INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A New Vowel Normalization for Sociophonetics

Wilbert Heeringa[1], *Hans Van de Velde*[12]

[1]Fryske Akademy, The Netherlands
[2]Utrecht University, The Netherlands

WHeeringa@fryske-akademy.nl, HVandeVelde@fryske-akademy.nl

Abstract

Several studies have shown that in sociophonetic research Lobanov's speaker normalization method outperforms other methods for normalizing vowel formants of speakers. An advantage of Lobanov's method compared to the method that was introduced by Watt & Fabricius in 2002 is that it is independent of the shape of the vowel space area, and also normalizes to the dispersion of the vowels. However, it does depend on the distribution of the vowels within the vowel space. When using Lobanov normalization the formant values are converted to z-scores. We present a method where the μ in the z-score formula is replaced by the center of the convex hull that encloses the vowels, and the σ is obtained on the basis of the points that constitute the convex hull. When normalizing measurements of two real data sets, and of a series of randomly generated data sets, we found that our method improved in matching vowel spaces in size and overlap.

Index Terms: vowel formants, vowel space, normalization, sociophonetics

1. Introduction

Sociophonetics aims to identify and explain language variation and change by using modern phonetic methods [1] [2] [3] [4]. A classical methodological issue in sociophonetics is the normalization of vowel formants [5] [6] [7] [8]. Vowel formant normalization methods are used in order to eliminate acoustic differences resulting from anatomical differences between speakers, but have to preserve socio-geographic and cross-variety distinctions, as well as phonological distinctions between vowels. From a theoretical point of view, it is argued that a good normalization procedure should also model the normalization processes used by human listeners. In this paper, we propose a new normalization method.

When the speaker's complete vowel system has been measured, vowel-extrinsic, formant-intrinsic and speaker-intrinsic normalization methods are preferably used [5] [9] [10] [11] [12]. Within this groups of methods Adank et al. [5] found that "Lobanov is the best procedure for language variation research.". Fabricius et al. [10] found that "Lobanov is the most successful technique with regard to improving overlap and optimizing area ratios between pairs of speakers." Kohn & Farrington [12] wrote that Lobanov was the most effective technique at eliminating variation attributable to age. Volín [13] found also that "Lobanov showed the best performance", having the highest success rates in discrimination analyses of 3000 vowels. Van der Harst [7] concludes that "Lobanov stands out, mainly because it is best in dealing with the normalization of diphthongs" and "Lobanov is the best method to normalize formant values." (p. 122). However, according to [14] Lobanov may overnormalize formant measures and remove sociolinguistically relevant information.

The normalization methods of Watt & Fabricius [8], Fabricius et al. [10] and Bigham [15] depend on the size of the vowel space. The formant values are expressed as values relative to the centroid of a speaker's vowel space. The centroid is obtained on the basis of the corner points of the vowel space. The methods of [8] and [10] assume the vowel space to be triangular, and the method of Bigham [15] assumes the vowel space to be a trapezium. It is difficult to decide what method to use when the shape of a vowel space is something in between a triangle or trapezium, or if the shape is something else.

The advantage of Lobanov's [16] method is that a particular shape of the vowel space is not assumed, and that formants are also normalized to the dispersion of the vowels in the vowel space. A speaker's mean formant frequency is subtracted from a formant value and then divided by the standard deviation for that formant. The formant values are thus converted in z-scores which have a mean of 0 and a standard deviation of 1. The formula for normalizing values of formant i is:

$$F_i^{Lobanov} = \frac{F_i - \mu_i}{\sigma_i} \tag{1}$$

A weakness of Lobanov's method is that μ depends on the distribution of the vowels within the vowel space. If vowel spaces have (about) the same shape and the same set of different vowels, but have differently distributed vowels, their μ's are different.

In this paper we introduce a new normalization method that tries to handle the disadvantages of both normalization techniques: It does not depend on the shape of the vowel space neither on the distribution of the vowels within the vowel space. We introduce the method in Section 2. In Section 3 we describe the procedures and data sets that we use for evaluating the new normalization method. In 4 the results are presented. In 5 we mention our free web program Visible Vowels in which the new normalization method is available. Conclusions and discussion are found in 6.

