ASA 169th Meeting
Pittsburgh, Pennsylvania
Session 2pSC, 19 May 2015
http://waisman.wisc.edu/vocal

High Fidelity Analysis of Vowel Acoustic Space

Michael H. Coen[a], Houri K. Vorperian[b], Ray D. Kent[b],

[a] Departments of Biostatistics & Medical Informatics and Computer Sciences, and [b] Waisman Center

Vocal Tract Development Lab, University of Wisconsin-Madison

(1) Introduction and Purpose

Vowel acoustic space is often characterized by polygons, whose vertices are determined by summary statistics such as mean values of the formant frequencies of distinct phonemes. The F1-F2 quadrilateral is the most familiar of these. However, using summary statistics to represent formant-frequency data presents fundamental limitations. These representations are inherently ***lossy***—summarizing large amounts of data with single values. Furthermore, *mean* itself is a non-robust statistic and highly sensitive to outliers. Even robust statistics ignore *distributional information* within the data, which can vary markedly among different phonemes and age groups.

We introduce a new approach characterizing and measuring changes in formant spaces statically and developmentally. This approach treats acoustic spaces as ***point clouds*** of data, in which little to no information is abstracted or lost. Within this framework, we measure the ***spatial overlap*** of sets of formant data using two approaches. The first is based on the geometric concept of a ***convex hull*** (*Krein & Šmulian, 1940*) and measures the overall range of formant and articulatory movement *(Flipsen & Lee, 2012; Vorperian & Kent, 2007)* *during vocal production*.

The second combines optimization theory and computational statistics to provide extremely sensitive measures of **spatial overlap** using all available data (*Coen et al., 2011*; *Coen et al., 2010*). This enables the highest precision characterization of the articulatory working space for vowels of which we are aware.

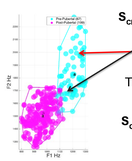
These ***point cloud*** approaches also enable detailed examination of developmental progression of individual vowels. In addition to providing new ways of charactering vowel acoustic space, they provide new tools for examining the fine-structure of the acoustic space associated with each vowel individually.

(2) Method

Acoustic data were collected from 150 participants between ages 4 to 40 years from monosyllabic words containing the corner vowels over /i u ae a /, as described in ().

Similarity between vowel acoustic spaces is a measure between 0 and 1, where 1 means the data are identical and 0 reflects they are entirely dissimilar according to some criterion. Similarity was measured in F1-F2 space using three methods:

1. **Vowel Space Area Similarity (S_{VSA})**, based on (*Vorperian & Kent, 2007*). For two vowel quadrilaterals V_1 and V_2, their similarity is defined as the area of their intersection divided by their larger area. This normalization insures $0 \le S_{VSA} \le 1$.
   $S_{VSA} = \text{Area}(V_1 \quad V_2)/\text{Max}(\text{Area}(V_1), \text{Area}(V_2))$
2. **Convex Hull Similarity (S_{CH})**, using a foundational concept in computational geometry of a convex polygon that minimally encloses a set of points (*Krein & Šmulian, 1940*). This can be computed over the entire vowel acoustic space or as a weighted average over the four point vowels in the /i u a ae/ quadrilateral. Similarity based on convex hulls relies on the extreme points in the convex representation of vowel acoustic data. To increase robustness, it may be computed as the weighted average over each vowel individually: For an individual comparison across two populations, $S_{CH}(v_i)$ is defined:

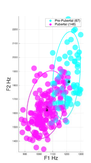


$$S_{CH}(v_i) \in [0,1] = \frac{\text{Area of convex hulls intersection}}{\text{Area of larger convex hull}} = \frac{4.9258 \times 10^3}{1.2622 \times 10^5} = .0390$$

