8-2014

An assessment of acoustic contrast between long and short vowels using convex hulls

Erin F. Haynes

Michael Taylor
Santa Clara University, mjtaylor@scu.edu

Recommended Citation

Haynes, E. F., & Taylor, M. (2014). An assessment of acoustic contrast between long and short vowels using convex hulls. The Journal of the Acoustical Society of America, 136(2), 883–891.

An assessment of acoustic contrast between long and short vowels using convex hulls

Erin F. Haynes[a)]
American Institutes for Research, Waltham, Massachusetts 02451

Michael Taylor
School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

(Received 24 November 2013; revised 22 June 2014; accepted 26 June 2014)

An alternative to the spectral overlap assessment metric (SOAM), first introduced by Wassink [(**2006**). J. Acoust. Soc. Am. **119**(4), 2334–2350], is introduced. The SOAM quantifies the intra- and inter-language differences between long–short vowel pairs through a comparison of spectral ($F1$, $F2$) and temporal properties modeled with best fit ellipses ($F1 \times F2$ space) and ellipsoids ($F1 \times F2 \times$ duration). However, the SOAM ellipses and ellipsoids rely on a Gaussian distribution of vowel data and a dense dataset, neither of which can be assumed in endangered languages or languages with limited available data. The method presented in this paper, called the Vowel Overlap Assessment with Convex Hulls (VOACH) method, improves upon the earlier metric through the use of best-fit convex shapes. The VOACH method reduces the incorporation of "empty" data into calculations of vowel space. Both methods are applied to Numu (Oregon Northern Paiute), an endangered language of the western United States. Calculations from the VOACH method suggest that Numu is a *primary quantity* language, a result that is well aligned with impressionistic analyses of spectral and durational data from the language and with observations by field researchers.
 [http://dx.doi.org/10.1121/1.4887479]

PACS number(s): 43.72.Ar [SAF] Pages: 883–891

I. INTRODUCTION

The Spectral Overlap Assessment Metric (SOAM) is a method introduced by Wassink (2006) in response to the need to quantify the differences between so called *primary quality* and *primary quantity* language classifications. The SOAM is an important tool for conducting both cross-language vowel system comparisons and within language comparisons of vowel contrasts. In this paper, we propose an improvement to the SOAM to increase its accuracy, particularly for use in endangered language documentation, an area of growing interest to linguists. The proposed method, called the Vowel Overlap Assessment with Convex Hulls (VOACH) method uses best-fit convex shapes, reducing the danger of incorporating "empty" data into calculations of vowel space.

A. The SOAM

In the SOAM, normalized $F1$ and $F2$ frequencies for a long and short vowel pair are compared in a two-dimensional system. Each vowel is plotted with a best-fit ellipse, and the two ellipses are analyzed for the percentage of tokens that fall within the overlapping regions of the ellipses (i.e., the regions of $F1 \times F2$ space occupied by both the long and short vowel). The overlap indicates to what degree the long–short vowel pair is similar in spectral measures. Next, duration is added to the model as a third dimension, and the long and short vowel data are plotted with best-fit ellipsoids. As in the two-dimensional model, the ellipsoids are analyzed for the percentage of overlapping tokens. This overlap indicates the degree to which the long and short vowel pair is the same as a function of both spectral and durational measures. If the overlap calculated in the two-dimensional (spectral) model is substantially larger than the overlap calculated in the three-dimensional (3D) model, we can conclude that the vowels are distinguished primarily by duration (*primary quantity*), because the addition of duration to the model decreases their co-occurrence. On the other hand, if there is little overlap in the spectral dimension and/or substantial overlap in the temporal dimension, the vowels are distinguished primarily by spectral information (*primary quality*).

The SOAM is an important tool in documenting and classifying vowel systems according to both their spectral and temporal features. However, as Wassink (2006) notes, the use of best-fit ellipses is perhaps not ideal:

> "There is a continued need within experimental phonetic research to critically consider the use of the ellipse and the ellipsoid to represent vowel distributions in acoustic space. This convention appears to exist out of convenience while having no basis in auditory or acoustic reality (p. 2343)."

She states that ellipses fit her data without leaving large distribution gaps, but calls for a method that is better motivated by the geometric realities of the data, ideally one that is fitted on a by-vowel basis. Wassink's call is even more urgent in cases where data are not widely available and

[a)]Author to whom correspondence should be addressed. Electronic mail: ehaynes@air.org

where the issue of distribution gaps may be more pronounced, for example in under-described and endangered languages. Morrison (2008) provides suggestions for improving the computation of vowel overlap in the SOAM by better utilizing the underlying assumption that the data are normally distributed. We argue later in this paper that assuming that the acoustic properties of vowels are described by a Gaussian distribution ignores non-random variation due to contextual and environmental factors. This variation is particularly acute in endangered languages and calls into question the appropriateness of such a statistical foundation in analyzing vowel properties.

In this paper, we attempt to improve upon the SOAM through the use of best-fit convex shapes, a method originally proposed by Brubaker and Altshuler (1959) to analyze overlap in $F1 \times F2$ space. Their shapes were apparently created and analyzed painstakingly by hand, making application to two-dimensional datasets very difficult and extension to three dimensions completely intractable. Today, this approach is much more feasible in both two and three dimensions due to the development of standardized and efficient convex hull algorithms in the field of computational geometry and their inclusion in widely available commercial numerical packages such as MATLAB.