2. Description of the method

In order to develop a normalization method that works for vowel spaces of any shape, we initially developed a generalization of the methods of Watt & Fabricius [8], Fabricius et al. [10] and Bigham [15]. We obtained the centroid on the basis of the convex hull that encloses the vowels in the vowel space. A convex hull is the smallest possible hull that encloses all points in a two-dimensional space. Assume we represent vowels as nails that are hammered in a wooden surface correctly representing their acoustic relationships. Then if we stretch a rubber band around the nails, this forms the edge of the convex hull [17]. Thus the convex hull metric tends to maximize the shape of the vowel space and is a more complete assessment of the vowel space area, since it allows for arbitrary shapes instead of only triangle or quadrilateral shaped areas [18].

http://dx.doi.org/10.21437/Interspeech.2021-1846

For finding the vowels that constitute the convex hull, we used the R function `chull`. For finding the centroid of the hull – actually the center of mass – we used the R function `poly_center` from the `pracma` package [19]. The result is shown in Figure 1 where 15 Dutch vowel are enclosed by a convex hull. The large '+' in the center marks the location of the centroid.

When formant values are measured at multiple times in the vowel interval, the convex hull is obtained on the basis of formant values that are averaged across the time points per vowel.

Just as Watt & Fabricius [8] did, we normalized F1 and F2 by dividing them by their respective centroid coordinates. We evaluated the method by using the evaluation methods that are described in Section 3. Compared to the method of Watt & Fabricius [8] and its derivations we found that our method improved in matching vowel spaces in overlap, but not in size. It performed worse than Lobanov's method. Characteristic for Lobanov's normalization is that it does not only center the vowels around (f1=0, f2=0), but also scales them by dividing them by the standard deviation of the vowel formants, which makes the sizes of the vowel spaces of the speakers more comparable. Therefore, we decided to develop a variant of Lobanov's method that does not depend on the distribution of the vowels in the vowel spaces of the speakers.

As a first step, we simply replaced the μ in Lobanov's z-score formula by the centroid coordinates. As as second step, we wanted to calculate the standard deviation on the basis of the formant values of the vowels that constitute the convex hull. However, as can be seen in Figure 1, those vowels are irregular distributed. The vowels [u], [o], [ɔ], [ɑu] and [ɑ] are found close to each other. But the vowels [ɛi] and [a] are distant to each other, not having any other vowels in between.

In order to solve this, we interpolated the number of points on the convex hull up to 1000 points. Next we classified the points in ten classes of equal width, both on the basis of F1 and F2. We found that by using ten classes there is an equilibrium between the even distribution of the points on the convex hull and providing sufficient detail. In our example the width of F1 classes was 54 Hz and of F2 classes 150 Hz. The ten F1 classes do not exactly correspond with the F2 classes since F1 and F2 differences of the pairs of two successive points do not exactly correlate. Therefore, points may have the same F1 class and different F2 classes, or the other way around. In our example we found 36 different F1 class/F2 class combinations. For each of the 36 combinations we averaged the points that had an F1 within the F1 class and an F2 within the F2 class. The result is shown in Figure 2. The 36 points on the convex hull are distributed a lot more evenly than the original ten points that constituted the convex hull (see Figure 1. On the basis of these 36 points the standard deviation – both for F1 and F2 – was calculated.

Our method requires that F1/F2 measurements are available for all vowels that constitute a speaker's vowel system. In our implementation of Lobanov normalization and our new method the centroid and the standard deviation (Lobanov) or convex hull (our own method) are calculated on the basis of averaged formant values, i.e. for each of the variables F1, F2 and F3 (Lobanov only) the values are averaged per combination of speaker, vowel and time point (25%, 50%, etc.) and subsequently averaged across time points.

