Vowel Space Density as a tool to guide language documentation

Eric Jackson, SIL International
UCLA American Indian Seminar, 2021-04-06

This talk presents a pilot study evaluating the concept of *Vowel Space Density* for use in dialectometry and early planning of language documentation and development efforts. The Vowel Space Density for a sample of connected speech can be calculated automatically from data that can be collected relatively quickly. Although initial results suggest that this may be a useful tool for characterizing variation within closely-related speech varieties, significant challenges remain to be addressed before it can be effectively and confidently applied to this problem, and different methods of data collection may be better for guiding documentation planning or language survey design.

1 Vowel Space Density: what and why

1.1 Statement of the problem

- The need for linguistic documentation:

  – 7139 ISO 639-3 language codes, 8516 Glottolog entries (Hammarström et al 2020)

  – WALS (Dryer and Haspelmath 2013) includes only 2662 language entries

  – PHOIBLE 2.0 (Moran and McCloy 2019) includes 3020 inventories from 2186 "distinct languages"

  – "*the vast majority of the world's languages remain under-documented (or even* **un***documented)*" (Maddieson 2016)

- SIL's perspective: good documentation is the foundation for efficient and sustainable language development
- This is lots of work! Let's work smarter, as well as harder. So…

(1) *How can we quickly estimate linguistic differences between under-described speech varieties, in order to better guide subsequent, more detailed documentation work?*

- My personal perspective: China
  – 280 indigenous languages with ISO codes, only 46 are “institutional” or “developing” (EGIDS scale 0-5 (Lewis and Simons 2010), used on Ethnologue.com)
  – a large fraction documented to some extent (they at least have an ISO code)
  – many are severely lacking in documentation
  – up to 150 vital enough to warrant community language development efforts, but which are under-documented or mostly undocumented
  – One language survey can take up to a year to plan and carry out
- How can we speed up the planning phase for data collection with methods that can quickly analyze large amounts of data with minimal processing or preparation?

1.2 Vowel Space Density (VSD): What is it? How is it calculated?

- I'm following methods described by Brad Story and Kate Bunton at the University of Arizona (Story and Bunton 2017; Story, Bunton, and Diamond 2018)
- Similar methods are found in Sandoval et al (2013) and Fox and Jacewicz (2017)
- The general procedure is:

(2) step 1: measure continuous F1 and F2 from connected speech
(step 1a: throw out the “bad” data points)
(step 1b: normalize the formant values for better comparison between speakers)
step 2: calculate a kernel density estimate (KDE) from F1-F2 bins
step 3: calculate a convex hull that encloses this KDE distribution

(the convex hull is typically calculated at a low but non-zero KDE value; Story & Bunton use a value of 0.25 on a normalized scale of 0-1)

- VSD has been used in speech pathology (Story & Bunton, including development in the same speaker over time) and in dialectometry (Fox & Jacewicz)

- *How might VSD be used to quickly estimate language variation, with just an untranscribed audio recording from multiple locations, for subsequent follow-up?*
  – would require not just an absolute area, but comparison of the distribution of formants within F1/F2(/F3) domain

1.3 What variables might affect VSD?

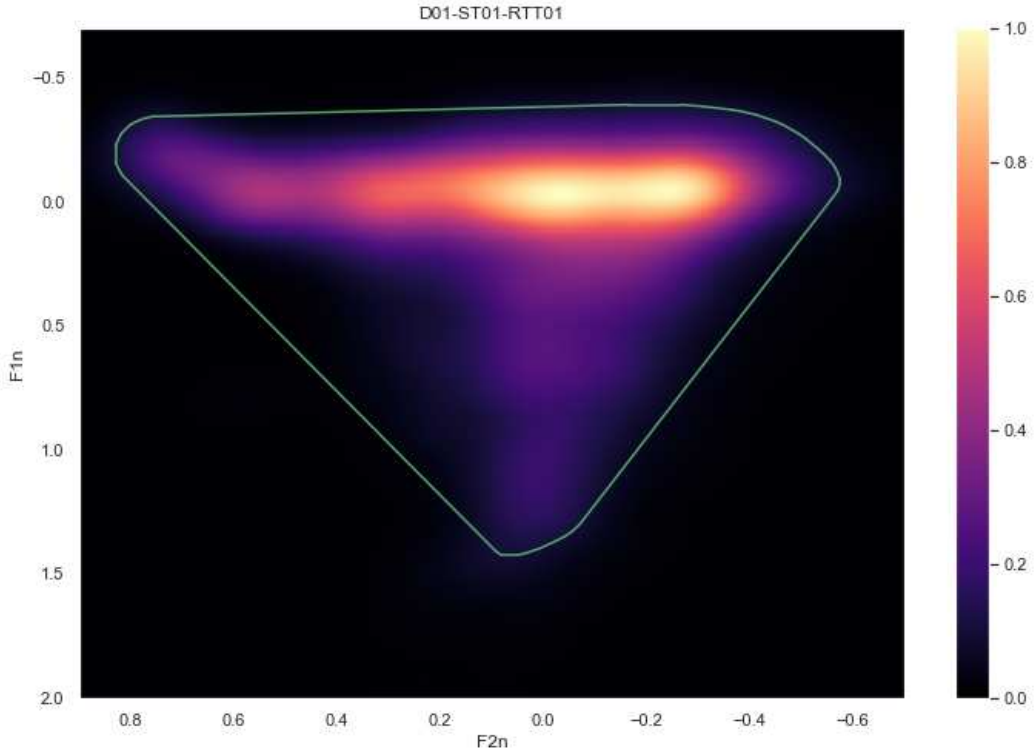
(3) **How does VSD vary for the same speaker in different (short) speech samples?**
– Does the lexical content of a story dominate VSD, or are there characteristics that are constant regardless of lexical or narrative content?

