Umeå University

This is a published version of a paper published in *Journal of the Acoustical Society of America*.

Citation for the published paper:

Karlsson, F., Doorn, J. (2012)

"Vowel formant dispersion as a measure of articulation proficiency"

Journal of the Acoustical Society of America, 132(4)

URL: <http://dx.doi.org/10.1121/1.4746025>

Access to the published version may require subscription.

Copyright 2012 Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-57715>

<http://umu.diva-portal.org>

Vowel formant dispersion as a measure of articulation proficiency

Fredrik Karlsson^{a)}

Department of Language Studies, Umeå University, SE-90187 Umeå, Sweden

Jan van Doorn

Department of Clinical Sciences, Division of Speech and Language Pathology, Umeå University, SE-90185 Umeå, Sweden

(Received 14 December 2011; revised 1 July 2012; accepted 23 July 2012)

The articulatory range of a speaker has previously been estimated by the shape formed by first and second formant measurements of produced vowels. In a majority of the currently available metrics, formant frequency measurements are reduced to a single estimate for a condition, which has adverse consequences for subsequent statistical testing. Other metrics provide estimates of size of vowel articulation changes only, and do not provide a method for studying the direction of the change. This paper proposes an alternative approach. Vowel formant frequencies are redefined as vectors originating from a defined center point of the vowel space fixed to a basic three-vowel frame. The Euclidean length of the vectors, the vowel formant dispersion (VFD), can be compared across conditions for evidence of articulatory expansions or reductions across conditions or speaker groups. Further, the angle component of the vowel vectors allows for analyses of direction of the reduction or expansion. Based on the range of investigations afforded by the VFD metric, and simulation experiments that compare its statistical properties with those of other proposed metrics, it is argued that the VFD procedure offers an enhanced view of vowel articulation change over rival metrics.

© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4746025]

PACS number(s): 43.70.Jt, 43.70.Dn [BHS]

Pages: 2633–2641

I. INTRODUCTION

Measurements of vowel formant frequencies are widely used in acoustic studies of articulatory settings in vowel production. The center formant frequencies of the first and second formants (F_1 and F_2) especially have been used to infer tongue position within the range of formant frequencies (and thus within the articulatory range) available for a specific speaker. Being readily available for extraction from acoustic recordings, formant frequency measurements afford studies of vowel articulation in situations and with speaker groups where direct physiological measurements would be infeasible or too invasive.

As F_1 and F_2 center frequencies have been linked with position of the tongue during vowel production, the articulatory range of a specific speaker for a specific condition has been estimated from the corner vowels on an $F_2 - F_1$ plane. The size of the shape enclosed by these corner vowels has been shown to correlate well with intelligibility of the speaker¹ and has predominantly been quantified by its area, the vowel space area (VSA).² VSA estimates, based on either three (VSA3) or four vowel-systems (VSA4), have formed the basis of many studies investigating the reduction of articulation in patient groups due to a specific condition, or the restoration of articulatory range as a result of treatment.^{2–10} The results have, however, been variable in terms of articulatory change shown by the VSA metric, with reductions in VSA at times failing to reach significance in cases

where a consistent trend of articulatory reduction is observed^{11–13} (see Sapir *et al.*⁹ for a review).

In an attempt to address the observed lack of power in the VSA measure, a reorganization of the formant frequency values has been proposed to form a quotient of formant frequency sums into a formant centralization ratio (FCR), so that “the formant frequencies in the numerator are likely to increase, and the formant frequencies in the denominator are likely to decrease with vowel centralization.”⁹ The FCR metric for a three-vowel acoustic space is defined as

$$\text{FCR3} = \frac{F_{2u} + F_{2a} + F_{1i} + F_{1u}}{F_{2i} + F_{1a}} = \frac{1}{\text{VAI3}}, \quad (1)$$

where the F (formant values) in each equation corresponds to the F_1 and F_2 measurements of the vowels /i,a,u/ (subscripts). As indicated in Eq. (1), the FCR is the inverse of an alternative measure, the vowel articulation index (VAI), proposed elsewhere⁸ for the same number of vowels. In the case of the VAI measure, a four-vowel variant, defined as

$$\text{VAI4} = \frac{F_{2i} + F_{2e} + F_{1a} + F_{1u}}{F_{1i} + F_{1u} + F_{2u} + F_{2a}}, \quad (2)$$

has been proposed to include the option of investigating the properties of four-vowel quadrilaterals.⁸ Based on the VAI/FCR metrics in Eqs. (1) and (2), studies have been able to show a significant reduction in VAI (or an increase in FCR) due to hypokinetic dysarthria.^{8,9,14} The VAI/FCR metrics have therefore been argued to provide an enhanced sensitivity over the VSA metric in quantifying vowel articulation change in dysarthric patients.^{8–10,14}