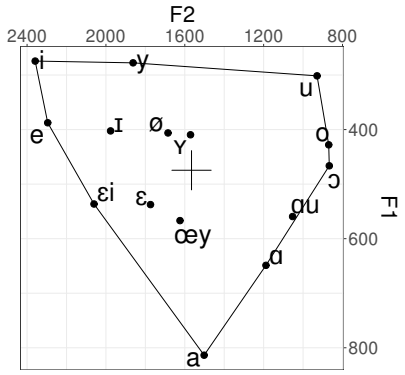

Figure 1: *15 Dutch vowels enclosed by a convex hull.*

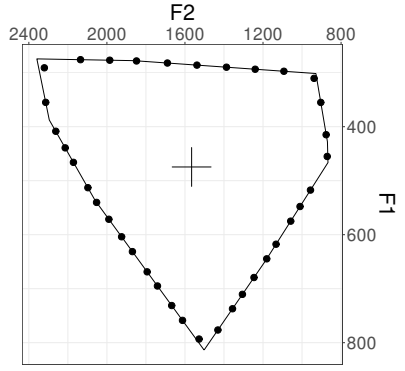

Figure 2: *A more regular distribution of points on the convex hull.*

3. Evaluation

3.1. Evaluation methods

We used two evaluation methods that were developed by Fabricius et al. [10]. Their methods try to derive normalized plots of vowel spaces for different speakers that were "as well matched in size and overlap as possible." (p. 414). They wrote:

> Rather than sorting between different possible sources of variability, and seeking to eliminate some and retain others, we simply seek to optimize the process of visual comparison between vowel plots from any two individuals, regardless of which sociolinguistically relevant factor lies behind the variability. (p. 417/418)

Factors like 'age' and 'gender' may be both anatomical and sociolinguistic factors. By making vowel spaces of different speakers maximally matching in size and overlap, we avoid the need to decide whether the variability of a factor should be minimized or maximized by a normalization procedure, or to what extent its variability should be minimized or maximized.

The two methods that we adopted from Fabricius et al. [10] were also used by Flynn [20] and Flynn & Foulkes [11]. The first method assesses the ability to equalize vowel spaces and the second to align vowel spaces.

The idea behind the first method is to quantify the equalization of the areas of the vowel spaces by examining the reduction of variance in the speakers' vowel spaces. In order to calculate the area of a vowel space, [11] assumed the vowel space to have the shape of a trapezium. Fabricius et al. [10] calculated the

area of a vowel space on the basis of its convex hull, which makes the procedure independent of any shape of the vowel space. In order to find the convex hull we used the R function `chull`. The area that is enclosed by the convex hull was calculated by the R function `polyarea` from the `pracma` package. Then the squared coefficient of variance (SCV) was calculated as:

$$SCV = \left(\frac{\sigma}{\mu}\right)^2 \quad (2)$$

Dividing σ by μ makes the SCV scale-invariant. Next, Fabricius et al. [10] divided each method's SCV by the Hertz SCV, which gave the proportion of variance that remained after normalization. This proportion was subtracted from 1, resulting in the proportional reduction in variance.

The second method proposed by Fabricius et al. [10] was also used by Flynn [20], Flynn & Foulkes [11] and Esfandiaria & Alinezhad [21]. When using this method the area of the intersection of the vowel spaces of the speakers is calculated and divided by the area of the union of the speaker's vowel spaces. This results in the proportion of area that overlaps. A higher proportion shows a better alignment. Again, the areas are found on the basis of their convex hulls, not assuming any particular shape a priori. However, different from what Fabricius et al. [10] proposed, Flynn & Foulkes [11] assumed a quadrilateral and Esfandiaria & Alinezhad [21] assumed a triangle.

Fabricius et al. [10] calculated overlap for each pair of speakers. Following Flynn & Foulkes [11] we divided the area of the intersection of the vowel spaces of all speakers by the area of the union of the vowel spaces of all speakers, thus obtaining one score for the complete set of speakers.

3.2. Data sets

We evaluated the new method on the basis of two data sets. The first is the classical data set of Peterson & Barney [22] and the second is a data set of Van der Harst [7].

Peterson & Barney [22] analyzed the sounds of General American English at Bell Telephone Laboratories. They measured the frequency and amplitude of F1, F2 and F3 for 10 vowels and 76 speakers. The speakers pronounced twelve different monosyllabic words, each beginning with [h] and ending with [d] and differing only in the vowel. The words were pronounced by male and female speakers and by children. We copied this data set from the computer program PRAAT [23] where it is freely available.

Van der Harst [7] measured f0, F1, F2, F3 and vowel duration of the 15 full vowels of Dutch on the basis of word list data, i.e. monosyllabic words, before coda [s] and [t]. The f0 and the formants were measured at the 13%, 25%, 38%, 50%, 62%, 75% and 88% time point in vowel intervals. The speakers are 160 teachers of Dutch at high schools and were recorded in 1999/2000. They were selected according to dialectological and socio-geographic criteria via schools in medium-sized cities. The data set is stratified by community (The Netherlands and Flanders), region (four regions in each community), gender (male, female) and age (old, young). The speakers in the youngest group were between 22 and 40 years old at the time of the interview and speakers in the oldest group were between 45 and 60 years old. As for the factor gender, the biological sex distinction was used. Each of the eight regions was represented by 20 speakers: five young male speakers, five older male speakers, five young female speakers, and five older female speakers. The data set of [7] is available at `https://fryske-akademy.nl/fa-apps/tutorial/#dataset`.