Then for a set of vowels V:

$$S_{CH}(V) = \frac{\sum_{v_i \in V} |v_i| \, S_{CH}(v_i)}{|V|}$$

3. **Cloud Similarity (S_{Cloud})** uses all available data to compute a measure of *spatial overlap* between two sets of points. Statistically, it is the most robust of the techniques here and is not sensitive to outliers. Note that because point sets have *zero measure*, the likelihood of two real data points actually overlapping is also zero. Nonetheless, one can compute a rigorous, well-defined notion of their spatial overlap using the methods in (*Coen et al., 2011*; *Coen et al., 2010*). This approach normalizes the Kantorovich-Wasserstein distance, which is commonly determined using the Transportation Simplex algorithm (*Dantzig, 1963*). S_{Cloud} may be computed over the entire acoustic space or taken as the weighted average over each vowel as with S_{CH}:



$$S_{Cloud}(v_i) \in [0,1] = 1 - \frac{\text{Transportation Simplex Distance between Data}}{\text{Average Distance between Sets}} = 0.1539$$

Then for a set of vowels V:

$$S_{Cloud}(V) = \frac{\sum_{v_i \in V} |v_i| \, S_{Cloud}(v_i)}{|V|}$$

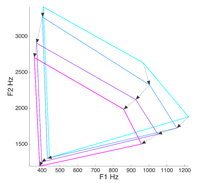
(3) Visualizing Similarity

Participants were divided in four age groups for the purposes of mathematically and visually evaluating the similarity measures:

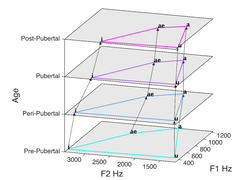
1. Pre-Pubertal: 48 to 96 months (4 to 8 years)
2. Peri-Pubertal: 96+ to 122.4 months (8 to 10.2 years)
3. Pubertal: 122.4+ to 174 months (10.2 to 14.5 years)
4. Post-Pubertal: 174+ to 301.2 months (14.5 to 25.1 years)

Pre-Pubertal
Peri-Pubertal
Pubertal
Post-Pubertal

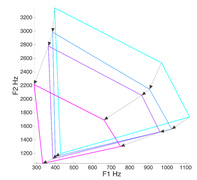
Vowel Space Area in 2D (Females)



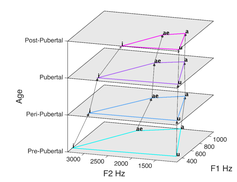
Vowel Space Area in 3D (Females)



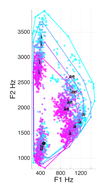
Vowel Space Area in 2D (Males)



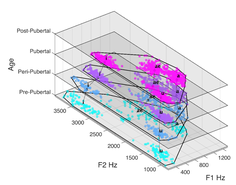
Vowel Space Area in 3D (Males)



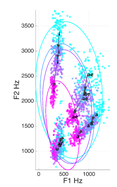
Convex Hulls in 2D (Females)



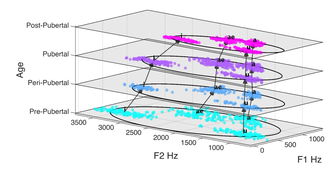
Convex Hulls in 3D (Females)



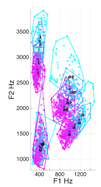
Point Cloud in 2D (Males)



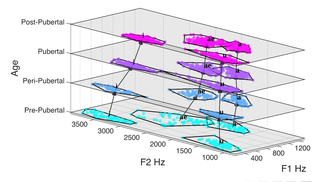
Point Cloud in 3D (Males)



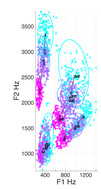
Convex Hulls Per Vowel 2D (Females)



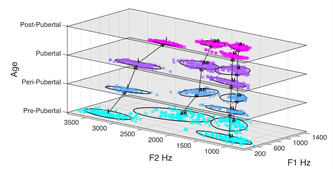
Convex Hulls Per Vowel 3D (Females)



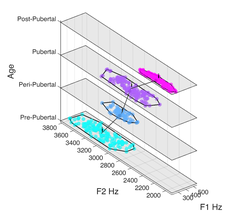
Point Clouds Per Vowel 2D (Males)



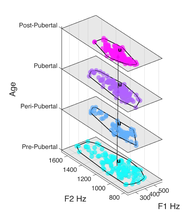
Point Clouds Per Vowel 3D (Males)



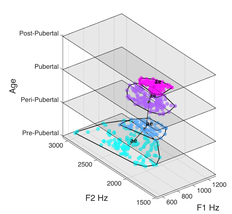
Convex Hulls for /i/ (Males)