^{a)}Author to whom correspondence should be addressed. Electronic mail: fredrik.karlsson@ling.umu.se

Apart from the literature concerned with the effect of dysarthria, alternative approaches for estimating the size of the active articulatory range in speakers focusing on other aspects of vowel space size have been investigated. Lane *et al.*¹⁵ proposed an average vowel spacing (AVS) metric, defined as

$$AVS = \frac{2}{n(n-1)} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sqrt{(F_{1i} - F_{1j})^2 + (F_{2i} - F_{2j})^2} \right\}, \quad (3)$$

to form the pairwise Euclidean distances between F_1 and F_2 frequencies of corner vowels (indexed by i) which is subsequently averaged to form a single separation quantity for the full set of n vowels. The AVS is predicted to increase in an expanding vowel space articulation and to decrease in reduced articulation range. A metric similar to the AVS was also shown by Neel *et al.*¹⁶ to be more powerful than VSA in the prediction of speech intelligibility, affirming that distance-based metrics may offer advantages over those estimating overall acoustic area.

In a parallel development, an alternative approach focused on a reformulation of formant frequency values from being points in a vowel space to distances from a vowel space center. Bradlow *et al.*¹ placed a center of the $F_2 - F_1$ plane at the point (F_{2c}, F_{1c}) , where $F_{1c} = (1/n) \sum_{i=1}^n F_{1i}$ and $F_{2c} = (1/n) \sum_{i=1}^n F_{2i}$ for a set of n vowel productions, and found that the average distance from this center point to a point (F_{2i}, F_{1i}) for each vowel correlated well with the degree of intelligibility in normal speakers. A similar method has also been shown to be a useful basis for vowel space normalization across speakers in other fields.^{17,18} The approach used by Bradlow *et al.*¹ is seen as an advancement over rival approaches as it is intuitively related to a vowel space center point, and therefore agrees well with vowel centralization processes frequently observed in highly time-constrained articulation. It further considers the direction of the reduction in vowel articulation in the procedure, which provides a basic first step toward interpretation in terms of the articulatory structure that is affected by a decreased articulatory range.

While the metrics discussed so far have all been used recently in comparisons of articulatory range of speakers, they raise, however, two issues of statistical importance that have not yet received attention in the literature. The first issue concerns the overall nature of the metrics, and applies to the most frequently used VSA and VAI/FCR metrics. The metrics are calculated based on formant frequency measurements of corner vowels, each consisting of a single (F_2, F_1) coordinate tuple constructed from the within-vowel average. Thus, while the accuracy of the average corner vowel estimate may be strengthened by multiple measurements of phonetically comparable vowels, the result is in all cases a single (F_2, F_1) coordinate tuple for each corner. All information concerning the confidence interval of this mean estimate is discarded. Due to this issue, metrics similar to the AVS metric^{15,16} should be considered an advancement since formant measurements of individual productions can be directly included in the calculations of inter-vowel dispersions, adding to the stability of the measure.

The second issue of statistical importance concerns how the metrics are prepared for inclusion into statistical testing of group differences or developmental trends, and applies to all metrics discussed so far. As a part of the calculations, formant frequency values are condensed into a single value for each set of vowels, resulting in a substantial loss of information concerning the measurements on which the resulting metric is based. Further, as only a single value is returned from the calculation, determination of within-condition variation is not supported without a resampling procedure. This severely reduces the possibility of within-subject modeling and reliable repeated measures testing, and is therefore considered to be a strong negative factor in the usefulness of the metrics.

An additional issue of statistical importance relates specifically to the VAI/FCR metrics and concerns the reliability of results of statistical testing where these metrics are used as covariates. The VAI/FCR metrics were explicitly designed to be more sensitive to finding statistically significant reductions in vowel articulation range than the VSA.⁹ This design goal may unduly increase the risk of a Type I error of analysis by simply increasing the chance of finding significant differences, but this has not been evaluated in any report known to us.