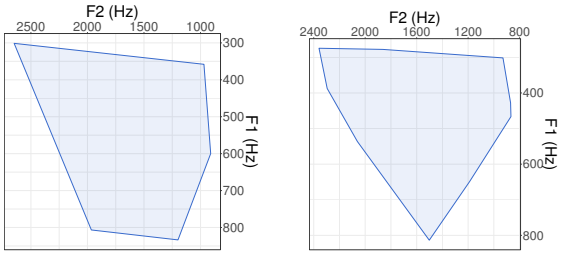

Figure 3: *Left: trapezium shaped convex hull of 10 vowels averaged over 76 speakers in the Peterson & Barney 1952 data set. Right: triangle shaped convex hull of 15 vowels averaged over 160 speakers in the Van der Harst 2011 data set.*

As can be seen in Figure 3 the shape of the averaged vowel spaces of the speakers in the Peterson & Barney [22] looks more like a trapezium and the shape of the averaged vowel spaces of the speakers in the Van der Harst [7] data looks more like a triangle.

Using the data set of Van der Harst [7] with 160 speakers, 15 different vowels per speaker, and two pronunciations per speaker/vowel as a template, we generated 20 data sets in which formant values were randomly generated. Per formant we measured the minimum and maximum value across the whole data set, and then we assigned a random value between those two extremes to each vowel pronunciation. In this way we obtained data sets in which the sizes and shapes of the (artificial) vowel shapes maximally vary. They can have any shape.

4. Results

In Figure 4 the convex hulls of the speakers' vowel spaces are superimposed per data set. The pictures suggest a strong effect of the Lobanov normalization compared to using raw Hertz values. The pictures obtained on the basis of our new method, referred to as 'Heeringa & Van de Velde', suggest a slightly tighter matching of the convex hulls.

Evaluation results are shown in Table 1 for the two data sets. When evaluating on the basis of the Van der Harst data set, we experimented with different combinations of percentage time points. We also left out the corner vowels [i], [u] and [a] one by one in order to get different shapes. In most cases we obtained higher percentages of variance reduction and overlap for our new method.

Additionally, we applied both Lobanov normalization and our new normalization method to the 20 data sets in which formant values were randomly generated (see 3.2). In Figure 5 the median and the spread of the percentages of variance reduction and overlap are shown per normalization method. Using a Welch two sample t-test the percentages of variance reduction were compared between the two normalization methods. Significantly higher percentages were found for our new normalization method compared to Lobanov normalization (t=-33.78, df=20.568, $p < 0.001$). Likewise we compared the percentages of overlap between the two normalization methods and found again significantly higher percentages for our new method (t=-16.701, df=32.223, $p < 0.001$).

Table 1: *Percentages of variance reduction of vowel space areas, and percentages of overlap of vowel space areas for the Lobanov and Heeringa & Van de Velde normalization procedures.*

	time points							% variance reduction		% overlap	
source	13%	25%	38%	50%	62%	75%	88%	Lobanov	H. & V.d.V	Lobanov	H. & V.d.V.
Peterson & Barney				x				97.4	99.51	31.7	45.8
V.d. Harst				x				95.2	99.4	29.8	50.8
V.d. Harst			x	x	x			94.1	97.5	29.6	44.6
V.d. Harst		x						93.9	99.6	31.1	51.1
V.d. Harst						x		94.6	99.4	30.6	48.6
V.d. Harst		x				x		90.0	93.4	27.6	38.2
V.d. Harst		x		x		x		91.3	94.4	28.1	39.9
V.d. Harst		x	x	x	x	x		92.3	95.3	28.5	40.9
V.d. Harst	x	x	x	x	x	x	x	89.0	91.6	26.9	36.1
V.d. Harst w.h. [i]				x				94.6	99.5	27.9	48.7
V.d. Harst w.h. [u]				x				94.5	99.0	27.7	43.8
V.d. Harst w.h. [a]				x				96.1	99.5	28.6	45.5

