakers. There was no significant difference between the AAVS at T2 and T1 ($p = .30$). On average, the AAVS was 0.7% larger at T2 compared to T1 for PwPD, and 5.6% larger for typical speakers. A main effect of group indicated that PwPD had a significantly smaller AAVS compared to typical speakers overall (β = -5,850 mel^2, T = -4.9, CI = [-8,035, -3,577], $p < .001$). The addition of an interaction between time and group did not improve the fit of the model ($\chi^2(2) = 1.14$, $p = .57$, see Figure 1B).

Our subsequent exploratory analysis revealed a significant effect of sex which indicated that males had a lower AAVS compared to females (β = -10,101 mel^2, T = -7.0, CI = [-13,055, -7296], $p < .001$). Secondly, a significant effect of age indicated that AAVS decreased with speaker age (β = -191 mel^2, T = -2.1, CI = [-375, -22], $p = .04$). The inclusion of the exploratory variables did not alter the significance levels of the terms included in our hypothesis model.

4. Discussion and conclusion

The purpose of this study was to investigate the effect of electromagnetic articulography (EMA) sensors on the articulatory-acoustic vowel space (AAVS) in both typical speakers and per-

---

[1] Additional manual filtering removed an extra 605 rows (0.2%). The results with and without manual filtering were nearly identical and we therefore use the AAVS with the median absolute deviation filter only.

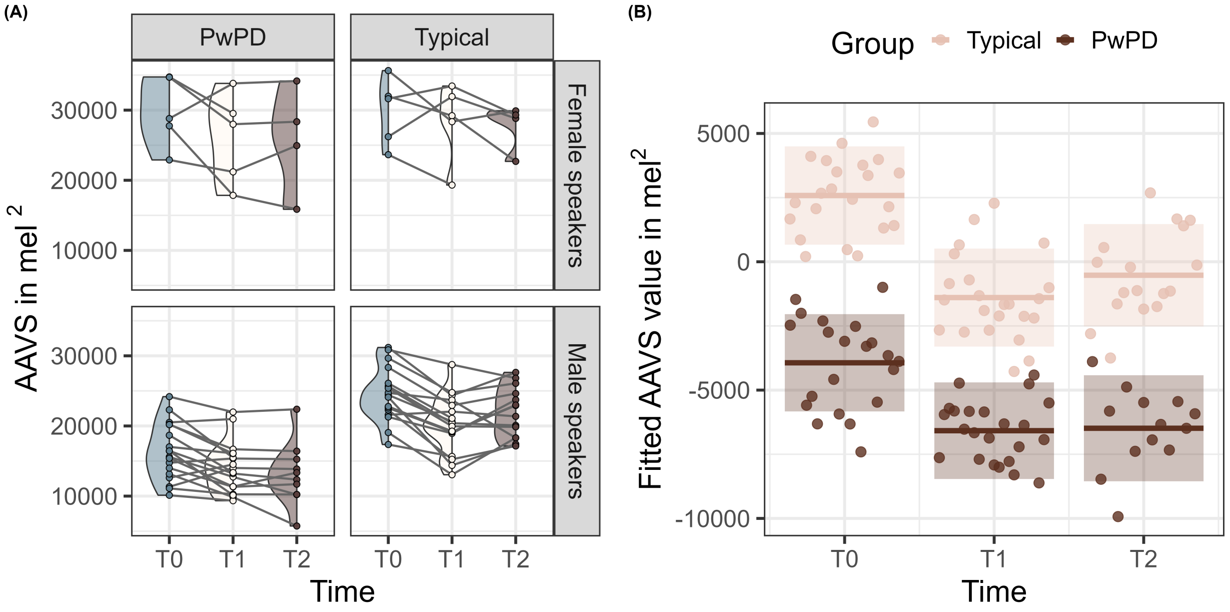

Figure 1: **(A)** Articulatory-acoustic vowel space (AAVS) per time point (T0, T1, T2) and group (Typical, PwPD), by sex (Male, Female). Different colours represent different time points (T0: blue, T1: white, T3: brown). Each point represents an individual speaker. **(B)** Model output showing the fitted mean-centered AAVS values of each group (Typical speakers; cream, Persons with Parkinson's disease: brown) per time point.

sons with Parkinson's Disease (PwPD). The results suggest that the AAVS is reduced after EMA sensor placement and does not significantly increase with habituation regardless of speaker group. This is in line with the results reported by Dromey, Hunter, and Nissen (2018), who previously reported no significant acoustic adaptation for /s/ and /ʃ/.