- Pilot data: various personal narratives from a dialect survey in China (Jackson, Jackson, and Lau 2012); full data in this set includes
  – wordlists, 144 or 490 words, transcribed and recorded with audio, 18 locations
  – personal narratives, transcribed and recorded, 9 locations (often, more than one narrative was recorded at a location, but usually only one was then transcribed)
  – comprehension testing of these personal narratives at 19 locations
  – individual and group questionnaires on sociolinguistic attitudes, 18 locations

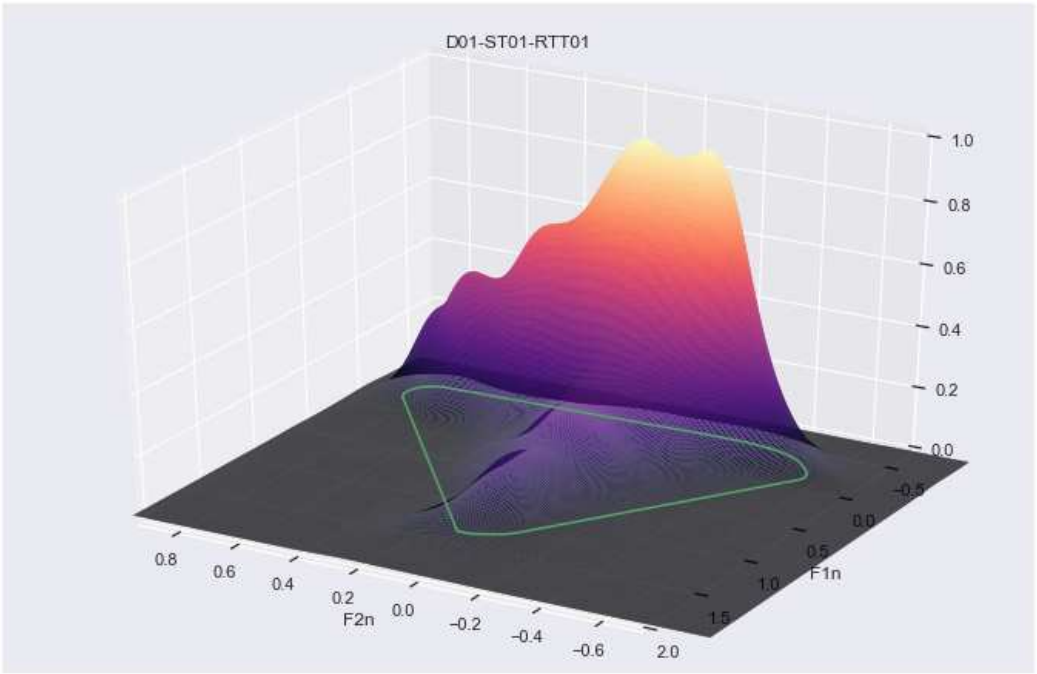
- Here, VSD calculated from two personal stories, about 3 minutes each, told by the same female speaker of Yang Zhuang [zyg] (a Tai language of southern China)
  – Orientation of these plots should be that of a standard vowel trapezoid, with high-front at the upper left

– story 1, D01-ST01-RTT01 “Gangsters,” in (4), (5)

(4)

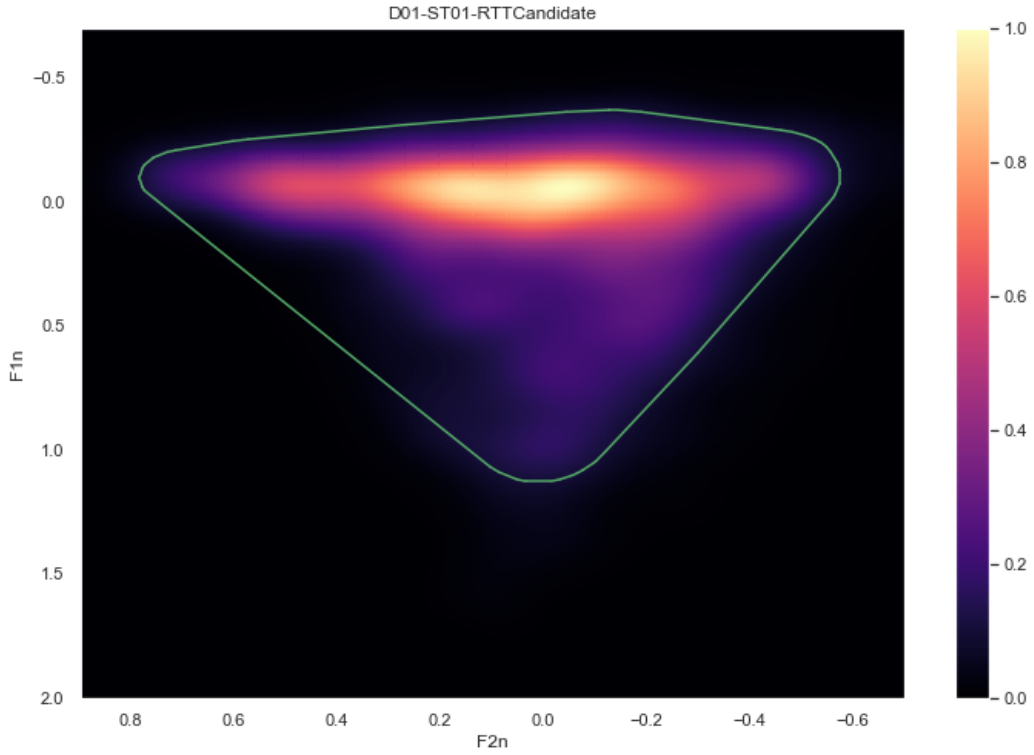


(5)