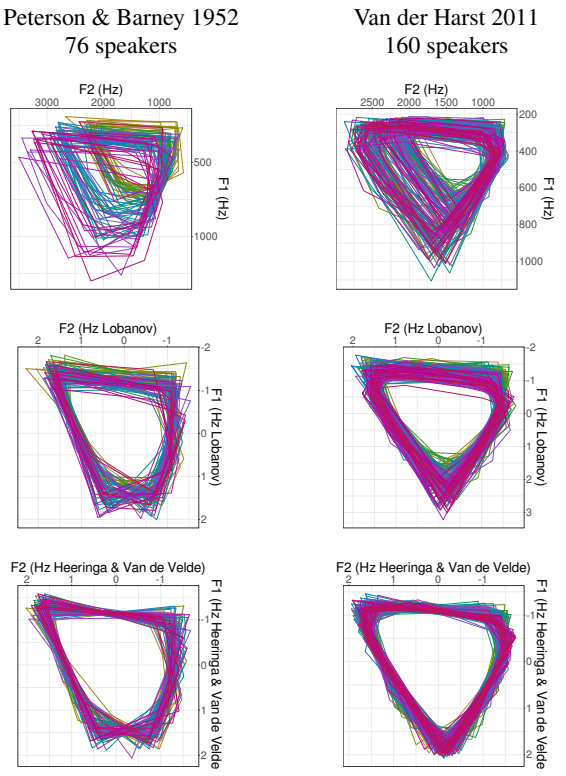

Figure 4: *Overlaying vowel space hulls of speakers of two data sets using raw Hertz data and two normalization methods.*

5. Web program

The new normalization method is available in the web program Visible Vowels [24] as 'Heeringa & Van de Velde II'. Its predecessor is included as 'Heeringa & Van de Velde I'. Visible Vowels is freely available at `visiblevowels.org`. Additionally, 14 other normalization methods including Lobanov's method, and the two evaluation methods of Fabricius et al. [10] that were used in this paper are available in this program. The evaluation methods can be used to evaluate any of the 16 normalization methods for a given data set. The normalization methods can be combined with different scales (Hz, Bark, ERB, logarithmic, mel). A standalone version of the web program can be used by installing the package `visvow` in R.

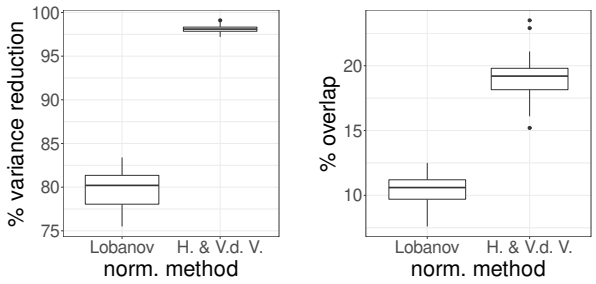

Figure 5: *Median percentages of variance reduction and overlap of vowel space areas per normalization method.*

6. Conclusions and discussion

We developed a method that works for vowel spaces of any shape, and which is independent of the distribution of the vowels in the vowel space. Based on 20 randomly generated data sets, we found that our methods make the vowel space of speakers more comparable, both in size and in overlap.

Instead of determining vowel space areas on the basis of convex hulls, we also considered using concave hulls. Unlike in a convex polygon, the interior angles in a concave hull may be greater than 180°. Therefore, the concave hull tends to eliminate unused regions at the periphery of the vowel space. [25]. Since this would undermine our goal of developing a vowel distribution independent method, we did not investigate the use of concave hulls any further for this paper, but will do so in future work.

The normalization method presented in this paper only works for normalizing F1 and F2 values. Therefore, we consider developing a 3D version that uses a 3D convex hull that is obtained on the basis of F1, F2 and F3. A 3D convex hull can be computed with the Quickhull [26]. This algorithm is implemented as the function `convhulln` in the R `geometry` package. It is less evident how to calculate the area of an arbitrary 3D polygon. It it would be interesting to see how F1 and F2 normalized by the 2D method compare to F1 and F2 normalized by a 3D method.

7. References

[1] M. Baranowski, "On the role of social factors in the loss of phonemic distinctions," *English Language and Linguistics*, vol. 17, no. 2, pp. 271–295, 2013.