We did not find evidence that PwPD are affected by the EMA sensors to a different extent than typical speakers, which suggests that group differences in AAVS were not impacted due to the placement of EMA sensors. Our results are consistent with prior work that showed comparable EMA sensor effects on sibilants between speakers with and without dysarthria, and extend the findings from sibilants and individual vowel formant metrics to sentence-level vowel metrics computed over running speech (Katz, Bharadwaj, and Stettler 2006; Hirsch, Thompson, and Kim 2024). Our results underscore the reliability of using EMA in assessing speech motor functions in PwPD despite possible sensory integration changes that arise as a consequence of PD (Conte et al. 2013). PwPD did have an overall lower AAVS than typical speakers when accounting for sex and age differences, which is in line with previous work (Skodda, Visser, and Schlegel 2011; Whitfield and Goberman 2014; Tjaden, Lam, and Wilding 2013).

The results further imply that sentence-level vowel metrics obtained from studies using both acoustic and kinematic methods might not be fully comparable to those obtained from purely acoustic designs. While Dromey, Hunter, and Nissen (2018) reported similar results for sibilants, a sound class that is actively hindered by the presence of sensors coils (i.e., through (near) sensor-palatal contact), we extend this finding by showing that EMA sensors also interfere with the vowel space as measured by the sentence-level AAVS, with average reductions of 13.5% for PwPD and 14.2% for typical speakers. This contrasts with Katz, Bharadwaj, and Stettler (2006), who reported no significant change in F_1 and F_2 measured with and without EMA sensors. However, the task also differed: we employed a reading passage whereas Katz, Bharadwaj, and Stettler (2006) used target words embedded in a carrier phrase, which might have elicited more clear speech. Our results further contrasts with those reported by Hirsch, Thompson, and Kim (2024) as they also did not report statistically significant reductions of the q-VSA following EMA sensor placement compared to pre sensor placement. One possible explanation for the difference is that the AAVS takes all vowels into account and provides an indication of general working space (i.e., the size of the space speakers tend to use the most), whereas previous studies looked at individual vowel formants or vowel formant metrics that provide more absolute indications of the vowel space (i.e., the maximum size of the vowel space).

Lastly, our results showed an effect of age such that the AAVS decreased with speaker age in our sample (age range: 52-81 years), regardless of speaker group. This finding might be explained by age-related atrophy of the orofacial and tongue musculature, which might result in smaller articulatory movements (Neel and Palmer 2012). However, it is important to note that the effects of aging on the size of the vowel space have been inconsistent, and that we did not test any young or middle aged adults (see e.g., Kent and Vorperian 2018; Hermes, Audibert, and Bourbon 2023).

A limitation of our study was that we could only assess habituation at the end of the experiment for a subset of participants (32/46 speakers). Considering that speakers were tested at home, the different locations may have resulted in different levels of background noise. To account for this, we checked the acoustic recordings and ensured an appropriate signal to noise ratio was present for all recordings (> 30 dB; Deliyski, Shaw, and Evans 2005), and levels ranged from 33.6-59.4 dB (mean: 44.5 dB).

In conclusion, we show that passage-level vowel formant metrics are reduced as a result of EMA sensor placement, with an average reduction of 13.5% for PwPD and 14.2% for typical speakers. The AAVS did not increase after a long period of habituation regardless of speaker group. As a result, articulatory-acoustic vowel metrics from studies with and without parallel EMA data acquisition might not be comparable. Moreover, our results show that individuals with and without PD are impacted by the presence of EMA sensors in a similar manner, underscoring its reliability in assessing the speech motor functions in PwPD.

5. Acknowledgements

We are most grateful to all speakers included in the original study who kindly agreed to have their data be used for scientific purposes. This work was supported by a University of Groningen Center for Language and Cognition PhD grant awarded to the first author and by the Research School of Behavioral and Cognitive Neurosciences of the University of Groningen. This work was further supported by the International Macquarie University Research Excellence Scholarship (iMQRES) grant awarded to the third author.

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.