As we show, this method allows for more precision in calculating overlap by minimally accounting (with respect to other convex shapes, e.g., ellipsoids) for space that is not occupied by data. As a result, this method, the VOACH method, is favorable for datasets with relatively few data points. While the issue of small datasets is unlikely to pose a problem for well-studied languages and languages with large numbers of speakers (e.g., English), the majority of the world's languages today are endangered, with rapidly dwindling numbers of speakers (Krauss, 1992). Unfortunately, the fact that they are endangered makes it even more important to accurately document their linguistic features (Linguistic Society of America, 1994), including phonetic features. We apply the VOACH method to peripheral vowel data from Numu (Oregon Northern Paiute), an endangered Uto-Aztecan language, qualitatively described as a *primary quantity* language (Thornes, 2003). This paper thereby serves a secondary purpose of quantifiably characterizing the vowel system of an under-examined language.

B. Endangered language documentation

Language documentation has only recently been widely recognized as both an essential and academically rigorous task, both for formal linguistic practice (Linguistic Society of America, 2010) and for endangered language revitalization (Hinton, 2001). An emerging theory of documentation holds that language data should be collected not only for use by the documenter, but for a range of possible uses (Woodbury, 2003). Indeed, as several of the articles in Grenoble and Furbee (2010) discuss, one of the chief issues in endangered language documentation is ensuring the adequacy of the data collection. Ethical and practical considerations are also critical (see Dwyer, 2006), including other demands on speakers' and researchers' time.

Given a shortage of qualified linguistic field workers (Grinevald, 2003), researcher and speaker time constraints, and the demand for data that meet a wide range of community and researcher needs, it is likely that the kind of detailed and systematic documentation of language sound systems called for by Ladefoged (2003) will frequently be eschewed in favor of broader and more general descriptions. As a result, methods of phonetic analysis must be tailored to the needs of unsystematically collected datasets. This paper addresses this issue in the realm of qualifying *primary quality* versus *primary quantity* language by modifying the SOAM to better address datasets from small speech communities. Results are shown for Numu, for which acoustic measurements are drawn from data collected for other purposes.

C. Numu (Oregon Northern Paiute)

Numu, as it is known to its speakers, is a Uto-Aztecan language of the Western Numic branch. Dialects of Northern Paiute are spoken in scattered communities throughout Oregon, Nevada, and California, but this study is concerned primarily with the language as it is currently spoken on the Confederated Tribes of Warm Springs Reservation in central Oregon. At the time of this study, Numu had fewer than ten speakers in Warm Springs. Overall, dialects of Northern Paiute are spoken by fewer than 500 people and it is considered endangered (Thornes, 2003).

Numu has ten monophthongs comprised of five vowel qualities (i, i̱, u, ɔ, a), each exhibiting a phonemic length contrast. This research and other phonetic documentation by the lead author (Haynes, 2010) contributes a phonetic record of salient features of Numu vowels for future generations of learners and researchers, adding to Waterman's (1911) phonetic description of Oregon dialects of Northern Paiute. Due to limitations in equipment more than a century ago, Waterman was able to provide only a small range of acoustic measurements of Numu. This research expands on his work with the aid of improved technology.

II. NUMERICAL METHOD

A. Motivation

As noted in Sec. I, Wassink (2006) questions the motivation for using ellipses (or ellipsoids) to characterize vowel distributions (at least from the standpoint of acoustics). However, it is possible to motivate the use of ellipsoids as a natural outgrowth of an assumption that the vowel data are random samples taken from a population assumed to be described by a multivariate Gaussian distribution. This assumption is the basis of the SOAM. To the authors' knowledge, the connection between normal distributions in multiple dimensions and elliptical shapes has not been explicitly shown in the literature related to computing vowel overlap. We provide a brief summary (based upon Orechovesky, 1996) of this connection here that may be of benefit to this community and that will help motivate our own approach.

In one dimension, a Gaussian (or normal) distribution can be characterized by its mean and its standard deviation

(Rosner, 2000). In higher dimensions, this concept is generalized to what is called a multivariate Gaussian distribution, defined by its probability density

$$p(\mathbf{x}) = k * \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)\cdot\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}, \quad (1)$$

where $\mathbf{x}$ is a vector of variables, μ is the vector of mean values, k is a normalization factor, and $\boldsymbol{\Sigma}$ is the covariance matrix. The components of a general $n \times n$ covariance matrix can be written as

$$[\boldsymbol{\Sigma}] = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix}, \quad (2)$$

where σ_i is the standard deviation of the variable x_i and ρ_{ij} is the Pearson correlation coefficient (Orechovesky, 1996). A convenient property is that the covariance matrix is symmetric and positive semi-definite, meaning that its eigenvalues are real and non-negative. Moreover, the eigenvectors are mutually orthogonal. Assuming the eigenvalues are distinct, the eigenvectors form a basis (Murdoch, 2012).

A contour of constant probability density can be identified from Eq. (1) by setting the exponent equal to a constant scalar function,

$$(\mathbf{x}-\mu)\cdot\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu) = \chi_n^2(c^2), \quad (3)$$

where χ_n^2 is itself a probability distribution of dimension n and c is the normal score. To illustrate, in the bivariate case, Eq. (3) becomes

$$\frac{1}{1-\rho_{12}^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho_{12}\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right] = \chi_2^2(c^2), \quad (4)$$

which is the equation of an ellipse centered at the mean (Orechovesky, 1996). In higher dimensions, Eq. (3) furnishes an equation for an ellipsoid (or hyper-ellipsoid). The eigenvectors of the covariance matrix are the principal axes, while the eigenvalues are the principal radii. The choice of normal score, c, acts to scale the size of the ellipse. In the SOAM, the linear regressions used to compute rotation angles are essentially finding an approximation to the eigenvectors of the underlying (but not explicitly computed) sample covariance matrix. The data are then rotated into the approximate eigenbasis and principal radii are found by assuming a length of 2σ in each principal direction. If the samples constituted the entire population, a 2σ ellipse (corresponding to the choice of $c = 2$) would represent a 86.47% tolerance interval (i.e., 86.47% of the data are contained within the ellipse) (Rosner, 2000). Of course, the vowel data are assumed to be random samples covering only some small portion of the overall population. Thus, tolerance intervals need to be embedded within a confidence interval (Burrows, 1963; Young, 2010). The analyst must choose a desired confidence and tolerance for a given set of data in order to appropriately scale the ellipse. In addition, the tolerance interval in this case is dependent on the number of samples taken (Burrows, 1963; Young, 2010). In the SOAM, it would be desirable to fix the confidence and tolerance interval to be the same for all considered vowels. This means that vowel ellipses should be scaled differently if they were based on datasets of different size. This issue was not addressed in Wassink (2006) but is an important point.