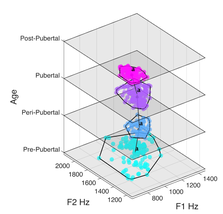
Convex Hulls for /u/ (Males)



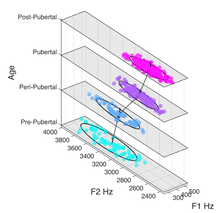
Convex Hulls for /ae/ (Males)



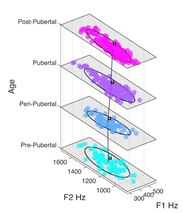
Convex Hulls for /a/ (Males)



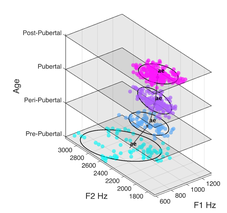
Point Cloud for /i/ (Females)



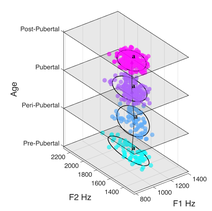
Point Cloud for /u/ (Females)



Point Cloud for /ae/ (Females)



Point Cloud for /a/ (Females)



(4) Results

Contrasting methods by examining sequential developmental changes, divided into three stages. In these tables, the **F1-F2** column examines the vowel format space in its entirety. The individual similarity measures for each quadrilateral vowel is contained where appropriate. The **Vowels** column presents the weighted averages of these, ignoring all "white space" between the vowels that arbitrarily inflate measurements in unpredictable ways due to empty space and can lead to unexpected non-monotonic dynamics simply due to epiphenomenal shrinkage in the acoustic space.

VSA Similarity Males

	F1-F2
$S_{VSA}(I)$	0.65148
$S_{VSA}(II)$	0.80222
$S_{VSA}(III)$	0.35161

Convex Hull Similarity Males

	F1-F2	/i/	/u/	/ae/	/a/	Vowels
$S_{CH}(I)$	0.61848	0.27476	0.45744	0.17621	0.20642	0.31865
$S_{CH}(II)$	0.88375	0.39324	0.7382	0.58359	0.54156	0.64511
$S_{CH}(III)$	0.49506	0.016544	0.52912	0.0026668	0.057704	0.19941

Point Cloud Similarity Males

	F1-F2	/i/	/u/	/ae/	/a/	Vowels
$S_{Cloud}(I)$	0.74405	0.079819	0.71267	0.12801	0.21623	0.31235
$S_{Cloud}(II)$	0.85313	0.27588	0.78048	0.49561	0.55653	0.54096
$S_{cloud}(III)$	0.57244	0.008813	0.44926	0.045385	0.055396	0.17237

VSA Similarity Females

	F1-F2
$S_{VSA}(I)$	0.82057
$S_{VSA}(II)$	0.67057
$S_{VSA}(III)$	0.73617

Convex Hull Similarity Females

	F1-F2	/i/	/u/	/ae/	/a/	Vowels
$S_{CH}(I)$	0.81027	0.50759	0.58935	0.23337	0.60982	0.435
$S_{CH}(II)$	0.78804	0.39603	0.59226	0.48777	0.57994	0.49853
$S_{CH}(III)$	0.81749	0.64317	0.76078	0.62962	0.49712	0.62498

Point Cloud Similarity Females

	F1-F2	/i/	/u/	/ae/	/a/	Vowels
$S_{Cloud}(I)$	0.83086	0.58767	0.74736	0.22431	0.45448	0.48071
$S_{Cloud}(II)$	0.78615	0.068099	0.52765	0.20071	0.42529	0.35721
$S_{Cloud}(III)$	0.82457	0.26949	0.76768	0.43768	0.30048	0.44578

Subjects consisted of 58 males and 73 females, who provided 3926 data points. 22 of these were deemed outliers for being further than 2σ from their vowel and demographic means.