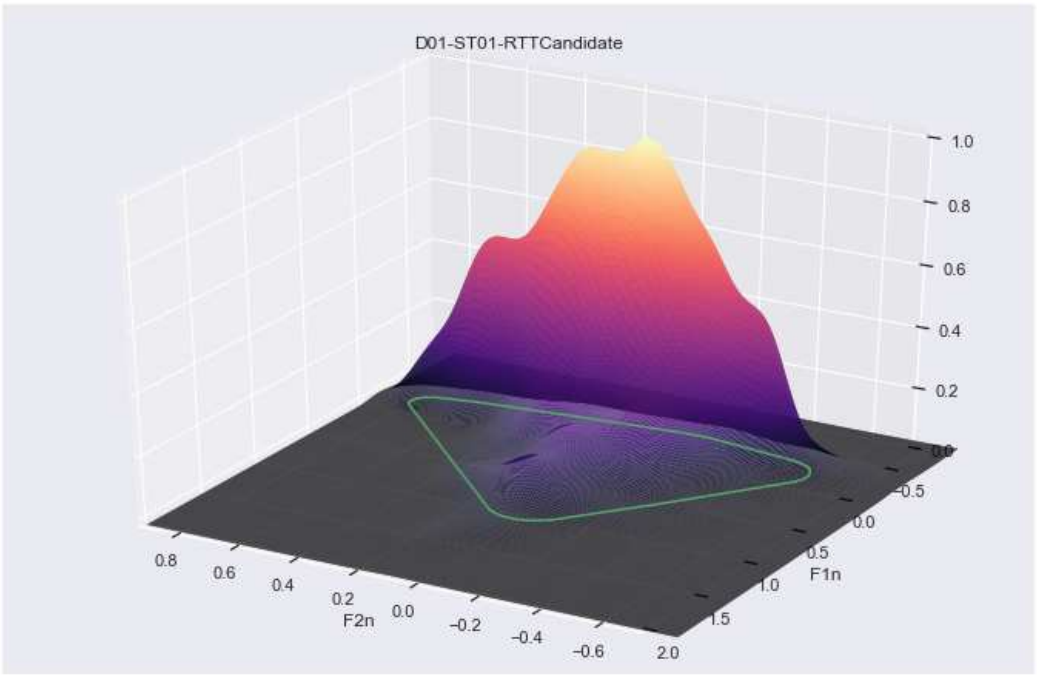


– story 2, D01-ST01-RTTCand “Being cheated,” in (6), (7)

(6)

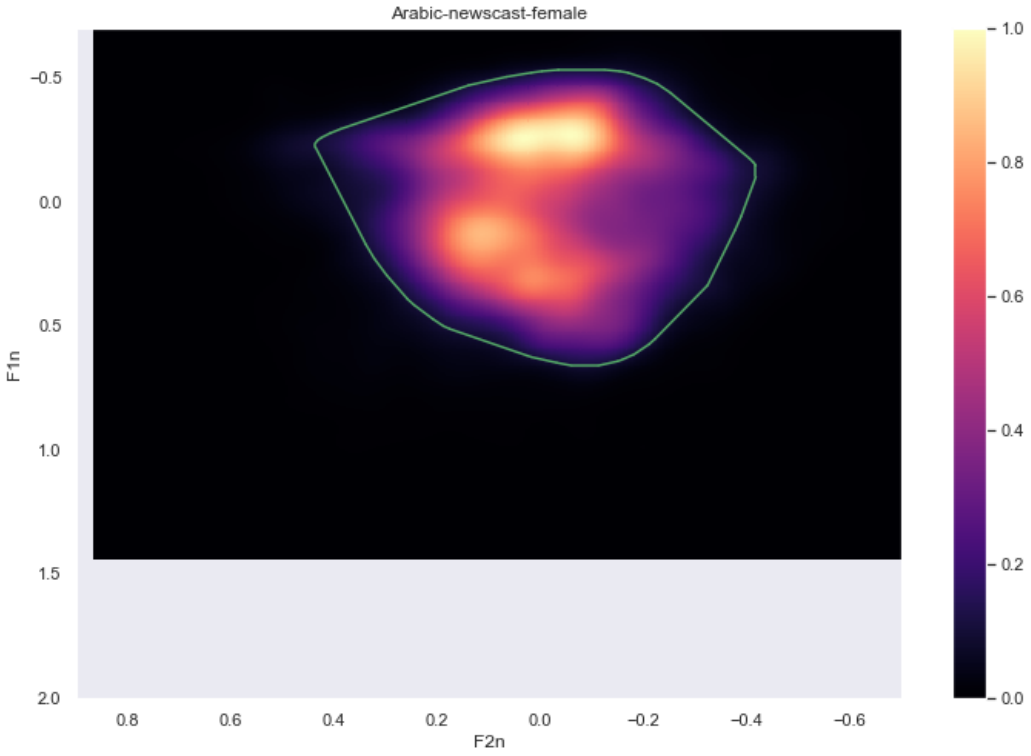


(7)

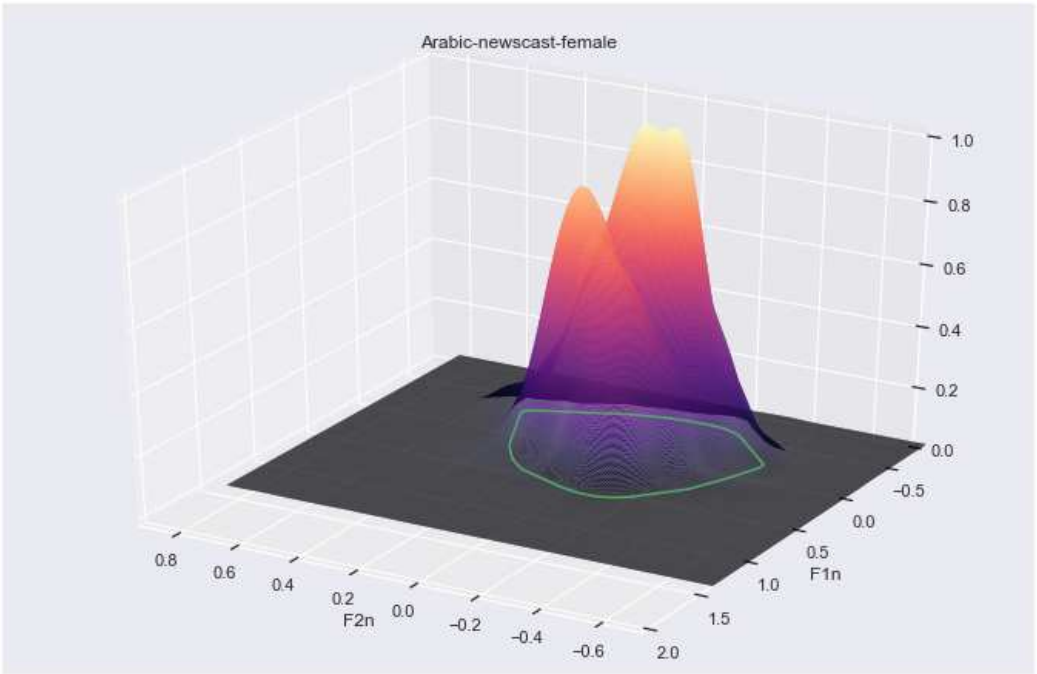


(8) **How does VSD vary for similar speakers speaking very different languages?**
– Compare the VSD for the female Zhuang storyteller in (4) – (7) with that of a female Arabic-speaking newscaster in (9), (10) (taken from YouTube)