The question is whether or not it is appropriate to assume vowel data are random samples from a population well described by a multivariate Gaussian distribution—particularly when those data are samples from an endangered language. Variation in data samples are partially random due, for example, to differences in vocal tract sizes and properties of the air (temperature, humidity). There are also significant sources of non-random variation due to both social and consonant context. Researchers have for decades recognized the importance of social context on the pronunciation of words (Labov, 1972). Systematic factors that influence pronunciation include the formality of a speech act, knowledge of other languages (see Chang, 2013), and a complex set of other lexical and perceptual factors (see Pardo *et al.*, 2013). In addition, for endangered languages it may be possible to sample the entire population of speakers if there are a small number. Unless the analyst is lucky, the distribution will almost certainly not be well described by a Gaussian distribution. Thus, basing an analysis of vowels in endangered languages on statistical assumptions is open to significant question.

These reasons motivate the development of an alternative to the SOAM where vowel distributions are not characterized by any assumed statistical foundation, but rather on the geometry of the distribution. Our method is based on fitting convex polygons (or in three dimensions, polyhedrons) to the vowel data. A shape is convex if the line that connects any two points within the shape is itself entirely contained within the shape. An ellipse (or ellipsoid) is an example of a convex shape. In particular, we use what are called *convex hulls* in the field of computational geometry (Preparata and Shamos, 1985; O'Rourke, 1998). Convex hulls are the smallest convex shapes that contain a given set of points. The idea is to minimize the amount of empty space within the shape while still respecting the outer boundary of the dataset. A common analogy in two dimensions is to think of data points as pins on a cork board. The convex hull would be the shape formed by fitting a rubber band around the outer-most pins. In three-dimensions, the convex hull can be thought of as a shrink-wrap fit to data.

Brubaker and Altshuler (1959) first applied convex shapes to vowel data in order to study overlap in the $F1 \times F2$ space. These shapes appear to be convex hulls, but the authors did not formalize this aspect. The authors do not give much detail on the construction of the shapes and the calculation of overlap, but it appears that this was done by hand. They note

"…the boundaries of such data can be drawn in a number of ways to form a variety of configurations. Consequently, plot areas and their overlapping portions can be changed according to an investigator's caprice. Therefore, it is desirable that a single method be consistently applied in forming configurations of point plots (p. 1364)."

Algorithms for computing convex hulls efficiently and in a consistent and standardized way have been developed in the computational geometry community (e.g., O'Rourke, 1998) and included in such commercial numerical packages such as MATLAB (The MathWorks, Inc., Natick, MA), which we employ here. These are easy to implement in both two and three dimensions.

B. Procedure

All numerical results have been generated using a base installation of MATLAB R2011a (no additional toolboxes). In this section, we give an overview of how we generate convex hulls for a given set of vowel data and compute the overlap for long and short vowel pairs. A general algorithm in three-dimensions is given in the Appendix. First, each set of n normalized tokens corresponding to a particular vowel are arranged into an $n \times 2$ ($n \times 3$ in three dimensions) array, with columns corresponding to $F2$ and $F1$ (and duration in three dimensions). Next, for each vowel array, we compute the Delaunay triangulation (Preparata and Shamos, 1985). In a Delaunay triangulation, no vowel token lies within the circumcircle of any triangle. This triangulation is then used to compute the associated convex hull and hull area (volume) for each vowel. While the Delaunay triangulation is not necessary to compute the convex hull in MATLAB (it could be computed directly from a vowel array), it is useful for visualization as well as overlap computation.

Suppose we wish to compute the percentage overlap in two vowels A and B. In the SOAM, Wassink (2006) uses a rectangular grid of test points to approximate the areas of both ellipses (or volumes of ellipsoids in 3D). The number of test points in at least one ellipse (n_A and n_B) as well as points in both ellipses (n_{both}) are tallied. The overlap is taken as

$$\Omega = \max\left[\frac{n_{\text{both}}}{n_A}, \frac{n_{\text{both}}}{n_B}\right]. \tag{5}$$

Morrison (2008) offers an improvement to this procedure by better utilizing the underlying statistical assumptions that we argue against here. Both methods for computing overlap rely on generating a number of test points. In general, an analysis would need to be undertaken to ensure that enough points have been generated such that the computed overlap converges to a particular value. This is a small but important point since, particularly in three dimensions, this may add significantly to the computation time. Another statistical tool that has been used to compute vowel overlap is Pillai's trace (Hay *et al.*, 2006), which is based on computations using covariance matrices.

In the VOACH method, we compute the points in hull A (i.e., the convex hull associated with vowel A) that are contained with hull B and vice versa. These points are used to generate a third convex hull and associated area/volume, V_{OL}. To facilitate comparison with the SOAM, we use a formula for overlap that is similar to Eq. (5),

$$\Omega = \max\left[\frac{V_{\text{OL}}}{V_A}, \frac{V_{\text{OL}}}{V_B}\right]; \tag{6}$$

however, in principle we could use other measures. This measure of overlap has the advantage over that used in the SOAM in that it does not rely on a user-defined test grid. It depends solely on the measured data. Thus, a minimal number of points are used to compute the overlap, resulting in a substantial reduction in computing time.

