

Comment on “A geometric representation of spectral and temporal vowel features:
Quantification of vowel overlap in three linguistic varieties” [J. Acoust Soc. Am. 119,
2334–2350 (2006)] (L)

Geoffrey Stewart Morrison^{a)}

Department of Cognitive & Neural Systems, Boston University, 677 Beacon Street, Boston,
Massachusetts 02215

upload date 1 October 2007

Prepublication version of letter to appear (2008) *Journal of the Acoustical Society of America*
123 (1) including footnotes removed from published version because of space constraints.

^{a)}Now at School of Language Studies, Australian National University, Canberra, Australian Capital
Territory 0200, Australia. Electronic mail: geoff.morrison@anu.edu.au

ABSTRACT

In a recent paper by Wassink [J. Acoust Soc. Am. 119, 2334–2350 (2006)] the spectral overlap assessment metric (SOAM) was proposed for quantifying the degree of acoustic overlap between vowels. The SOAM does not fully take account of probability densities. An alternative metric is proposed which is based on quadratic discriminant analysis and takes account of probability densities in the form of a posteriori probabilities. Unlike the SOAM, the a-posterior-probability-based metric allows for a direct comparison of vowel overlaps calculated using different numbers of dimensions, e.g., three dimensions (F1, F2, and duration) versus two dimensions (F1 and F2).

PACS number: 43.70.Jt

I. INTRODUCTION

Wassink (2006) presented a metric for quantifying the degree of overlap between pairs of vowel categories in a two-dimensional (2-D) first and second formant (F1–F2) space, and a three-dimensional (3-D) F1–F2–duration space. Wassink’s spectral overlap assessment metric (SOAM) is based on the assumption that the acoustic properties of vowel categories can reasonably be represented by multivariate normal distributions of normalized formant and duration values. SOAM calculation consists of the following steps: 1) The mean vectors and covariance matrices of the distributions of the two vowel categories are estimated via least-squares fits to sample data. 2) Ellipsoids extending two standard deviations from the category means are calculated. 3) A regular grid of points is projected into the 2-D / 3-D space. 4) The proportion of the number of points falling within the intersection of the two ellipsoids relative to the number of points falling within a single ellipsoid is calculated. This results in two proportions, one for when the single ellipsoid corresponds to category *A* and one for when it corresponds to category *B*. 5) The larger of the two proportions is used as the overlap metric (Ω_{SOAM}).

Wassink’s SOAM has some similarities with quadratic discriminant analysis (Hastie, Tibshirani, and Friedman, 2001, §4.3), but the use of a regular grid does not fully exploit information about the probability densities of the two distributions (an earlier vowel-overlap metric based on planimetric convex polygons, Brubaker and Altshuler, 1959, also failed to take account of probability densities). This results in a practical problem with the SOAM: It cannot be used to compare the overlaps of pairs of vowels in different numbers of acoustic dimensions (this problem was acknowledged by Wassink, 2006, p. 2340). If two vowel categories differ in their F1 and F2 distributions, but have identical duration distributions, then duration does not contribute to the separation of the two categories. In this situation, a desirable property for an overlap metric would be that the value calculated on the first two dimensions (F1 and F2) be the same as the value calculated on all three dimensions (F1, F2, and duration). However, the 2-D Ω_{SOAM} value would be greater than the 3-D Ω_{SOAM} value. The 2-D SOAM procedure is somewhat akin to measuring the

overlap of two spheres by assuming that they are instead two cylinders of equal height and thus reducing the problem to the calculation of the overlap of two circles. This will always overestimate the overlap of the two spheres except when overlap is complete or nil.¹ When one considers probability density, the situation is further complicated by the fact that the density of each sphere is not constant and the overlap of more dense regions should count more than the overlap of less dense regions. Hence, except at the two extremes, the 2-D Ω_{SOAM} value will always be greater than the 3-D Ω_{SOAM} value irrespective of whether the third dimension contributes to the separation of the two vowel categories. Therefore, the values calculated using the SOAM cannot be used to determine whether the degree of overlap between two vowel categories is smaller if one considers vowel duration in addition to F1 and F2 as opposed to considering only F1 and F2.

In this letter an alternative vowel-overlap metric is proposed. Using the proposed metric, overlap values can be directly compared irrespective of the number of acoustic dimensions considered. The proposed metric is based on quadratic discriminant analysis and exploits probability density information in the form of a posteriori probabilities.² If homogeneity of covariance matrices were assumed, then standard analytic statistics such as Pillai's Trace or Wilks's Lambda could be used as metrics of vowel overlap, and both have in fact been used in this way (Hay et al., 2006; Morrison, 2004). Both the SOAM and the proposed a-posterior-probability-based overlap metric differ from Pillai's Trace and Wilks's Lambda in that they are based on quadratic discriminant analysis and therefore use separate covariance matrices for each category. They are thus more flexible in that they are applicable in situations where the assumption of homogeneity of covariance is questionable (as appears to be the case in the data plots presented in Wassink, 2006, Fig. 2).