(9)

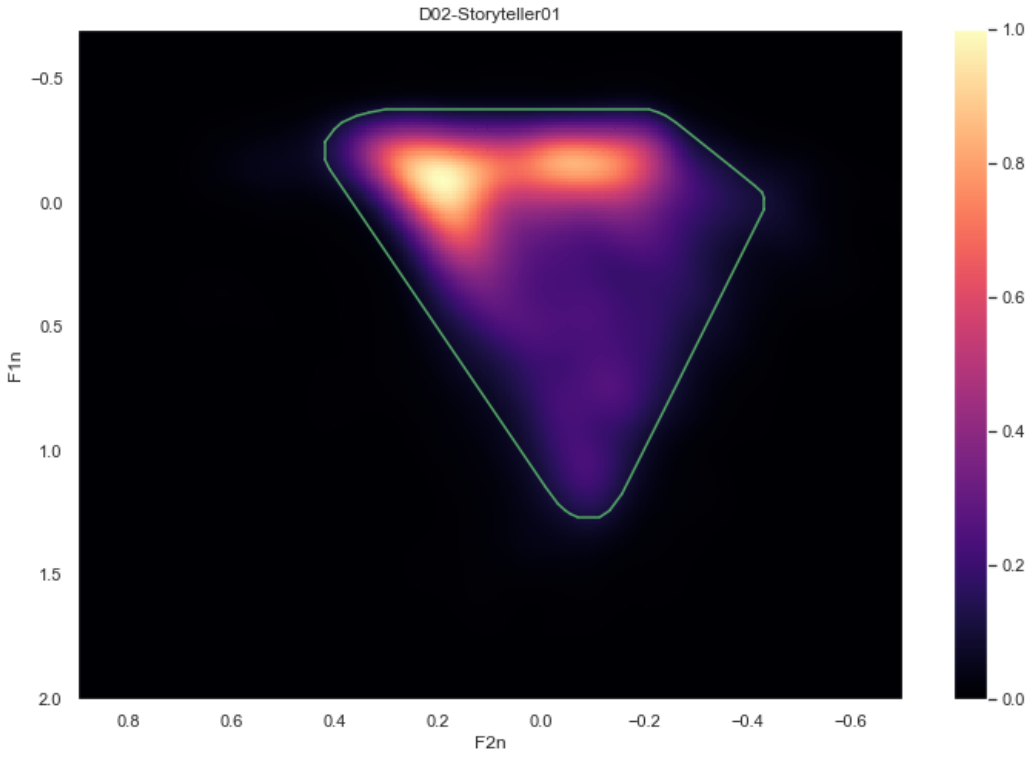


(10)

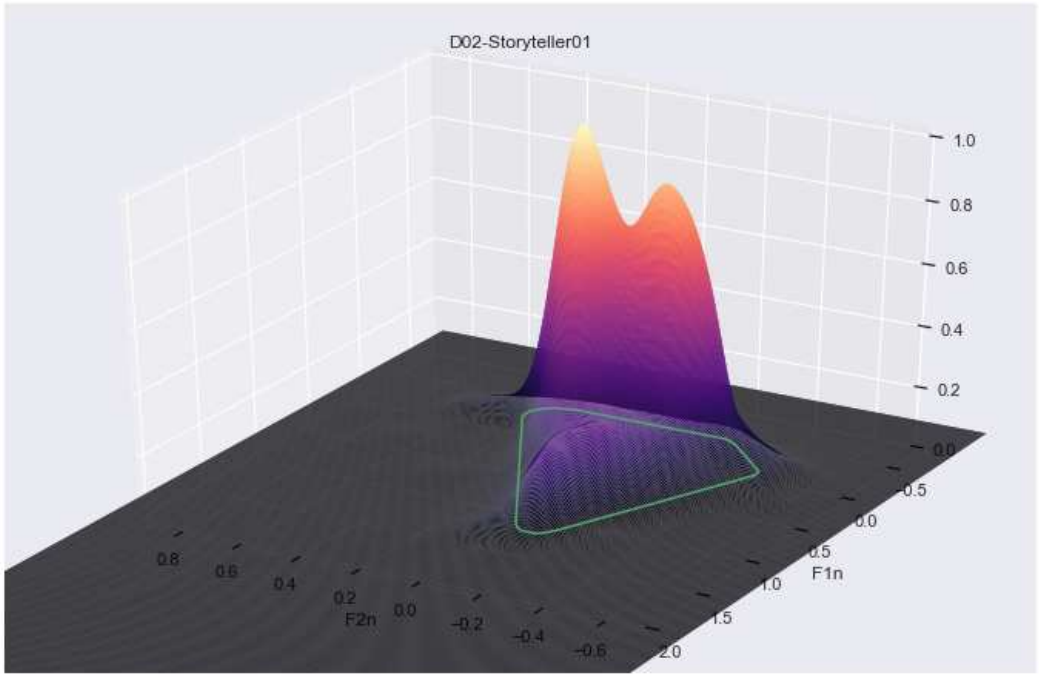


(11) **How does VSD vary for different speakers of the same local variety?**
– speaker D02-ST01, male, “Lifesaving” (story D02-ST01-RTTCand)

(12)