Figure 1 compares the SOAM and the VOACH method for English (/u, uː/) (i.e., [ʊ], [u]). Data are from men's and women's vowel productions in the vowel database of Hillenbrand *et al.* (1995) and include data from 93 tokens each of the monophthongs [i], [ɪ], [ɑ], [æ], [u], [ʊ]. All vowel spectral data were normalized using Nearey's (1977) formant-extrinsic log-mean normalization method with the Vowels package in R (R Development Core Team, 2009; Kendall and Thomas, 2010). This method was chosen because it has been found to reduce differences in spectral data due to speakers' physiological differences (e.g., longer or shorter vocal tracts), while preserving sociolinguistic differences (Adank *et al.*, 2004; Clopper, 2009). Vowel duration data were normalized using the segmental duration Z-score normalization method described in Wassink (2006) in order to match her procedure as closely as possible. The SOAM figures were generated using a rectangular test grid with 100 evenly spaced points in each direction. Note that (/u uː/) have been selected because they are the only monophthongal short–long vowel pair in this dataset that display overlap in both dimensions.

As we see in the two-dimensional images [Fig. 1(a)], the SOAM provides a very uniform view of the vowels, without accounting for several data points in the higher $F2$ region for /uː/, or for points in the lower $F2$ region for /u/. As a result, the overlap region is calculated as a higher percentage of the whole using the SOAM, at 34.01%, as compared to 32.59% for the VOACH method. In three dimensions [Fig. 1(b)],[1] the uniform SOAM ellipsoids exhibit more overlap than the angular VOACH method convex hulls, and an even larger overcalculation for the overlap results: 4.98% overlap for the SOAM and 1.18% overlap for the VOACH method. While these differences are not substantial for these data, they can have more drastic repercussions for other datasets, as shown in the analysis of Numu vowel data, below.

III. APPLICATION TO NUMU

As we have argued, the VOACH method is particularly applicable for languages with few speakers, and for which datasets are not collected for phonetic analysis. In this section, we present results from applying the method to Numu.

A. Data

Numu data are drawn from a set of recordings that were collected over the course of 1 year for the purpose of creating an audio repository of Numu for the Confederated Tribes of

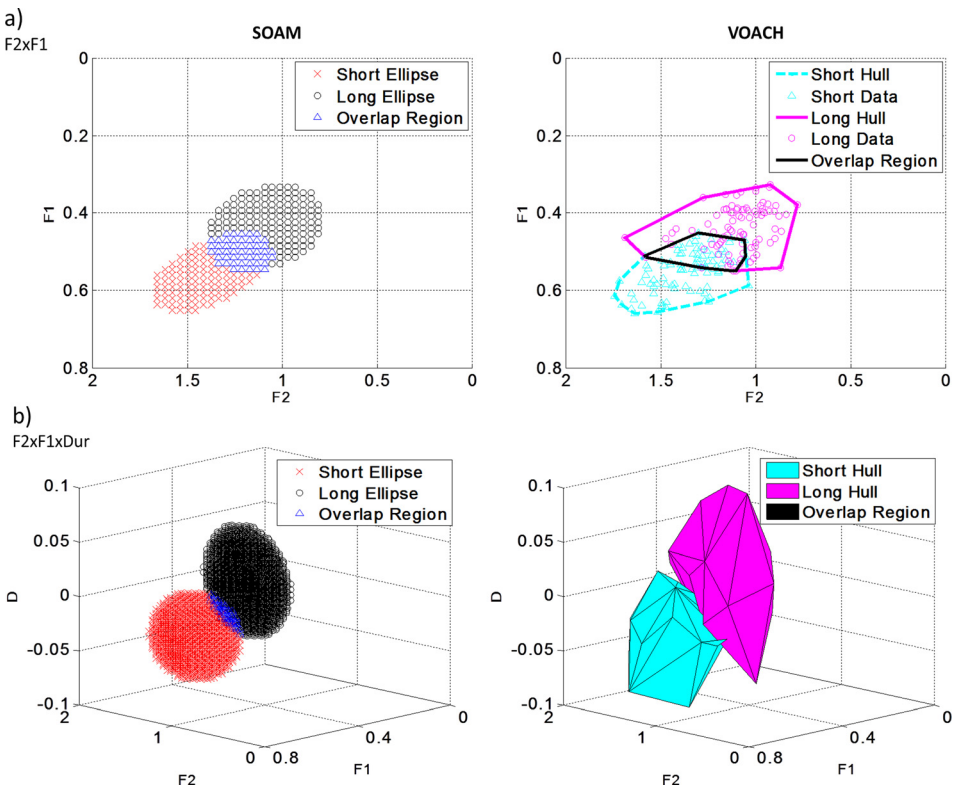

FIG. 1. (Color online) Comparison of (a) two-dimensional and (b) 3D results for the SOAM and the VOACH method applied to American English (/u uː/).

Warm Springs. A total of four fluent speakers were recorded saying Numu words and phrases, including the head and two assistant Numu teachers for Warm Springs Language Program. The head teacher was originally from Burns, Oregon, but lived most of her adult life in Warm Springs. The other three speakers were originally from McDermitt, Nevada (McDermitt is 10 miles from the Oregon border). According to these speakers, all four speak Numu similarly due to their long residence in Warm Springs.