Besides the statistical issues raised above, there is a final point related to the interpretation of any experimental changes found for the metrics discussed so far. For all metrics, a change in the resulting value is proposed to indicate an underlying reduction or expansion of overall vowel articulation range. An analysis into which part of the vowel space is affected is, however, not afforded by any of the metrics. As such, even in research designs where the statistical deficiencies are remedied by controlled resampling, the metrics provide a limited view regarding the nature of the indicated change in articulation.

In light of the reviewed issues with the more frequently used metrics, a considerable advancement is seen in the approach used by Neumeyer *et al.*¹⁹ in the estimation of vowel production range in children with cochlear implants. In their study, Neumeyer *et al.* used the center point formed by the average frequencies of F_1 and F_2 , from which the Euclidean distance of each vowel was calculated. However, in contrast to other metrics,^{1,3} Neumeyer *et al.*¹⁹ used formants from all productions of each corner vowel, and not their average, to create a full set of distances that were used as the basis for comparisons between groups. This approach offers significant advantages over previously proposed metrics as it overtly includes the full set of corner vowels, and thereby also an estimate of the within-vowel variability, into the statistical comparison of groups.

The selection of the centroid as a point of origin for each vowel, however, has a significant disadvantage when it comes to deducing the direction of a changed articulatory range within the $F_2 - F_1$ space. Any directional systematic change in articulation also affects the position of the center point, which in turn propagates a change in vowel distances to other vowels as well. This point was raised by Turner *et al.*³ previously in connection to their metric. Further, the positioning of the centroid is dependent on the number of corner vowels included. As a result, the distances of a specific set of vowel productions within a space enclosed by

four corner vowels would likely differ substantially from distances computed for the same vowel productions, but positioned within a system enclosed by three corner vowels. In light of the varying number of corner vowels used in previous studies of speakers with reduced articulatory range,^{2-8,10,14} this instability in the placement of the center point offers a significant reduction in reliability of the metric. This lack of reliability will affect the study of significant articulatory impairment, as well as severely reduce comparability across studies.

A. Vowel formant dispersion

Against this background, a new metric of vowel articulation is proposed, vowel formant dispersion (VFD). In the VFD view of vowel space, vowels are considered as vectors originating from a center point $(F_{2_{wm}}, F_{1_m})$ that is calculated as a weighted midpoint in a stepwise procedure. In the first step, the coordinate of the midpoint along the vertical axis is calculated according to

$$F_{1_m} = \frac{1}{n} \sum_{i=1}^n F_{1_i} \quad (4)$$

for all vowels $V(F_2, F_1)$. This is consistent with previously proposed procedures.^{1,3,19}

In contrast, the position of the vowel space center placement along the F_2 axis is defined as a weighted midpoint

$$F_{2_{wm}} = \frac{1}{n} \sum_{i=1}^n F_{2_i} \{V(F_2, F_1) : F_1 < F_{1_m}\}, \quad (5)$$

and is therefore based only on vowels that are more closed than a vowel positioned at the vowel space midpoint along the F_1 axis, i.e., all vowels $V(F_2, F_1)$ with an F_1 value less than the value of the F_1 midpoint. This approach is similar to vowel space normalization strategies used previously,^{20,21} and accomplishes two objectives. First, the vowel space center position is kept constant across systems containing three and four investigated vowels, which facilitates comparison of results from three- and four-vowel studies. Further, three-vowel systems have been proposed to approximate a minimal set of vowels across languages,²² with systems similar to (/i/, /a/, /u/) being proposed as an approximate base set, suggesting that the method could be used in most contexts. Thus, calculations of the vowel space center point are based on the subset of measurements that forms the vowel triangle which is available in most studies. This provides an increased level of comparability across conditions over previously proposed approaches in which any vowel articulation change affects the placement of the center point, and therefore also the distances of all vowels.

Second, the placement of the center point is made more robust to missing productions of a suitable front, open vowel. Additional levels of robustness are further afforded by inclusion of vowels that are not at the edge of the vowel space across conditions in the calculations of vowel space center. These vowels are argued to be less affected by

changes in vowel articulation proficiency, and therefore add robustness to the center point placement. This leads to increased robustness of the system as a whole when it is included in all conditions compared.