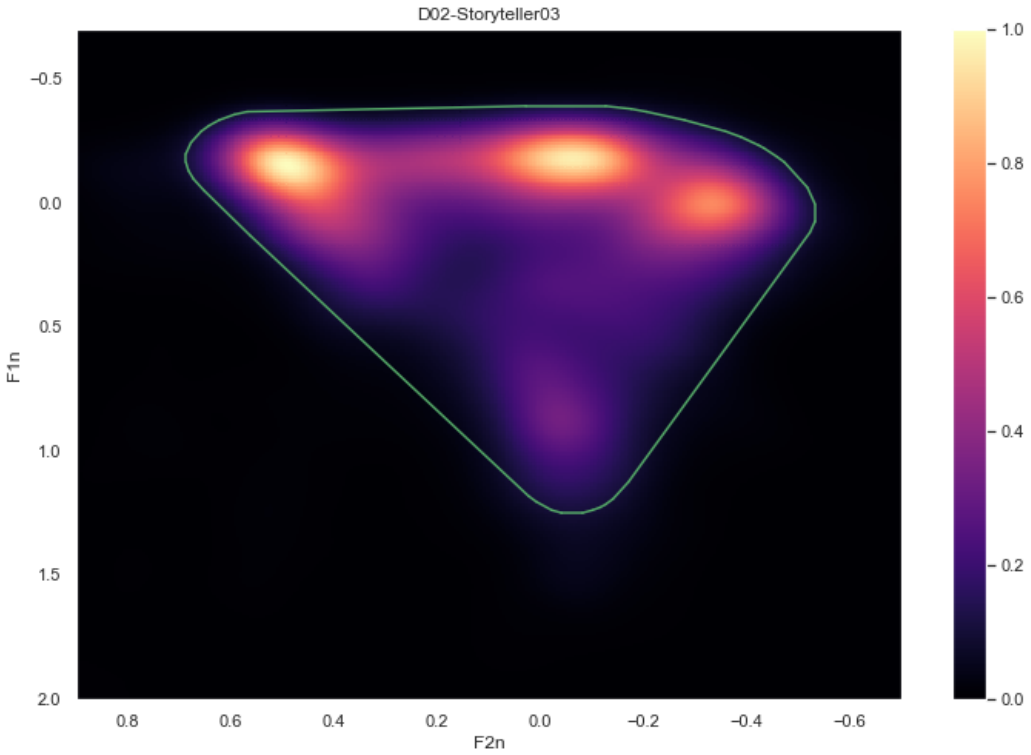


(13)

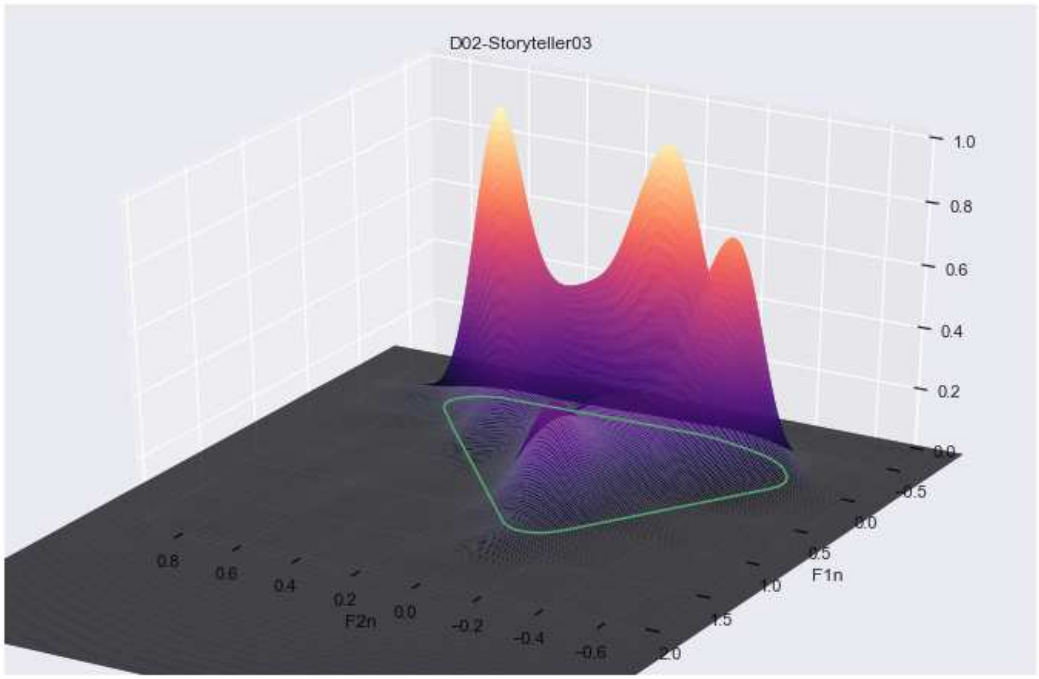


– speaker D02-ST03, male, “Bitten by Bees” (story D02-ST03-RTT02)

(14)

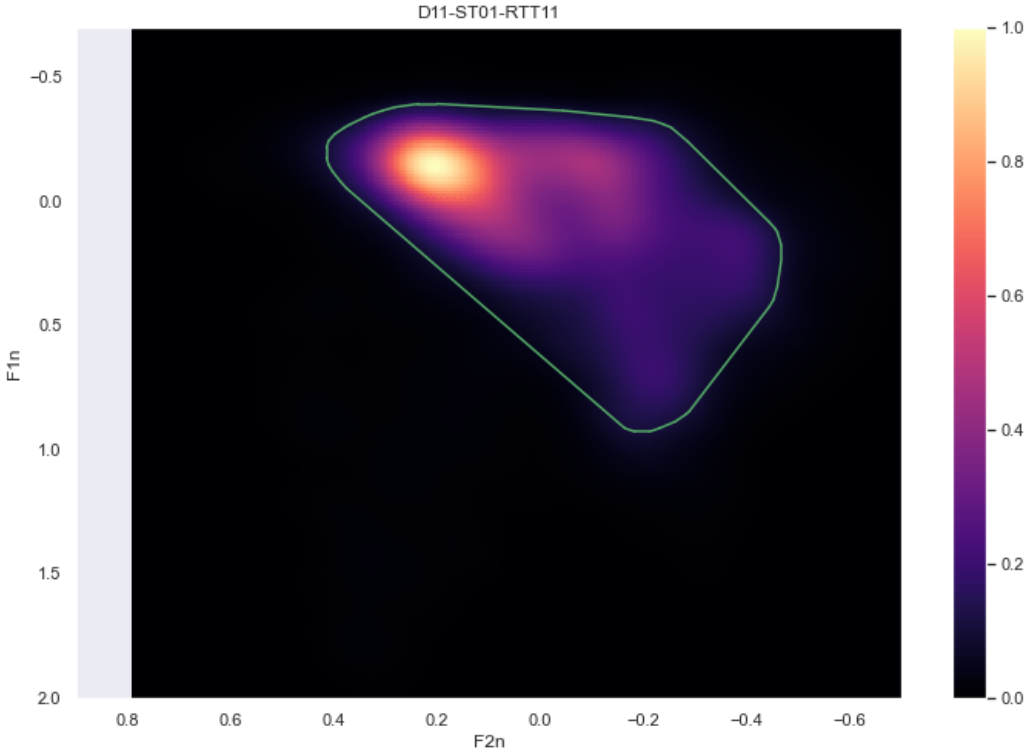


(15)