Recordings took place at the Culture and Heritage Department on the Confederated Tribes of Warm Springs Reservation in a quiet basement room (no sound-proof room was available). All four speakers were recorded on either a Marantz (Marantz, Mahwah, NJ) PMD660 solid-state recorder with an AKG C420 (AKG Acoustics, Vienna, Austria) head mounted condenser microphone, or an M-AUDIO (inMusic M-AUDIO, Cumberland, RI) Mobile-Pre USB preamp audio interface with an AKG C520 head mounted condenser microphone. All data were sampled at 44.1 kHz. Because the recordings were collected for other purposes, vowel pairs were not elicited explicitly. Rather, all vowel measurements are taken from stressed consonant-vowel-consonant or consonant-vowel syllables that occurred in isolated words in the dataset (i.e., from lists of isolated words that speakers produced for inclusion in the audio repository).

All four speakers produced similar lists of words, but consonant contexts varied within speaker, such that all vowels occurred before and after consonants at all possible Numu places of articulation (labial, alveolar, velar, and glottal), with one exception: None of the tokens of [a] was followed by a velar consonant. Consonant contexts can influence both vowel spectral and duration data, and as we discuss later, variable consonant contexts are preferred for the VOACH method because of its sensitivity to missing data at a vowel's boundaries.

Each speaker produced 10 to 14 tokens of [i], three to 15 tokens of [iː], 12 to 16 tokens of [u], 8 to 17 tokens of [uː], 12 to 17 tokens of [a], and 11 to 25 tokens of [aː]. The dataset included a total of 49 tokens of [i], 37 tokens of [iː], 56 tokens of [u], 51 tokens of [uː], 61 tokens of [a], and 61 tokens of [aː]. An additional 94 combined tokens of [ɔ, ɔː, ɨ, and ɨː] were also analyzed for use in the qualitative description below. Vowel duration was measured from the start of the first vocalic glottal pulse to the end of the last vocalic glottal pulse, as determined by surrounding consonant constriction release and onset landmarks. First and second formant measurements were taken by hand using linear predictor coefficients spectra from a 25.6-ms window over the vowel midpoint (i.e., at 50% of the vowel's duration). All vowel analyses were carried out in PRAAT (Boersma and Weenink, 2008).

Vowel spectral data were normalized to reduce cross-speaker differences. Because the VOACH method is sensitive to the distribution of peripheral tokens within a vowel category, the unbalanced number of tokens in each vowel distribution could give undue influence to one speaker's system if we chose a vowel-extrinsic method. We therefore selected a vowel-intrinsic, speaker-intrinsic, formant-intrinsic normalization technique (Bark Scale). For vowel duration, we used Wassink's (2006) segmental duration *Z*-score normalization method in order to facilitate comparisons with SOAM results.

B. Qualitative description

We begin with a presentation of the Numu vowel data, along with a traditional, impressionistic analysis of the

TABLE I. Mean formant values in Hertz for short and long Numu vowels.

	*F*1 (Hz)			*F*2 (Hz)	
	n	mean	sd	mean	sd
i	49	399.82	58.58	2363.16	291.44
iː	37	398.86	74.63	2546.32	143.41
ɨ	32	497.59	110.24	1558.63	331.02
ɨː	23	449.57	86.38	1742.00	205.42
u	56	453.57	83.77	1027.48	189.91
uː	51	404.59	59.22	1049.53	286.41
a	61	756.23	83.08	1496.03	174.76
aː	61	843.52	67.63	1445.15	132.37
ɔ	30	625.70	73.71	1113.77	138.48
ɔː	9	626.78	134.37	1065.67	123.55

language's classification as *primary quantity* or *primary quality*. Table I shows the mean values and standard deviations for the first two formants of each long and short Numu vowel (including measurements taken for long and short /ɨ/ and /ɔ/) in Hertz. Based on these values, it appears that Numu is likely a *primary quantity* language. *F*1 value differences for each of the short-long vowel pairs range from 0.96 Hz for /i/ to 87.29 Hz for /a/; for all but the /a, aː/ vowel pair, *F*1 differences are well within one standard deviation. For the /a/ pair, the short vowel is slightly raised in comparison to the long vowel, but still within two standard deviations. *F*2 vowel differences, ranging from 22.05 Hz (/u/) to 183.37 Hz (/ɨ/), are within one standard deviation for every vowel pair.

Figure 2 shows a plot of first and second formant values for the peripheral vowels (/i, u, a/), with shapes representing the individual tokens and the mean represented by a large character. This figure also suggests overlap in the spectral dimensions for each peripheral vowel pair.

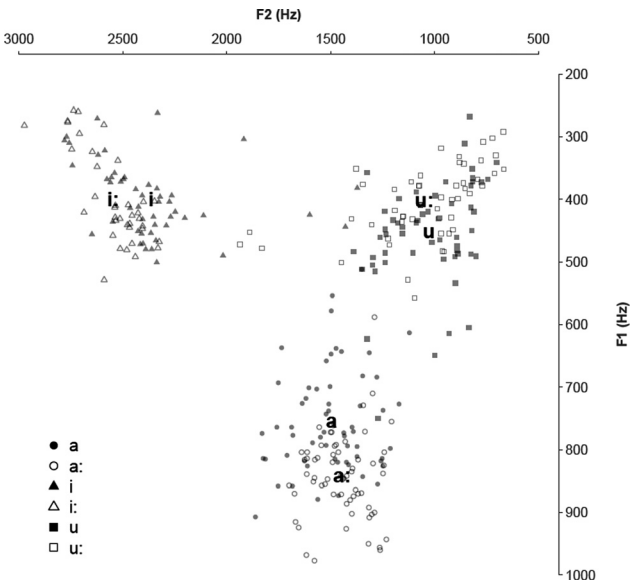

FIG. 2. Scatterplot of *F*2 vs *F*1 values for peripheral Numu vowel tokens (mean values are represented by large characters).