Bartholomew, Emily Adelaide (2020). *Kinematic and Acoustic Adaptation in Response to Electromagnetic Articulography Sensor Perturbation*. MA Thesis. Brigham Young University.

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`.

Broadfoot, C. K., D. Abur, J. D. Hoffmeister, C. E. Stepp, and M. R. Ciucci (2019). “Research-Based Updates in Swallowing and Communication Dysfunction in Parkinson Disease: Implications for Evaluation and Management”. In: *Perspectives of the ASHA Special Interest Groups* 4.5, pp. 825–841.

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

Chen, Yu-Wen and Peter J. Watson (2017). “Speech production and sensory impairment in mild Parkinson’s disease”. In: *The Journal of the Acoustical Society of America* 141.5, pp. 3030–3041.

Conte, Antonella, Nashaba Khan, Giovanni Defazio, John C. Rothwell, and Alfredo Berardelli (2013). “Pathophysiology of somatosensory abnormalities in Parkinson disease”. In: *Nature Reviews Neurology* 9.12, pp. 687–697.

Deliyski, Dimitar D., Heather S. Shaw, and Maegan K. Evans (2005). “Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements”. In: *Journal of Voice* 19.1, pp. 15–28.

Dromey, Christopher, Elise Hunter, and Shawn L. Nissen (2018). “Speech Adaptation to Kinematic Recording Sensors: Perceptual and Acoustic Findings”. In: *Journal of Speech, Language, and Hearing Research* 61.3, pp. 593–603.

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.

Hermes, Anne, Nicolas Audibert, and Angélina Bourbon (2023). “Age-related vowel variation in French”. In: *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague, Czech Republic, pp. 2045–2049.

Hirsch, Micah E., Austin Thompson, and Yunjung Kim (2024). “The Effects of EMA Sensors on Speech in Individuals with and without Dysarthria”. In: Poster presented at the 22nd biennial Motor Speech Conference, San Diego, CA, United States of America.

Jacobi, Jidde (2022). “Coordination and timing of speech gestures in Parkinson’s disease”. PhD thesis. University of Groningen.

Katz, William F., Sneha V. Bharadwaj, and Monica P. Stettler (2006). “Influences of Electromagnetic Articulography Sensors on Speech Produced by Healthy Adults and Individuals With Aphasia and Apraxia”. In: *Journal of Speech, Language, and Hearing Research* 49.3, pp. 645–659.

Kent, Raymond D. and Houri K. Vorperian (2018). “Static measurements of vowel formant frequencies and bandwidths: A review”. In: *Journal of Communication Disorders* 74, pp. 74–97.

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.

Lee, Jimin, Meghan Anne Littlejohn, and Zachary Simmons (2017). “Acoustic and tongue kinematic vowel space in speakers with and without dysarthria”. In: *International Journal of Speech-Language Pathology* 19.2, pp. 195–204.

Mefferd, Antje S. and Jordan R. Green (2010). “Articulatory-to-Acoustic Relations in Response to Speaking Rate and Loudness Manipulations”. In: *Journal of Speech, Language, and Hearing Research* 53.5, pp. 1206–1219.

Neel, Amy T. and Phyllis M. Palmer (2012). “Is Tongue Strength an Important Influence on Rate of Articulation in Diadochokinetic and Reading Tasks?” In: *Journal of Speech, Language, and Hearing Research* 55.1, pp. 235–246.

Opara, Józef, Andrzej Małecki, Elżbieta Małecka, and Teresa Socha (2017). “Motor assessment in Parkinson‘s disease”. In: *Annals of Agricultural and Environmental Medicine* 24.3, pp. 411–415.

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/`.

Skodda, Sabine, Wenke Visser, and Uwe Schlegel (2011). “Vowel Articulation in Parkinson’s Disease”. In: *Journal of Voice* 25.4, pp. 467–472.

Thompson, Austin and Yunjung Kim (2019). “Relation of second formant trajectories to tongue kinematics”. In: *The Journal of the Acoustical Society of America* 145.4, EL323–EL328.

Tjaden, Kris, Jennifer Lam, and Greg Wilding (2013). “Vowel Acoustics in Parkinson’s Disease and Multiple Sclerosis: Comparison of Clear, Loud, and Slow Speaking Conditions”. In: *Journal of Speech, Language, and Hearing Research* 56.5, pp. 1485–1502.

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.