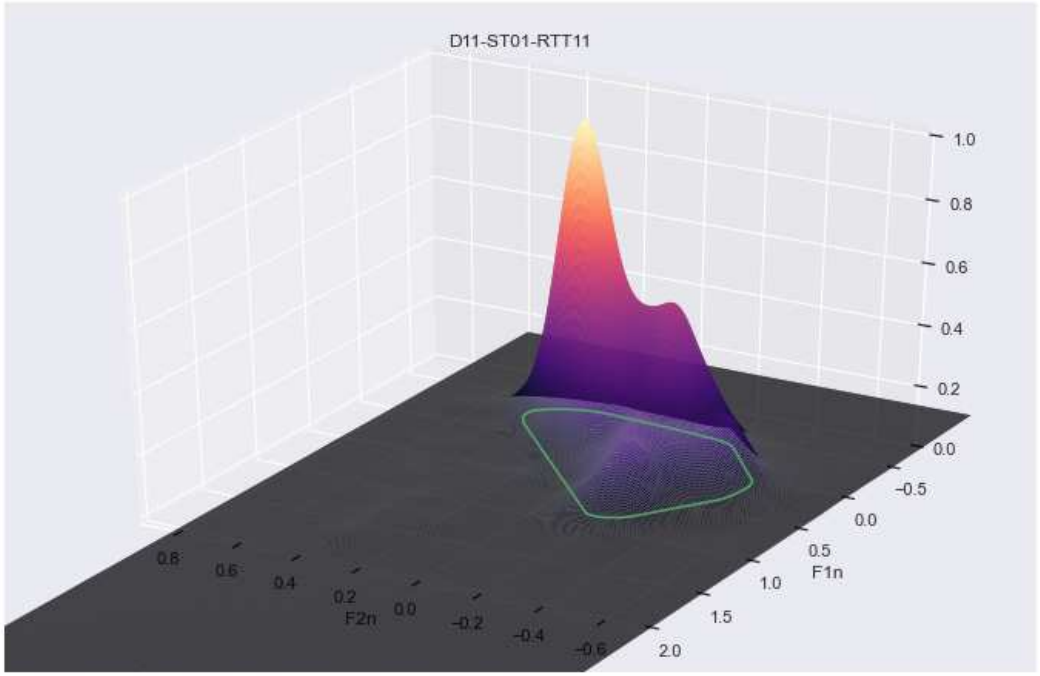


(16) **How does VSD vary for similar speakers from different local varieties?**
– speaker D11-ST01, male, “Kicking donkey” (D11-ST01-RTT11), Minz Zhuang [zgm]

(17)

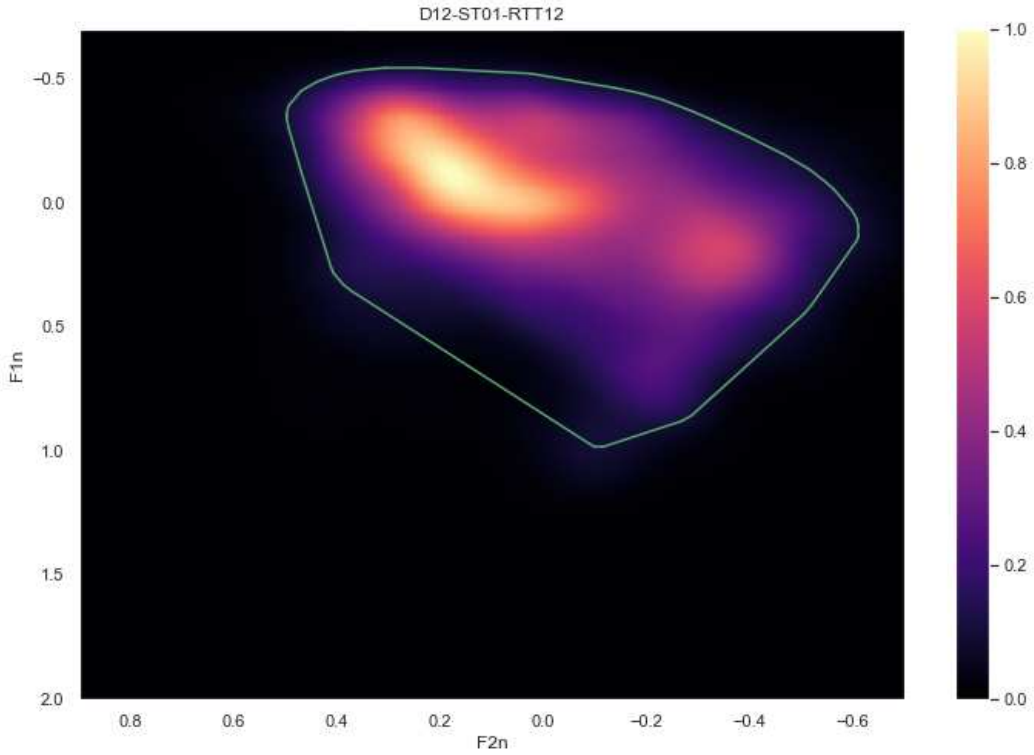


(18)