II. CALCULATION OF A-POSTERIORI-PROBABILITY-BASED OVERLAP METRIC

The numerical procedure for calculating the a-posterior-probability-based overlap metric is presented below. See EPAPS Document No. ____ for a Matlab function implementing the procedure. This document can be reached through a direct link in the online article's HTML

reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).

- 1) On the basis of experimental sample data, use least-squares fits to estimate
 - a) the mean vector and covariance matrix for category A .
 - b) the mean vector and covariance matrix for category B .
- 2) On the basis of the estimated mean vectors and covariance matrices, use a multivariate sample generator to generate a large number of fresh points for each category.
- 3) Train a quadratic discriminant analysis model on the generated data.
- 4) For every generated point from category A , calculate its a posteriori probability for membership of category B .
- 5) Calculate the mean of the a posteriori probabilities from step 4.
- 6) Repeat steps 4 and 5 to calculate the mean a posteriori probability of generated points from category B as members of category A .
- 7) Add the results of steps 5 and 6. Use this as the overlap metric (Ω_{app}).

The procedure can be expanded to any number of dimensions and any number of categories. For calculation of Ω_{app} for more than two categories: At stage 4, calculate a posteriori probabilities for all $\neg A$ points as members of category A . At stage 6, cycle through all other categories B, C, D , etc.. At stage 7, divide the sum of the a posteriori probabilities by one less than the number of categories.

III. CALIBRATION OF A-POSTERIORI-PROBABILITY-BASED OVERLAP METRIC AND COMPARISON WITH SOAM

The a-posterior-probability-based overlap metric is a number between zero and one. An Ω_{app} close to one indicates a high degree of overlap, and an Ω_{app} close to zero indicates very little overlap. To provide a calibration of the a-posterior-probability-based overlap metric and a comparison with the SOAM, a series of category A and category B distributions were generated using predetermined parameter values, and overlaps were calculated. Distributions A and B were both spherical with

variances set to one and covariances to zero. The means of distribution A were fixed at the origin, and the means of distribution B were roved. Five-hundred-thousand sample points were generated for each category in the a-posterior-probability-based procedure,³ and a total of one-million grid points were used in the SOAM procedure. The 2-D and 3-D Ω_{app} values were calculated using the same set of generated samples: The 2-D Ω_{app} values were calculated using a quadratic discriminant model trained only on the first two dimensions $[x, y]$ from the generated samples, and the 3-D Ω_{app} values were calculated using all three dimensions $[x, y, z]$. Table I presents the results, which are discussed in the following paragraphs.

Distribution B_1 is identical to A , so Ω_{app} is very close to one and Ω_{SOAM} is exactly one.

As distribution B moves away from distribution A along the x axis (B_2, B_3, B_4) both metrics decrease towards zero. By design, Ω_{SOAM} is exactly zero for separations of two standard deviations or greater. In contrast, Ω_{app} does not suffer from this arbitrary floor effect.

Distributions B_5 , and B_6 illustrate that an equal-magnitude shift away from A in one-, or two-dimensions results in the same value for 2-D Ω_{app} and 3-D Ω_{app} . The third dimension does not contribute to the separation of the two categories, and this fact is reflected in the identical values for the 2-D Ω_{app} and 3-D Ω_{app} . In contrast, the 3-D Ω_{SOAM} value is smaller than the 2-D Ω_{SOAM} value. The SOAM therefore incorrectly indicates that the third dimension contributes to the separation between distribution A and distribution B_5 or B_6 , whereas the a-posteriori-probability-based metric correctly indicates that the third dimension does not contribute to the separation between distribution A and distribution B_5 or B_6 .

Distributions B_5, B_6 , and B_7 illustrate that an equal-magnitude shift away from A in one-, two- or three-dimensions results in the same 3-D Ω_{app} values. The 3-D Ω_{SOAM} values are also the same for the three distributions. Distribution B_7 differs from distributions B_5 , and B_6 in that the magnitude of the shift in two-dimensions is less than the magnitude of the shift in three-dimensions. This difference is reflected in larger values for both 2-D Ω_{app} and 2-D Ω_{SOAM} for distribution B_7 compared to distributions B_5 , and B_6 .

Distributions B_5 , B_6 , and B_8 illustrate that an equal-magnitude shift away from A in one- or two-dimensions results in the same 2-D Ω_{app} values. The 2-D Ω_{SOAM} values are also the same for the three distributions. Distribution B_8 differs from distributions B_5 , and B_6 in that the magnitude of the shift in three-dimensions is greater than the magnitude of the shift in two-dimensions. This is reflected in the fact that the 3-D Ω_{app} value is smaller for distribution B_8 compared to distributions B_5 , and B_6 . This is also true for the 3-D Ω_{SOAM} value.