Subsequent to the establishment of a vowel space center point $(F_{2_{wm}}, F_{1_m})$, each vowel is conceived as a vector originating at this point. An illustration of the switch from perceiving vowels as points on a $F_2 - F_1$ plane to representing them as vectors is provided in Fig. 1 using vowel formant measurements from Peterson and Barney.²³

Each vowel vector V_i is defined to have a length according to

$$VFD_i = \|\vec{V}_i\| = \sqrt{(F_{1_i} - F_{1_m})^2 + (F_{2_i} - F_{2_{wm}})^2}, \quad (6)$$

indicating the dispersion of vowel formants (F) from the vowel space center, and referred to as the VFD. Similar to the method used by Neumeier *et al.*,¹⁹ whole-system expansions or reductions of the vowel spaces can be studied by statistical testing of differences between the set of VFD values in each contrasted condition, and possibly within each vowel category. The VFD values are proposed to increase in expanded vowel spaces and decrease in vowel spaces with reduced articulation range. Similarly, as the complete set of measured vowels are included in the calculations, a substantial increase in statistical power is argued to be achieved over methods that use VSA, VAI/FCR, and AVS metrics that result in a single value for each speaker and condition.

The VFD view is argued to offer a significant advantage over the method of Neumeier *et al.*¹⁹ due to the increased robustness in the placement of the center point by the angle of a vowel vector against the F_2 axis in the $F_2 - F_1$ vowel plot. The vowel angle ω_i is defined as

$$\omega_i = \begin{cases} \tan^{-1} \left(\frac{F_{1_i} - F_{1_m}}{F_{2_i} - F_{2_{wm}}} \right) & \text{for } F_{2_i} > F_{2_{wm}} \\ \tan^{-1} \left(\frac{F_{1_i} - F_{1_m}}{F_{2_i} - F_{2_{wm}}} \right) + \pi & \text{for } F_{2_i} < F_{2_{wm}} \text{ and } F_{1_i} \geq F_{1_m} \\ \tan^{-1} \left(\frac{F_{1_i} - F_{1_m}}{F_{2_i} - F_{2_{wm}}} \right) - \pi & \text{for } F_{2_i} < F_{2_{wm}} \text{ and } F_{1_i} < F_{1_m} \\ \frac{\pi}{2} & \text{for } F_{2_i} = F_{2_{wm}} \text{ and } F_{1_i} > F_{1_m} \\ -\frac{\pi}{2} & \text{for } F_{2_i} = F_{2_{wm}} \text{ and } F_{1_i} < F_{1_m} \\ 0 & \text{for } F_{2_i} = F_{2_{wm}} \text{ and } F_{1_i} = F_{1_m} \end{cases} \quad (7)$$

in a $-\pi < \omega_i < \pi$ interval and may be obtained by calculating $\text{atan2}(F_{1_i} - F_{1_m}, F_{2_i} - F_{2_{wm}})$ in many computer languages and software packages. The angle ω_i defines in which direction a given vowel is located compared to the vowel center point. It may therefore be used to categorize the vowel in terms of articulatory properties by the quadrant of a unit circle in which the vowel is found for the purpose of systematic evaluation of vowel articulation properties. Back, closed, and half-closed vowels are found in the $-\pi < \omega_i < -\pi/2$ range; front, closed, and half-closed vowels are found in the $-\pi/2 < \omega_i < 0$ range; front, half-open, and open vowels are found in the $0 < \omega_i < \pi/2$ range; and back, half-open, and open vowels are found in the $\pi/2 < \omega_i < \pi$ range.

FIG. 1. Illustration of the shift of view of vowels from points (top) to vectors originating at a vowel space center point (bottom). Vowel formant measurements were obtained from Peterson and Barney (Ref. 23).

As shown in Fig. 2, the computed vector angles remain robust to changes in vowel space size as they are comparable within most vowel categories across male, female, and child speakers in the Peterson and Barney²³ data set. One caveat that should be noted, however, is that back, half closed, or half-open vowels may receive a full $-\pi < \omega_i < \pi$ angle

range due to individual cross overs from the $-\pi < \omega_i < -\pi/2$ quadrant to the $\pi/2 < \omega_i < \pi$ quadrant, and should not be considered robustly categorized by the angle proposed here. For vowels not positioned close to this crossover point, however, the data presented in Fig. 2 suggest that the vowel center positioning scheme results in vowel angles that are transferable to a full range of vowel space sizes.

Combined with the VFD values resulting from Eq. (6), the computed vowel vector angles afford detailed analysis of both size and direction of vowel space expansion or centralization. Figure 3 presents box plots of VFD values for vowels in the Peterson and Barney²³ data set divided by the angle component range, and with the phonetic target corresponding to the corner vowel of the angle range indicated for the sake of clarity. This type of display affords specific interpretations concerning the nature of observed changes in vowel articulation. In within-speaker investigations, reductions in proficiency of reaching articulatory targets in a certain direction, or as a result of a reduction involving a certain articulator (e.g., the mandible), are observed in the VFD values in the corresponding quadrant. Further, a quantification of articulatory precision is made available through descriptive statistics of variance within the quadrant in a controlled research design.