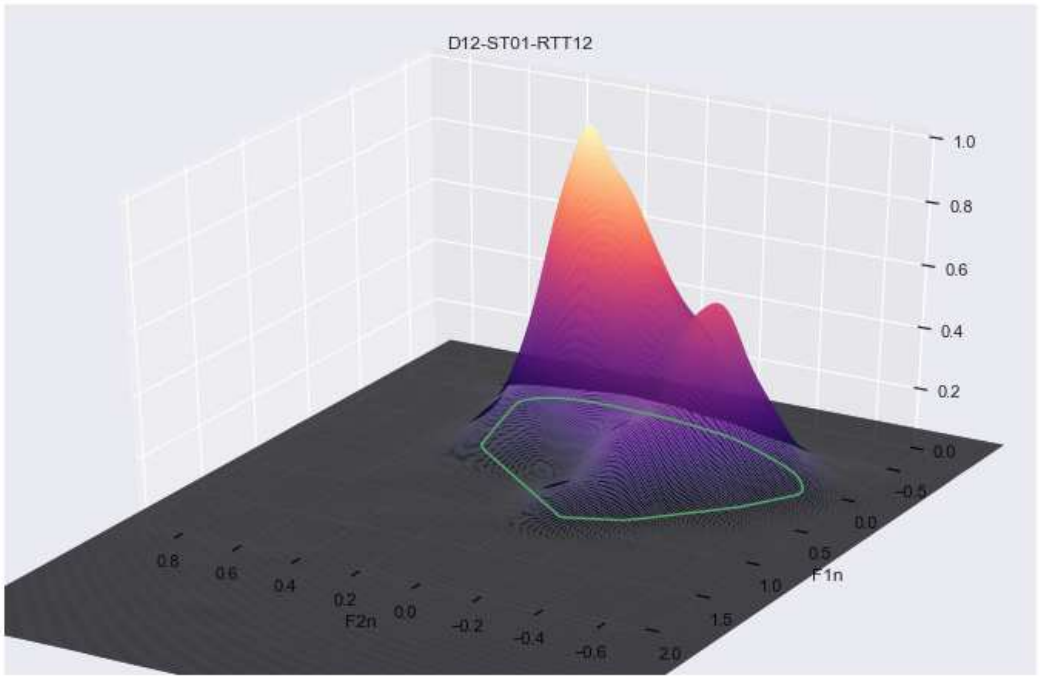


– speaker D12-ST01, male, “Bitten by Wasps” (D12-ST01-RTT12), Nong'an Zhuang

(19)

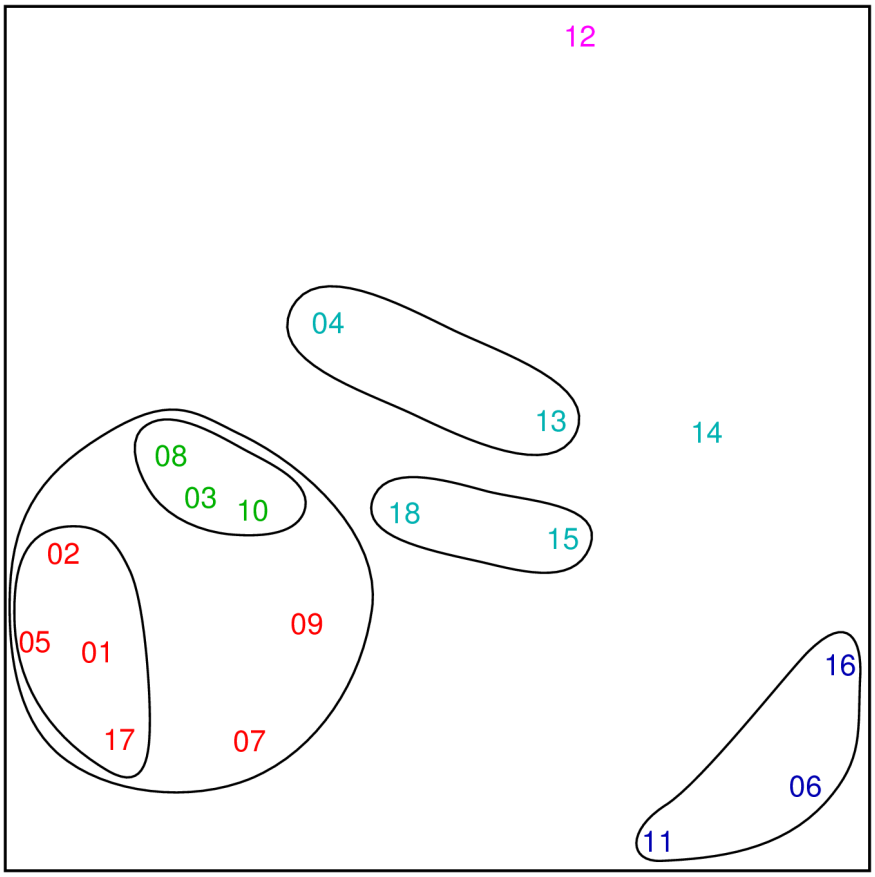


(20)



- For comparison, this figure from Jackson, Jackson, and Lau (2012) shows a clustering analysis of these language varieties (and others) based on wordlist similarity
  – points represented by VSD plots so far are 01, 02, 11, and 12
  – VSD for locations 01 and 02 (figures in (4) – (7), (12) – (15)) look more similar than any of these does to locations 11 or 12 (figures in (17) – (20)), but we need to characterize this by more than simple inspection

(21)



2 Stepping back: The problems

- Beautiful plots (thanks, MatPlotLib!), but can this tool really do what we want?

Problem 1: Representativeness of VSD extent

- For a proper characterization of the extent of used vowel space (ie, the convex hull area), we need to ensure all the sounds of the language are represented in the speech sample.

  – Plausible that a 3-minute sample of natural speech might be representative in this way, but need a way to verify (without resorting to transcription)

Problem 2: Calculating differences in VSD extent

- Convex hull area can give a scalar value for a speaker and speech sample, but how can we characterize not just a different overall area, but different *shapes* in vowel space?

Problem 3: Representativeness of VSD distribution

- For a proper characterization of the way that vowel space is used *in the language as a whole*, we need a speech sample with accurate representation of the token frequency of the entire lexicon.

  – This will obviously take a larger sample than a 3-minute personal story! Would hours or a full day of ambient conversation reflect this well?

- Short (~3 minute) samples of connected narrative will have certain lexical items occurring with higher-than-normal frequency, skewing the density

  – Because of this, in short samples, it's probably not realistic to draw conclusions from individual peaks in the VSD rather than the general extent of the VSD.

Problem 4: VSD is sensitive only to dialect differences in sonorants

- VSD characterizes differences only in formants; it cannot represent differences in obstruents (example: Iberian Spanish /s/), tone systems, or prosodic features

Problem 5: Is normalization masking inter-speaker differences?