Finally, distributions B_7 , B_8 , and B_9 illustrate that a shift including a shift in the third dimension can result in 3-D Ω_{app} values which are substantially smaller than the corresponding 2-D Ω_{app} values. The 3-D Ω_{SOAM} values are also smaller than the 2-D Ω_{SOAM} values; however, this is also true for distributions B_5 and B_6 which do not differ from category A on the third dimension. Thus, the difference between the 2-D and 3-D Ω_{app} values, but not the difference between the 2-D and 3-D Ω_{SOAM} values, can be taken as an indication of the magnitude of the contribution of the third dimension to the separation between the two distributions.

The relationship between the SOAM and the a-posteriori-probability-based overlap metric is further illustrated in Fig. 1 which graphs the relationship between the overlap metrics and a shift of distribution B along the x axis. The Ω_{app} (solid line) is identical whether calculated on two or three dimensions. The 3-D Ω_{SOAM} (dotted line) is smaller than the 2-D Ω_{SOAM} (dashed line), even though the third dimension does not contribute to the separation of the two distributions. Due to not fully taking into account probability densities and using an arbitrary floor, both 2-D and 3-D Ω_{SOAM} underestimate the degree of overlap and have discontinuities at an x -axis separation of two standard deviations.

IV. EXAMPLE OF USE OF A-POSTERIORI-PROBABILITY-BASED OVERLAP METRIC WITH NATURAL DATA

Wassink (2006) discussed the overlap between /æ/ and /ɛ/ in Midwestern US English (Hillenbrand et al., 1995).⁴ She noted that for steady-state F1 and F2 values there appeared to be a

larger overlap between the two vowel categories in adult male speakers' productions than in adult females speakers' productions. She also noted that women appeared to produce greater vowel inherent spectral change (formant movement) during /æ/, whereas men appeared to produce similar magnitudes of vowel inherent spectral change in /æ/ and /ɛ/.

As a demonstration of the a-posterior-probability-based procedure applied to natural data, tests were conducted to determine the degree of overlap between /æ/ and /ɛ/ in male and in females speakers' productions. Overlap values for /æ/ and /ɛ/ were calculated for the adult speakers in the Hillenbrand et al. (1995) data using a one-dimensional space (duration), a two-dimensional space (steady-state F1 & F2), a three-dimensional space (steady-state F1 & F2, plus duration), a four-dimensional space (F1 & F2 at 20% of the duration of the vowel, plus the change in F1 & F2 from 20% to 80% of the duration of the vowel), and a five-dimensional space (the four-dimensional space plus duration).⁵ First and second formant values for /æ/ and /ɛ/ were log-interval normalized (Nearey, 1978; Nearey and Assmann, 2007), and vowel durations were independently log-interval normalized.⁶ Data from 44 men and 47 women were included. Plots of the two- and three-dimensional distributions are provided in Figure 2, and Table II provides the calculated overlap values for the one- through five-dimensional spaces. Table II also provides Ω_{SOAM} values; however, one should keep in mind that Ω_{SOAM} values cannot be directly compared across different numbers of dimensions. Also, note that the SOAM floor effect results in several Ω_{SOAM} values of zero where Ω_{app} values range from 0.02 to 0.07. The significance levels reported in Table II are based on randomization tests on the Ω_{app} values.⁷ Monte Carlo simulations were conducted on Ω_{app} values to determine whether there were significant reductions in vowel overlap when additional dimensions were considered. For both men and women, there was a significant reduction in overlap when dynamic spectral properties were compared to steady-state spectral properties, and when duration properties were considered in addition to spectral properties.⁸

In terms of F1 and F2 steady-states (two dimensions), the men produced a substantially and significantly greater overlap between /æ/ and /ɛ/ than did the women (Ω_{app} of 0.87 versus 0.63).

When dynamic spectral information was considered (four dimensions), both men and women had a substantial drop in vowel overlap compared to when only steady-state values were considered (Ω_{app} dropped from 0.87 to 0.32 for men, and from 0.63 to 0.17 for women). Although it just failed to meet the customary significance level of 0.05, the gender difference in overlap values calculated on dynamic formant data was still relatively large (Ω_{app} of 0.32 versus 0.17). These results support Wassink's observation that the men produced a greater spectral overlap between /æ/ and /ɛ/ than did the women.