[2] C. Celata and S. Calamai, *Introduction: Sociophonetic perspectives on language variation*. John Benjamins, 2014.

[3] M. Di Paolo and M. Yaeger-Dror, *Sociophonetics: A Student's Guide*. Routledge, 2011.

[4] P. Foulkes and G. Docherty, "The social life of phonetics and phonology," *Journal of Phonetics*, vol. 34, no. 4, pp. 409–438, 2006.

[5] P. Adank, R. Smits, and R. van Hout, "A comparison of vowel normalization procedures for language variation research," *Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3099–3107, 2004.

[6] D. Watt, A. Fabricius, and T. Kendall, "More on vowels: Plotting and normalization," in *Sociophonetics: A student's guide*. Routledge, 2010, pp. 107–118.

[7] S. Van der Harst, "The vowel space paradox: A sociophonetic study on Dutch," Ph.D. dissertation, Radboud University, Nijmegen, 2011.

[8] D. Watt and A. Fabricius, "Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1-F2 plane," *Leeds Working Papers in Linguistics and Phonetics*, vol. 9, no. 9, pp. 159–173, 2002.

[9] C. G. Clopper, "Computational methods for normalizing acoustic vowel data for talker differences," *Language and Linguistics Compass*, vol. 3, no. 6, pp. 1430–1442, 2009.

[10] A. Fabricius, D. Watt, and D. E. Johnson, "A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics," *Language Variation and Change*, vol. 21, no. 3, pp. 413–435, 2009.

[11] N. Flynn and P. Foulkes, "Comparing vowel formant normalization methods," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII): August 17-21, 2011*, W. Lee and E. Zee, Eds. City University of Hong Kong, 2011, pp. 683–686.

[12] M. E. Kohn and C. Farrington, "Evaluating acoustic speaker normalization algorithms: Evidence from longitudinal child data," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2237–2248, 2012.

[13] J. Volín, "Normalization of the vocalic space," in *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Springer, 2009, pp. 190–200.

[14] W. Rankinen and K. de Jong, "The entanglement of dialectal variation and speaker normalization," *Language and Speech*, vol. 64, no. 1, pp. 181–202, 2021.

[15] D. S. Bigham, "Dialect contact and accommodation among emerging adults in a university setting," Ph.D. dissertation, University of Texas at Austin, Austin, 2008.

[16] B. M. Lobanov, "Classification of russian vowels spoken by different speakers," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 606–608, 1971.

[17] Wikipedia contributors, "Convex omhulsel — Wikipedia, the free encyclopedia," 2021, accessed 22 March 2021. [Online]. Available: https://nl.wikipedia.org/wiki/Convex_omhulsel.

[18] S. Sandoval, V. Berisha, R. L. Utianski, J. M. Liss, and A. Spanias, "Automatic assessment of vowel space area," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. EL477–EL483, 2013.

[19] H. W. Borchers, *pracma: Practical Numerical Math Functions*, 2019, R package version 2.2.9. [Online]. Available: https://CRAN.R-project.org/package=pracma.

[20] N. Flynn, "Comparing vowel formant normalisation procedures," *York Papers in Linguistics*, vol. 2, no. 11, pp. 1–28, 2011.

[21] N. Esfandiaria and B. Alinezhadb, "Evaluating normalization procedures on reducing the effect of gender in Persian vowel space," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, vol. 13, no. 2, pp. 303–316, 2014.

[22] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America*, vol. 24, no. 2, pp. 175–184, 1952.

[23] P. Boersma and P. Weenink, "Praat: doing phonetics by computer [computer program]," 2021, version 6.1.40, retrieved 27 February 2021. [Online]. Available: http://www.praat.org/.

[24] W. Heeringa and H. Van de Velde, "Visible vowels: a tool for the visualization of vowel variation." in *Proceedings CLARIN Annual Conference 2018, 8 - 10 October, Pisa, Italy*, I. Skadin and M. Eskevich, Eds. CLARIN ERIC, 2018, pp. 124–127. [Online]. Available: https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf.

[25] R. A. Fox and E. Jacewicz, "Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English," *The Journal of the Acoustical Society of America*, vol. 142, no. 1, pp. 444–459, 2017.

[26] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software (TOMS)*, vol. 22, no. 4, pp. 469–483, 1996.