It can be argued that the VFD and vowel angle offer an enhanced view of vowel production properties over previously proposed metrics. In a controlled setting, it is comparable to the metric proposed by Neumeyer *et al.*¹⁹ However, since the vowel space center remains constant across conditions, it adds an additional opportunity for separate study of direction of change that is not afforded by the method of Neumeyer *et al.* due to the limitations of the placement of

FIG. 2. Vowel angles obtained from the data set of Peterson and Barney (Ref. 23). Within each vowel, the group of box plots show angles obtained from children (left), males (middle), and women (right).

FIG. 3. Box plots of VFD values categorized by vowel space corners. The corner categories are defined by the angle component of the vowel vector: $-\pi < [u] < -\pi/2$, $-\pi/2 < [i] < 0$, $0 < [ae] < \pi/2$, and $\pi/2 < [a] < \pi$.

the vowel space center.³ VFD is argued to be superior to VSA, VAI/FCR, and AVS metrics for the study of changes in vowel articulation range as it retains more information that would be of benefit in construction of a solid statistical model.

Conversely, it could also be argued that the VSA, VAI/FCR, and AVS offer an enhancement over VFD in that the process of averaging and reducing measurements into a single value may help prevent within-vowel variability and measurement errors that would otherwise increase the likelihood of a Type II error of analysis. In order to resolve these opposing views, two simulation experiments were conducted. Experiment I aimed at investigating the level of bias in the metrics while Experiment II investigated the power of each metric as a method for reliably finding reductions in produced vowel spaces.

II. EXPERIMENT I: POTENTIAL STATISTICAL BIAS IN THE PROPOSED METRICS

The aim of this experiment was to investigate whether a statistical bias exists in the proposed metrics. The relative frequency with which group differences would be found where none actually existed was investigated using a procedure of random sampling of speakers from a homogenous data set. It was expected that an unbiased metric would find differences between two random subsets of a homogenous group in 5% of investigations.

A. Material

The speech material used in the simulation consisted of the vowel formant frequency values obtained from 50 male speakers of Dutch²⁴ made available in the Praat software package.²⁵ This data set was selected because it constitutes the largest publicly available data set of comparable speakers, even though it contains measurements of only one production per vowel. F_1 and F_2 formant measures for the four

vowels that constitute vowel space corners for each speaker were extracted to form the basis of the experiment.

B. Procedure

In the first step of the procedure, the size of the measurement set obtained for each vowel was varied in a resampling procedure. For each vowel and vowel formant, 2...8 new vowel measurements were simulated by sampling from a normal distribution centered at the estimate given in Pols *et al.*²⁴ for the vowel formant in question, and with the standard deviation of that vowel formant across all speakers in the original data.²⁴ This initial step of the procedure simulated the scenario of an increasing number of comparable vowel measurements being available during statistical testing.

As a result of the initial resampling procedure, new data sets consisting of 1...8 estimates of each corner vowel and speaker were provided. For each new data set, the vowel space metrics VSA3, VSA4, VAI3, VAI4, FCR3, and AVS were estimated on the mean corner vowels of the newly constructed vowel sets, simulating the situation of each metric being calculated by averaging an increasing number of productions of each corner vowel. In addition, VFD values were computed within each speaker, based on the F_1 and F_2 formant frequencies of each vowel production in the data set.

In the second step of the procedure, the generated data sets were repeatedly divided into two sets of 25 speakers and compared. In each iteration, 25 speakers were selected at random to form set *A*, and the remaining speakers were considered as set *B*. The two sets were then compared in terms of within-speaker VSA3, VSA4, VAI3, VAI4, FCR3, AVS, and VFD metrics by testing group differences with a student's *t*-test. The second step in the procedure was repeated 10 000 times, giving results in terms of probabilities that the samples came from the same populations (the *p*-value of the *t*-test) for each metric and iteration. Identical data sets formed the basis for the calculations for each metric.

C. Results

The relative frequency of tests showing a significant difference at the $p < 0.05$ level between the two tested sets (A and B) for each proposed metric is shown in Fig. 4. The $p < 0.05$ level was chosen as it constitutes an often used cutoff point where a significant difference would be reported. Test results are divided into separate panels by the number of vowel productions generated in step one of the simulation procedure, in groups of two.