Participants and Data	Pre-Pubertal	Peri-Pubertal	Pubertal	Post-Pubertal
Males/# of Data Points	14/498	7/276	18/550	19/522
Females/# of Data Points	11/302	10/273	21/635	31/848

In addition to sequential values, matrix-based pairwise comparisons among all stages can reveal developmental trajectories. These matrices are necessarily symmetric as all measures of similarity in Panel (1) are symmetric functions. Diagonals must contain only values of 1, as each data set (i.e., demographic) is identical to itself. For example, the cloud point measure, taken as a weighted average over the four point vowels has values:

S_{Cloud}	Pre-Pubertal	Peri-Purbertal	Pubertal	Post-Pubertal
Pre-Pubertal	1	0.48071	0.23175	0.15379
Peri-Purbertal	0.48071	1	0.35721	0.18662
Pubertal	0.23175	0.35721	1	0.44578
Post-Pubertal	0.15379	0.18662	0.44578	1

The matrix view compares the demographics pairwise. It demonstrates the monotonic convergence during development toward the post-pubertal stage and can illustrate the temporally distinct rates of change over the individual point vowels. The tables above can be derived from the first off diagonal of these matrices, which compare subsequent stages.

Interpreting the values in this matrix is straightforward. Consider the first column of numbers labeled **Pre-Pubertal**. It is most similar to itself, reflected in S_{Cloud}=1. Moving down this column reflects the acquisition of increasing articulatory control and anatomic growth, leading to each subsequent age grouping being **less** similar to the **Pre-Pubertal** stage. At the bottom of that column, the similarity between Pre-Pubertal and Post-Pubertal reaches its nadir value of 0.15379, reflecting maximum dissimilarity between the most juvenile and mature ages examined.

(5) Discussion/Conclusions

This study introduced new, robust methods for characterizing vowel acoustic space. Among the most important observations of this work is that approaches for measuring acoustic vowel space in its entirety can be extremely *insensitive* even if they do not make use of non-robust statistics such as *mean* in VSA. More subtly**, en masse formant space approaches "reward" the empty space between the vowels** because of their reliance on area. Doing so has no meaningful interpretation, varies enormously by coordinate system (e.g., when conducting the analysis in a logarithmically compressed space modeling cochlear compression; does not enable formal evaluation of articulatory evolution in individual vowels during development; and makes comparisons with higher formants such as F3 and F4 impossible to interpret.

Thus, the methods in this study:

1. Provide vowel-level as well as acoustic space-level measures of change.
2. Are statistically robust with respect to outliers and noise and are not "lossy." The cloud point exploits all available data rather than employing proxy statistics.
3. Exploits distributional information, as with the cloud point model. In contrast, the convex hull approach is useful for gauging movement extremes but is oblivious to the actual distribution of the data points themselves *within* the convex hull. The cloud point approach gives all data points "equal" importance and reflects changes in any of them. Because it makes use of all points, outliers and noise and increasingly negligible effects as the amount of data grows.
4. Enable tracking of differential development of vowel production during the lifespan.
5. Reveal complex interior structure within vowels individually in acoustic space.

These may provide useful approaches for studying articulatory development, modeling impaired speech, and elucidating distinct "clusters" of development patterns.

References

1) Coen, M.H., Ansari, M. H., and Fillmore, N. (2011). Learning from Spatial Overlap. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence* (AAAI-11). San Francisco, CA.
2) Coen,M.H., Ansari, H, and Fillmore, N. (2010). Comparing Clusterings in Space. In *Proceedings of the 27th International Conference on Machine Learning* (ICML'10). Haifa, Israel.
3) Dantzig, G. (1963). Linear programming and extensions. Princeton University Press. Princeton, NJ.
4) Flipsen, P., and Lee, S. (2012). Clin. Linguist Phon., 26, 926-33.
5) Krein, M., and Šmulian, V. (1940), "On regularly convex sets in the space conjugate to a Banach space", Annals of Mathematics, 2nd ser. 41: 556–583
6) Vorperian, H.K., and Kent, R.D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *J. Speech Lang. Hear. Res.*, *50*, 1510-1545.

Acknowledgments

Work was supported by NIH grants #R01-DC 006282 & P30-HD03352. Special thanks for participating in the development of the mathematical and *in-situ* statistical framework for point cloud similarity to: Hidayath Ansari, Marissa Phillips, and Theodora Hinkle. Thanks for data collection and acoustic analysis to: Carlyn Burris, Erin Douglas, Ekatarini Derdemezis, Sara Kurtzweil, Katie Lester, Jen Lewandowski, Erin Nelson, Allis on Petska, and Alyssa Wild.