TABLE II. Mean duration in seconds for short and long Numu vowels, and duration ratios.

	Short vowels			Long vowels			Long:Short
	n	duration (s)	sd	*n*	duration (s)	sd	Vowel ratios
i	49	0.139	0.043	37	0.235	0.060	1.69:1
ɨ	32	0.146	0.047	23	0.218	0.046	1.49:1
u	56	0.137	0.034	51	0.240	0.059	1.75:1
a	61	0.136	0.031	61	0.237	0.071	1.74:1
ɔ	30	0.128	0.040	9	0.271	0.084	2.12:1
ALL	228	0.137	0.038	181	0.237	0.064	1.73:1

Mean and standard deviations for Numu short and long vowel durations are presented in Table II. This table also includes the ratio of long to short vowel durations. The average ratio of long to short vowel duration is 1.73:1, which is a shorter ratio than that found in *primary quantity* languages like Japanese and Thai, but longer than that found in *primary quality* languages like English and German (see Crothers, 1978). Impressionistically, the spectral data indicate that spectral distinctions between long and short vowel pairs are unlikely; mean spectral measurements for each of the vowels are very close and there is a great deal of overlap among individual tokens of long and short vowels. However, a more rigorous classification of the data can be obtained through application of the SOAM, and as we show here, a more robust result is obtained through application of the VOACH method.

C. Application of the SOAM

Table III presents two-dimensional and 3D overlap values obtained from applying the SOAM and the VOACH method to the Numu peripheral vowel data,[2] as well as the difference in results obtained from the two methods; in all cases, the SOAM overlap calculation is higher than the VOACH method calculation. The 3D over-calculation is the most drastic, ranging from 17.37 percentage points to 24.06 percentage points higher for the SOAM.[3] The column below each calculation indicates the degree of vowel overlap, based on Wassink (2006, p. 2341–2342)'s classifications: No overlap (0% to 20%); partial overlap (20% to 40%); complete overlap (greater than 40%). The SOAM over-calculations in the two-dimensional model do not change the vowel classifications, all of which are completely overlapping, but the over-calculations in the 3D model are sufficient to change the classification. According to the SOAM, the third dimensional results suggest partial overlap for each vowel pair (i.e., adding duration to the model only partially distinguishes the vowels), while the VOACH method suggests that there is no overlap in the third dimension (i.e., adding duration to the model completely distinguishes the vowels).

The over-calculation in the traditional SOAM is due, in part, to the inclusion of empty data points in its calculation, as discussed for English data above (Fig. 1). That is, points

TABLE III. Two-dimensional and 3D overlap values for the SOAM and the VOACH method applied to Numu peripheral vowels

	*F*2X*F*1 SOAM	*F*2X*F*1 VOACH	Difference	*F*2X*F*1XDur SOAM	*F*2X*F*1XDur VOACH	Difference
i/iː	88.93%	73.01%	15.92	26.68%	3.32%	23.36
OVERLAP:	*full*	*full*		*partial*	*none*	
u/uː	79.94%	67.26%	12.68	25.80%	1.74%	24.06
OVERLAP:	*full*	*full*		*partial*	*none*	
a/aː	76.50%	62.14%	14.36	27.96%	10.59%	17.37
OVERLAP:	*full*	*full*		*partial*	*none*	

from a best fit ellipse are included that would not be included in the fitted shape created by a convex hull. As explained earlier, SOAM ellipses are scaled based on user-defined tolerance and confidence intervals. In the case where two vowels overlap, a larger scaling would result in a higher percentage overlap. Figure 3 compares the two methods in two dimensions and in three dimensions for the Numu (/i, iː/) pair.

In the two-dimensional images [Fig. 3(a)], the SOAM excludes several data points in the higher *F*2 region for /i/, while at the same time filling in empty data points in the higher *F*2 region for /iː/, resulting in a much larger overlap region than is obtained using convex hulls in the VOACH method. In three dimensions [Fig. 3(b)], empty durational data points in the ellipsoids for both vowels result in more overlap than in displayed by the angular VOACH method convex hulls, again resulting in a larger overlap region.

IV. CONCLUDING REMARKS

This paper has presented an update to the SOAM, an important method developed by Wassink (2006) to quantify the differences between *primary quantity* and *primary quality* languages. The main advantage of the method presented here, the VOACH method, is that it is not reliant on an assumption of normally distributed vowel data. It is unlikely that vowel data in general have a Gaussian distribution, due to social and contextual factors. Moreover, the VOACH method, which uses convex hulls to calculate vowel distribution and overlap, reduces empty data in its calculations, thus avoiding over-calculation of overlap. We have compared the SOAM and the VOACH method in application to Numu, an endangered Uto-Aztecan language. Results from the VOACH method are better aligned with impressionistic analyses of the language's spectral and durational data and with observations by field researchers (Thornes, 2003) that vowel pairs are distinguished primarily by duration. Finally,