As seen in Fig. 4, all proposed metrics were shown to be biased when one or two productions per vowel were included in the calculations preceding the statistical testing. All metrics showed a worse than expected ability to find group differences. At $3 \leq n \leq 4$ productions per vowel, all metrics except VSA4 and AVS approximate an unbiased testing in that differences between subgroups were found in relatively close to 5% of the 10 000 trials. When larger sets of productions ($n \geq 5$) of each vowel were used to provide the mean estimate for the corner vowels, all metrics except VFD showed a considerable bias toward finding differences between the two randomly assigned sets of speakers of identical origin.

D. Discussion: Experiment I

Experiment I aimed at investigating whether a statistical bias existed in the proposed metrics. The experiment was conducted using a procedure of random sampling of speakers from an assumed homogenous data set,²⁴ which was then

compared through statistical testing based on the proposed metrics of vowel space articulation.

The expected result of an unbiased metric is that statistical testing should show statistical within-group differences at the 0.05 level in 5% of investigated cases. This expectation was, however, not borne out for all investigated metrics, and further, the simulation indicated that outcomes may be dependent on the number of productions included in the calculations. With less than three productions of each vowel, all metrics found group differences much less often than the assumed 5% of cases. At higher numbers of included productions, only the VFD metric remained close to the 5% level, while other metrics found differences to a considerably higher degree than what would be assumed for an unbiased metric.

The results of the simulation therefore indicated that, with less than three productions of each corner vowel included in the calculations, all investigated metrics of vowel space expansion found differences between groups where it could be assumed that there was no difference less often than expected. VFD stabilized at around 5% of tests at $n \geq 3$ productions and performed in a reasonably predictive manner throughout the investigated range of productions of each vowel. The outcome of testing using the FCR3, VAI3, VAI4, and VSA3 appear to behave as expected with three or four included productions (five or six for AVS). Outside of these ranges, all previously proposed metrics appear to be approximately three times as likely to produce either a Type

FIG. 4. The relative frequency (in percent) for which statistical testing of a homogenous data set divided at random into two sets would indicate a significant difference between the sets at the 0.05 level. The metric on which the testing was conducted is indicated on the horizontal axis. The 5% level expected in an unbiased testing is indicated by a dashed horizontal line. A dotted white line indicates a 1% level for the purpose of visual reference.

I or a Type II error of analysis compared to the generally assumed 5% of cases.

III. EXPERIMENT II: PERFORMANCE OF THE METRIC IN FINDING EXISTING GROUP DIFFERENCES

The aim of this experiment was to investigate the statistical power of each metric in finding an artificially constructed reduction in vowel space. It was expected that a reliable metric would give stable probabilities for statistical differences between sets given a controlled reduction of the actual size of the vowel space, while being able to detect the change earlier than rival metrics.

A. Material

The data set used in this experiment was the same as in Experiment I.

B. Procedure

This experiment was conducted in three steps, each one performed 1000 times. The first step was identical to that of Experiment I. As a result, eight data sets consisting of 50 speakers and from one to eight productions of each corner vowel were generated.

In a second step, the centroid of the vowel space, defined in accordance with some previous studies,^{3,19} was established as the within-speaker mean of F_1 and F_2 formant frequencies for vowels produced. The distance of each vowel to this vowel space center point was subsequently calculated. The selection of the centroid as a point toward which centralization would occur was made due to its intuitive nature and in order to not specifically bias any subsequent comparisons toward the VFD metric due to similarities in procedure.

In the final step, a reduction in articulatory range was simulated by moving each vowel closer to the center of the vowel space in steps of 1% of the original vowel distance. After each applied reduction of vowel space, statistical testing of each metric (FCR3, VAI3, VAI4, VSA3, VSA4, AVS, and VFD) between the original productions and the artificially reduced set was performed using one-tailed t -tests.

C. Results

The results of the simulation tests showing when a reduction of produced vowel space size would be detected is presented in Fig. 5. The results are presented in terms of box plots of p -values obtained by statistical testing of differences between the original vowel space and the reduced vowel space. Each metric is displayed separately, as well as each range of the number of produced vowels included in the calculations. The same vowel spaces were compared across all metrics.