- Normalization here follows Story & Bunton (2017), who follow just one normalization method suggested in Disner (1980); for formant pairs [F_1(n), F_2(n)]:

(22) $$F_j^*(n) = \frac{F_j(n) - F_j^{median}}{F_j^{median}}, \; j=\{1,2\}, \; n=\{1 \ldots N\}$$

- For female speakers in (4) (Yang Zhuang) and (9) (Arabic), how much difference in "low" vowel extent is due to normalization against a different median F1?

Problem 6: Is "mass-produced" formant accuracy sufficient?

- Recent studies examining accuracy of formant measurements (Harrison 2013, Derdemezis et al 2016) with tools including Praat (Boersma and Weenink 2021) find average errors as high as hundreds of Hz

  – Errors can be reduced by tailoring LPC analysis parameters—so using a common set of parameters across a large speech sample, or multiple samples, will produce errors

  – How sensitive is VSD to variations or errors in the input data? How big can these become before they significantly alter the convex hull area calculation?

Problem 7: More data is needed to verify generalizability of the method

- This data set is still too small to verify that single speakers from a speech community are likely to be representative of that community as a whole

  – In this data, there are at most three speakers from one location (all male), and some locations have only one speaker.

  – Improved data: TED talks? (with a verifiable way to determine dialect of speakers)

3 Where to go from here

- Competent speakers of a language, with past exposure to variants, can quickly identify a variant based on a short speech sample (for example, Ruch 2018)
  - Doesn't this require a speaker who knows the language(s)/variants? Humans can't usually do this for languages and varieties they haven't been exposed to
  - Suggests: Lexical knowledge is also important in this evaluation—such as, *this* lexical item/phoneme has *this* phonetic realization

- How to combine a raw phonetic model like this with lexical knowledge? Maybe string-edit distance of wordlists, or raw phonetic comparison of cognate wordlist items, would better approximate what speakers can do
- Automated Dialect Identification is an area of active research (for a recent survey, see Etman and Beex 2015), but typically requires training on existing data

- How can we create workflows to handle large amounts of audio data and perform useful tasks on them?
  - This method is widely available and widely replicable: Praat, Python, Parseltongue (getting data into a Pandas DataFrame means that many tools can be applied to it)
  - All code from this study is available at http://github.com/euangeleo/vsd-verify
- Is this a sufficiently useful workflow for finding formants in connected speech? What is still needed?
  - Next: implement a filter to remove outliers / bad formant values
  - If LPC parameters strongly affect formant measurements, is there any hope of automating parameter selection to improve accuracy of results?
- What other purposes might the typical range of data from an intelligibility survey be useful for?
- What methods can improve the efficiency of intelligibility survey data, or improve the speed and quality of documentation of language variation?
  - Can this method suggest that speech at location A may closely resemble speech at location B? *Maybe*
  - Is this the best or easiest way to collect that kind of information? *Maybe not; asking local speakers their sociolinguistic judgments might provide similar information with more confidence*

4 References

Boersma, Paul and David Weenink. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.41, retrieved 25 March 2021 from http://www.praat.org/

Derdemezis, Ekaterini, Houri Vorperian, Ray Kent, Marios Fourakis, Emily Reinicke, and Daniel Bolt. 2016. Optimizing Vowel Formant Measurements in Four Acoustic Analysis Systems for Diverse Speaker Groups. *American Journal of Speech and Language Pathology* 25(3):335-354. doi:10.1044/2015_AJSLP-15-0020

Disner, Sandra. 1980. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*, 67(1), 253–261.

Dryer, Matthew S. & Martin Haspelmath (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info)

Etman, Asmaa and A. A. (Louis) Beex. 2015. Language and Dialect Identification: A survey. Paper presented at the SAI Intelligent Systems Conference. doi:10.1109/IntelliSys.2015.7361147.

Fox, R. A., & Jacewicz, E. 2017. Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. *Journal of the Acoustical Society of America*, 142(1), 444–459. http://doi.org/10.1121/1.4991021

Hammarström, Harald, Robert Forkel, Martin Haspelmath, & Sebastian Bank. 2020. *Glottolog 4.3*. Jena: Max Planck Institute for the Science of Human History. https://doi.org/10.5281/zenodo.4061162. (Available online at http://glottolog.org)

Harrison, Philip. 2013. Making accurate formant measurements: An empirical investigation of the influence of the measurement tool, analysis settings and speaker on formant measurements. York, UK: University of York dissertation. (Available online at http://etheses.whiterose.ac.uk/7393/1/Harrison%20-%20PhD%20Thesis.pdf)

Jackson, Eric, Emily Jackson, and Shuh-Huey Lau. 2012. A Sociolinguistic Survey of the Dejing Zhuang Dialect Area. SIL Electronic Survey Reports 2012-036. (Available online at http://www.sil.org/resources/archives/50901)

Lewis, Paul and Gary Simons. 2010. Assessing endangerment: Expanding Fishman's GIDS. *Revue roumaine de linguistique*. 55 (2): 103–120. (Available online at https://www.lingv.ro/RRL-2010.html)

Maddieson, Ian. 2016. Under-documented languages expand phonetic typology. *Journal of the Acoustical Society of America*, 139(4): 2212. DOI: 10.1121/1.4950614. (Slides available at https://www.researchgate.net/publication/302070096_Under-documented_languages_expand_phonetic_typology)

Moran, Steven & Daniel McCloy (eds.) 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. (Available online at http://phoible.org)

Ruch, Hanna. 2018. The Role of Acoustic Distance and Sociolinguistic Knowledge in Dialect Identification. *Frontiers of Psychology* 9:818. doi:10.3389/fpsyg.2018.00818

Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. 2013. Automatic assessment of vowel space area. *Journal of the Acoustical Society of America*, 134(5), EL477–83. http://doi.org/10.1121/1.4826150

Story, Brad and Kate Bunton. 2017. Vowel space density as an indicator of speech performance. *Journal of the Acoustical Society of America* 141(5), EL458-464. http://dx.doi.org/10.1121/1.4983342

Story, Brad, Kate Bunton, and Rebekkah Diamond. 2018. Changes in vowel space characteristics during speech development based on longitudinal measurements of formant frequencies. Paper presented at the 175th Meeting of the Acoustical Society of America.

Eric Jackson, SIL International
eric_jackson@sil.org
This handout may be downloaded from https://www.academia.edu/45687918/