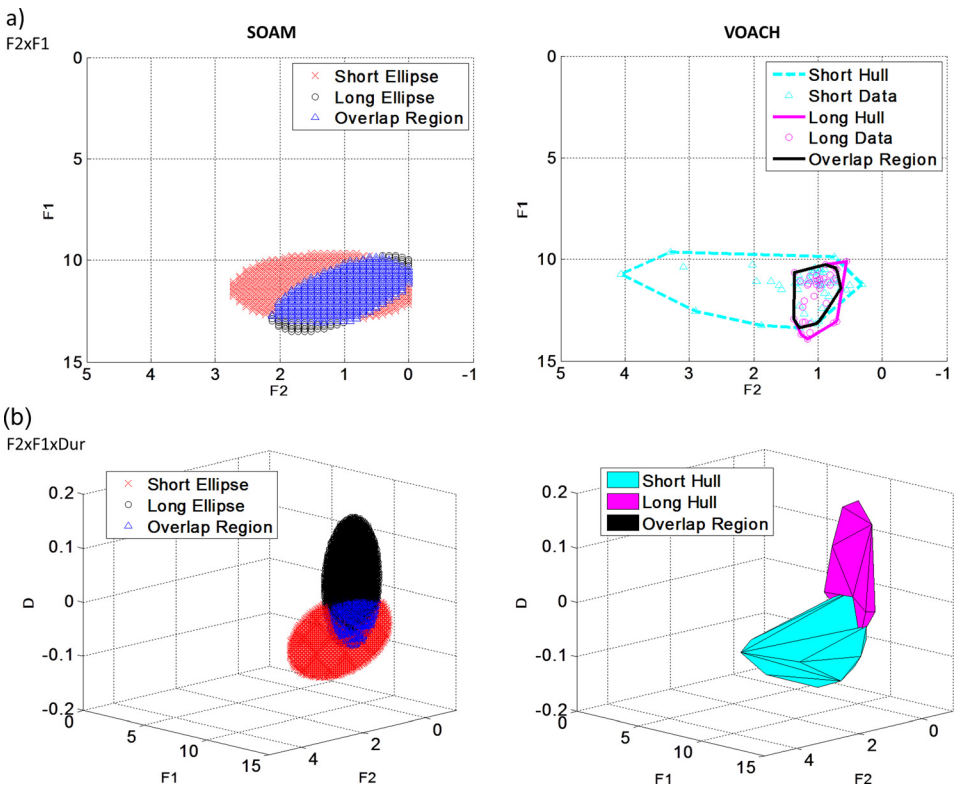

FIG. 3. (Color online) Comparison of (a) two-dimensional and (b) 3D results for the SOAM and the VOACH method applied to Numu (/i iː/).

the VOACH method requires less computing time and does not require the analyst to make decisions regarding the scaling of shapes based on statistical assumptions.

This method has two general limitations that we believe can be resolved through further study. First, as Wassink (2006) points out, the SOAM would benefit from expansion to multiple temporal points, such that spectral analyses are not limited to a single point (or range of points) in the duration of a vowel. This would benefit the VOACH method as well. Second, this analysis has relied on Wassink's (2006) estimates of overlap cutoff ranges, which she derives arbitrarily, such that overlap in less than half but more than a fifth of the volume of the shortest vowel (20% to 40%) is partial overlap, and no overlap or full overlap are protrusions into less or more of the vowel, respectively. These estimates seem reasonable, but would be better supported by cross-linguistic evidence through further study in this area.

The VOACH method also poses limitations for datasets that are very small, or whose distribution does not accurately define a vowel's outer boundaries, because it is sensitive to missing boundary data. In cases where an analyst has data from a small segment of a population, or from a limited number of contexts, and where the analyst can be reasonably sure that the vowel data are normally distributed, the SOAM might be a more appropriate method. In the case of Numu, we have analyzed data from a substantial portion of the population of speakers, and from vowels spoken in a variety of consonant contexts, increasing the range of within-vowel tokens. The VOACH method is therefore more appropriate, and better approximates field researcher analyses. However, if more vowel data become available, it will be important to incorporate them into the analysis to increase descriptive knowledge of the language.

Continued study in this area is merited because having a rigorous, quantitative method like the SOAM for cross-language vowel system comparisons and within language comparisons of vowel contrasts is critical to support field researchers' impressionistic or qualitative analyses of under-documented languages. The VOACH method furnishes an alternative to the SOAM to provide more robust results for under-documented languages. It forms the basis for analyses of other languages, because field researchers studying under-documented languages can collect data for a wide range of analytic uses without necessarily sacrificing precision in their analysis of vowel distinctions.

ACKNOWLEDGMENTS

The authors wish to thank Myra Johnson, former director of the Warm Springs Language Program, as well as Shirley Tufti, Pat Miller, Pearl Stwyer, and Verleen Holliday. Gratitude is also owed to the members of Professor Katia Bertoldi's research group at Harvard University for helpful comments on the analysis. Numu data collection was supported by a grant from the Sven and Astrid Liljeblad Endowment Fund.

APPENDIX: VOACH ALGORITHM

3D convex hull generation procedure and overlap computation in MATLAB 2011a

Require: All normalized vowel tokens in $n \times 3$ arrays (columns are $F2, F1$, D)
  for all vowel tokens **do**
    Compute Delaunay triangulation: DT = DelaunayTri[*vowel*(:,1), *vowel*(:,2), *vowel*(:,3)];
    Compute convex hull and associate volume: [*convexHull*, *volume*] = convexHull(DT);
  end for
  To compare two sets of vowel tokens:
  for all vowel tokens in hull B **do**
    Check to see if token is in hull A: *state* = pointLocation(DT_A, $vowel_B$);
    if *state*! = NaN **then**
      Store token in array *overlap*
    end if
  end for
  for all vowel tokens in hull A **do**
    Check to see if token is in hull B: *state* = pointLocation(DT_B, $vowel_A$);
    if *state*! = NaN **then**
      Store token in array *overlap*
    end if
  end for
  Compute Delaunay triangulation of overlap points:
  DT_{OL} = DelaunayTri[*overlapl*(:,1), *overlap*(:,2), *overlapl*(:,3)];
  Compute convex hull and associate volume: [$convexHull_{OL}$, $volume_{OL}$] = convexHull(DT_{OL}).

[1]The 3D figures have been rotated to best highlight overlap. However, the overlap regions are small and therefore largely obscured.