D. Discussion: Experiment II

This experiment aimed at investigating the statistical properties of vowel space metrics when establishing a vowel space reduction. Three observations can be made from the results of the simulations (Fig. 5). First, all previously pro-

posed metrics showed a substantial variability in p -values obtained for a certain controlled reduction of the vowel space. This is indicated by the size of the boxes as well as the presence of outliers. The outcome of statistical testing to find differences when the vowel space is reduced using these metrics appears to be highly dependent on the shape of the vowel space tested. For the VFD metric, the boxes are substantially smaller and no outliers are marked. This metric therefore appears to be more stable than rival metrics, although it should be observed that this stability may cause a later finding of significant differences between sets (comparable to VSA3) if a small number of productions of each vowel is included in the calculations.

Second, the VSA3, VSA4, VAI3, VAI4, and FCR3 metrics generally did not benefit from the inclusion of more productions of each vowel when it came to stability of the p -value. A reduction in the number of outliers and box sizes may be observed for $n > 2$, but remained comparable across all the other ranges of included productions, with the exception of a significant difference being found somewhat earlier (i.e., after a less extensive induced reduction of the vowel space) by the FCR3 metric when more productions were included. The VFD metric, in contrast, benefited greatly from the inclusion of more productions into the calculations, achieving a progressive increase in p -value stability and earlier detection of vowel space reduction. AVS received some benefit similar to that of VFD by the inclusion of more measurements, but remained conservative throughout the conditions examined here.

Third, it may be observed from Fig. 5 that the VSA3 and AVS metrics appear to be of weakest power when it comes to finding differences between the two data sets.

E. Conclusions: Experiment II

The simulation experiment indicated that the VFD metric was more stable than rival metrics and benefited most from the inclusion of more productions of each vowel. The metric was, however, highly conservative when it came to finding differences when $n < 3$ productions of each corner vowel were included in the calculations. The previously proposed metrics showed a large variability in the results and did not benefit substantially from the addition of extra vowel productions into the calculations. FCR3 is the metric which was most prone to finding differences when a low number of vowel productions formed the basis of the comparison. VSA3 and AVS appeared to be the least statistical powerful metrics investigated.

IV. DISCUSSION

This paper provides a description of a new metric, the VFD, for estimating the size and direction of vowel articulation change. Vowel formant tuples are redefined as vectors originating from a vowel space center point, the position of which is defined to reduce variability of its placement and thereby increase comparability across vowel systems. The stability of the vowel center point placement further affords the interpretation of the vectors in terms of direction of change of articulation. Two simulation experiments were conducted to

FIG. 5. Obtained probabilities of no difference being detected in statistical testing (one tailed t -test) between the full vowel space (Ref. 24) and a reduced version of the same vowel space, categorized by metric (right side) and the range of number of productions of each vowel included in the calculations (top). The scaling factor (percent of original vowel distance from the vowel space centroid) applied to compute the reduced vowel space is shown on the x -axis, and the probability of the two sets being indicated to be the same (the p -value) is shown on the y -axis. The 5% alpha level where a statistically significant difference would predominantly be reported is indicated by a dashed horizontal line.

investigate the statistical properties of the proposed metric, and compare it with previously proposed metrics.

The results of these experiments suggest that the VFD metric is less statistically biased than previously proposed metrics, once an appropriate sample size of vowel productions has been obtained. Rival metrics performed differently depending on sample size, and were shown to be unstable in

terms of bias at opposite ends of the sample size range investigated. Further, the VFD metric was shown to be the most stable in prediction of vowel space reductions, and the metric that benefited most from the inclusion of more vowel measurements into the calculations.

The FCR3 was confirmed to be the most powerful metric in potentially finding differences between vowel spaces

of different sizes, but showed a substantial variation in probability estimates across all sample sizes. Thus, evidence is provided for FCR3 being powerful in finding differences, but not reliably so for all sizes of reductions and with an added potential bias toward Type I error for larger sample sizes. VSA3 and AVS appear to be the least statistically powerful metrics investigated.

V. CONCLUSIONS

The VFD metric outlined here provides an enhanced view of articulation changes compared to rival metrics based on vowel formant frequencies. It is a phonetically intuitive metric which affords easy transferal to articulatory interpretations, and was shown in simulations to be more efficient in terms of maintaining a level of statistical power consistent with an expanded data collection effort. The VFD view further introduces the possibility of constructing within-speaker models of articulation changes not afforded by previously proposed metrics. VFD is not specifically designed for finding differences between conditions, and was shown to be less likely to increase the risk of Type I error of analysis. It is therefore argued that the VFD method outlined here is superior to previously proposed metrics for the study of vowel production based on vowel formant frequency measurements.