[2]The SOAM data were generated using a rectangular test grid with 100 evenly spaced points in each direction.

[3]Recall that there are no /a/ tokens that precede a velar consonant in our dataset, which means that we may be missing some high $F2$ values for this vowel. These data could possibly decrease the overlap from the VOACH method even further, as the /a/ $F2$ values are already higher than the /aː/ $F2$ values. This would serve to increase the difference in results obtained for the SOAM and the VOACH method for the /a aː/ pair.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Boersma, P., and Weenink, D. (**2008**). *Praat: Doing Phonetics by Computer*, http://www.praat.org (Last accessed December 1, 2013).

Brubaker, R. S., and Altshuler, M. W. (**1959**). "Vowel overlap as a function of fundamental frequency and dialect," J. Acoust. Soc. Am. **31**(10), 1362–1365.

Burrows, G. L. (**1963**). "Statistical tolerance limits–what are they?," Appl. Stat. **12**, 133–144.

Chang, C. B. (**2013**). "A novelty effect in phonetic drift of the native language," J. Phonetics **41**(6), 520–533.

Clopper, C. G. (**2009**). "Computational methods for normalizing acoustic vowel data for talker difference," Lang. Linguist. Compass **3**(6), 1430–1442.

Crothers, J. (**1978**). "Typology and universals in vowel systems," in *Universals of Human Language*, edited by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford University Press, Palo Alto, CA), pp. 93–152.

Dwyer, A. M. (**2006**). "Ethics and practicalities of cooperative fieldwork and analysis," in *Essentials of Language Documentation*, edited by J. Gippert, N. P. Himmerlman, and U. Mosel (Mouton de Gruyter, New York), pp. 31–66.

Grenoble, L. A., and Furbee, N. L., Eds. (**2010**). *Language Documentation: Practices and Values* (John Benjamins, Philadelphia, PA), 358 p.

Grinevald, C. (**2003**). "Speakers and documentation of endangered languages," in *Language Documentation and Description*, edited by P. K.

Austin (Hans Rausing Endangered Languages Project, London), Vol. 1, pp. 52–72.
Hay, J., Warren, P., and Drager, K. (**2006**). "Factors influencing speech perception in the context of a merger-in-progress," J. Phonetics **34**(4), 458–484.
Haynes, E. F. (**2010**). "Phonetic and phonological acquisition in endangered languages learned by adults: A case study of Numu (Oregon Northern Paiute)," Doctoral dissertation, University of California, Berkeley, CA.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hinton, L. (**2001**). "Audio-video documentation," in *The Green Book of Language Revitalization*, edited by L. Hinton and K. Hale (Academic Press, San Diego, CA), pp. 265–271.
Kendall, T., and Thomas, E. R. (**2010**). "Vowels: Vowel manipulation, normalization, and plotting in R," R package, version 1.1, http://ncslaap.lib.ncsu.edu/tools/norm/ (Last accessed December 1, 2013).
Krauss, M. (**1992**). "The world's languages in crisis," Language **68**(1), 4–10.
Labov, W. (**1972**). "Some principles of linguistic methodology," Lang. Soc. **1**(1), 97–120.
Ladefoged, P. (**2003**). *Phonetic Data Analysis* (Blackwell, Malden, MA), Chap. 1.
Linguistic Society of America (**1994**). "The need for the documentation of linguistic diversity," http://www.linguisticsociety.org/files/lsa-stmt-documentation-linguistic-diversity.pdf (Last accessed December 1, 2013).
Linguistic Society of America (**2010**). "Resolution recognizing the scholarly merit of language documentation," http://www.linguisticsociety.org/resource/resolution-recognizing-scholarly-merit-language-documentation (Last accessed December 1, 2013).
Morrison, G. S. (**2008**). "Comment on 'A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties,' " J. Acoust. Soc. Am. **123**(1), 37–40.
Murdoch, A. I. (**2012**). *Physical Foundations of Continuum Mechanics* (Cambridge University Press, New York), 423 p.
Nearey, T. M. (**1977**). "Phonetic feature systems for vowels," Doctoral dissertation, University of Alberta, Canada.
Orechovesky, J. R., Jr. (**1996**). "Single source error ellipse combination," Master's thesis, Naval Postgraduate School, Monterey, CA.
O'Rourke, J. (**1998**). *Computational Geometry in C*, 2nd ed. (Cambridge University Press, Cambridge), 392 p.
Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., and Lewandowski, E. (**2013**). "Phonetic convergence in shadowed speech: the relation between acoustic and perceptual measures," J. Mem. Lang. **69**(3), 183–195.
Preparata, F. R., and Shamos, M. I. (**1985**). *Computational Geometry: An Introduction* (Springer-Verlag, New York), 390 p.
R Development Core Team (**2009**). "R: A language and environment for statistical computing," Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org (Last accessed December 1, 2013).
Rosner, B. (**2000**). *Fundamentals of Biostatistics*, 5th ed. (Duxbury, Pacific Grove, CA), 792 p.
Thornes, T. (**2003**). "A Northern Paiute grammar with texts," Doctoral dissertation, University of Oregon, Eugene, OR.
Wassink, A. B. (**2006**). "A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties," J. Acoust. Soc. Am. **119**(4), 2334–2350.
Waterman, T. T. (**1911**). *The Phonetic Elements of the Northern Paiute Language* (University Press, Berkeley, CA), Vol. 10, pp. 13–44.
Woodbury, T. (**2003**). "Defining documentary linguistics," in *Language Documentation and Description* (Hans Rausing Endangered Languages Project, London), Vol. 1, pp. 35–51.
Young, D. S. (**2010**). "Tolerance: An R package for estimating tolerance intervals," J. Stat. Software **36**(5), 1–39.