ACKNOWLEDGMENTS

This work was carried out as part of a research fellowship awarded to F.K. by the Swedish Research Council (Grant No. 2009-946). The valuable input on the design of Experiment I provided by Maria Karlsson at the Department of Statistics, Umeå University is gratefully acknowledged.

- ¹A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* **20**, 255–272 (1996).
- ²W. Ziegler and D. von Cramon, "Vowel distortion in traumatic dysarthria: A formant study," *Phonetica* **40**, 63–78 (1983).
- ³G. S. Turner, K. Tjaden, and G. Weismer, "The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis," *J. Speech Hear. Res.* **38**, 1001–1013 (1995).
- ⁴G. Weismer, J. S. Laures, J.-Y. Jeng, R. D. Kent, and J. F. Kent, "Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis," *Folia Phoniatr. Logop.* **52**, 201–219 (2000).
- ⁵P. A. McRae and K. Tjaden, "Acoustic and perceptual consequences of articulatory rate change in Parkinson disease," *J. Speech Lang. Hear. Res.* **45**, 35–50 (2002).

- ⁶K. Tjaden and G. E. Wilding, "Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings," *J. Speech Lang. Hear. Res.* **47**, 766–783 (2004).
- ⁷H.-M. Liu, F.-M. Tsao, and P. K. Kuhl, "The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy," *J. Acoust. Soc. Am.* **117**, 3879–3889 (2005).
- ⁸N. Roy, S. L. Nissen, C. Dromey, and S. Sapir, "Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy," *J. Commun. Disord.* **42**, 124–135 (2009).
- ⁹S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech," *J. Speech Lang. Hear. Res.* **53**, 114–125 (2010).
- ¹⁰S. Skodda, W. Grönheit, and U. Schlegel, "Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease" *PLoS ONE* **7**, e32132 (2012).
- ¹¹K. Bunton and M. Leddy, "An evaluation of articulatory working space area in vowel production of adults with Down syndrome," *Clin. Linguist. Phonetics* **25**, 321–334 (2011).
- ¹²S. Sapir, J. L. Spielman, L. O. Ramig, B. H. Story, and C. Fox, "Effects of intensive voice treatment (the Lee Silverman voice treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings," *J. Speech Lang. Hear. Res.* **50**, 899–912 (2007).
- ¹³G. Weismer, J.-Y. Jeng, J. S. Laures, R. D. Kent, and J. F. Kent, "Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders," *Folia Phoniatr. Logop.* **53**, 1–18 (2001).
- ¹⁴S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in Parkinson's disease," *J. Voice* **25**, 467–472 (2011).
- ¹⁵H. Lane, M. L. Matthies, J. S. Perkell, J. Vick, and M. Zandipour, "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation," *J. Speech Lang. Hear. Res.* **44**, 552–563 (2001).
- ¹⁶A. T. Neel, "Vowel space characteristics and vowel identification accuracy," *J. Speech Lang. Hear. Res.* **51**, 574–585 (2008).
- ¹⁷B. M. Lobanov, "Classification of Russian vowels spoken by different speakers," *J. Acoust. Soc. Am.* **49**, 606–608 (1971).
- ¹⁸D. Recasens and A. Espinosa, "Dispersion and variability in Catalan five and six peripheral vowel systems," *Speech Commun.* **51**, 240–258 (2009).
- ¹⁹V. Neumeyer, J. Harrington, and C. Draxler, "An acoustic analysis of the vowel space in young and old cochlear-implant speakers," *Clin. Linguist. Phonetics* **24**, 734–741 (2010).
- ²⁰D. 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* **9**, 159–173 (2002).
- ²¹N. Flynn and P. Foulkes, "Comparing vowel formant normalization methods," in *Proceedings of the 17th ICPHS*, Hong Kong (2011), pp. 683–686.
- ²²R. Lass, "Vowel system universals and typology: Prologue to theory," in *Phonology Yearbook*, (Cambridge University Press, Cambridge, 1984), pp. 75–111.
- ²³G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184 (1952).
- ²⁴L. Pols, H. Tromp, and R. Plomp, "Frequency analysis of Dutch vowels from 50 male speakers," *J. Acoust. Soc. Am.* **53**, 1093–1101 (1973).
- ²⁵P. Boersma and D. Weenink, "Praat, A system for doing phonetics by computer," *Glott Int.* **5**, 341–345 (2001).