When vowel duration was considered, there was a substantial decrease in vowel-category overlap for both men and women. The magnitude of the reduction was, however, greater for men than for women: For example, when duration was considered in addition to dynamic spectral values (five versus four dimensions) Ω_{app} dropped by 0.28 points for men (0.32 to 0.04), but only by 0.15 points for women (0.17 to 0.02). When duration was included in the calculation of vowel-category overlap, the gender difference was small and not significant. Thus although the men produced a greater spectral overlap between /æ/ and /ɛ/ than did the women, they produced a smaller duration overlap, and hence men and women had a similar degree of overlap between the two vowels when both spectral and duration properties were considered. Similar results could have been obtained via a direct application of quadratic discriminant analysis. The comparison of these two vowels indicated that they were better separated when duration properties were considered, as was the case with the discriminant analysis conducted on the whole vowel set by Hillenbrand et al. (1995). The advantage of using an overlap metric is that it provides an easily interpretable quantitative measure of the degree of overlap between vowels, which appears to have been an important objective in Wassink (2006). The a-posteriori-probability-based overlap metric is intended as a replacement for the SOAM, not as a replacement for straight forward quadratic discriminant analysis.

V. CONCLUSION

In conclusion, the a-posteriori-probability-based overlap metric is superior to Wassink's

(2006) spectral overlap assessment metric, because it more fully takes account of probability densities and can therefore be used to compare the degree of overlap between vowels when different numbers of acoustic dimensions are considered. Unlike Ω_{SOAM} values, Ω_{app} values can therefore be used to determine whether speakers produce a greater separation between a vowel pair if one considers vowel duration in addition to F1 and F2 values.

ACKNOWLEDGMENTS

This research was supported by the Social Sciences and Humanities Research Council of Canada. Thanks to James M. Hillenbrand, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler for making their data available. Thanks to the Associate Editor and anonymous reviewers for comments on earlier versions of this letter.

NOTES

1. When one considers probability density, the situation is further complicated by the fact that the density of each sphere is not constant and the overlap of more dense regions should count more than the overlap of less dense regions.

2. An alternative statistical technique on which to base an overlap metric is logistic regression. Werker et al. (2007) made use of hierarchical multi-level logistic regression, which directly deals with repeated measures issues and provides a parametric test of between-subject-group differences; however, as implemented, the model was unidimensional rather than multidimensional. Logistic regression would also be problematic in cases where there is complete separation between categories.

3. This resulted in reasonably consistent estimates of the a-posterior-probability-based overlap metric: One-thousand repetitions of the calculation of the overlap between distributions A and B_7 resulted in a 2-D Ω_{app} mean of 0.5724 with a standard deviation of 0.832×10^{-3} , and a 3-D Ω_{app} mean of 0.4496 with a standard deviation of 0.776×10^{-3} . Similarly, the Ω_{app} values calculated for the Hillenbrand et al. (1995) data had standard deviations of less than 1×10^{-3} .

4. Data from the Northwestern US dialect and the Jamaican dialects presented in Wassink (2006) were not available to the author of the present letter, hence an analysis of this data using the a-posterior-probability-based overlap metric could not be presented here.

5. The use of two sample points to represent dynamic formant information was the most successful procedure tested in Hillenbrand et al. (1995).

6. Each speaker-dependent displacement was the mean of the logs of steady-state F1 and F2 from the speaker's /æ/ and /ɛ/ productions. Had the displacements been based on the whole set of vowel categories rather than only the two vowels of interest, then differences between men's and women's

data in the randomization tests below could have been due to a possible shift of both vowels together rather than differences in overlap between the two.

7. The randomization tests were conducted as follows: An index was associated with each speaker's data set. The indices were randomly permuted one-thousand times. For each permutation, the data sets corresponding to the first 44 indices were assigned to one group (same number of indices as in the original male group), and the data sets corresponding to the remaining 47 indices were assigned to another group (same number of indices as in the original female group). The /æ/–/ɛ/ overlap metrics were calculated for each group, and the unsigned differences in the overlap metrics between the two groups were compared with the unsigned differences between freshly calculated /æ/–/ɛ/ overlap metrics for the original male and female groups. The reported p -values are the proportion of comparisons for which the differences between the randomly permuted groups was greater than or equal to the differences between the original groups.

8. The Monte Carlo simulations were conducted separately for men and women using the following procedure: Mean vectors and covariance matrices were calculated for each vowel category on the basis of the sample data. The sample mean vectors and covariance matrices were treated as population mean vectors and covariance matrices, and used to generate one-thousand simulated sample sets of the same size as the original sample data (44 tokens of each vowel for the men and 47 for the women). Overlap metrics were calculated for each of the one-thousand generated sample sets and used in paired-sample Wilcoxon signed rank tests comparing differences across different numbers of dimensions. (A non-parametric test was adopted since Ω_{app} values are markedly non-linear in the vicinity of the extreme values of 1 and 0.) In all cases investigated (four versus two dimensions, three versus two dimensions, and five versus four dimensions) the value of the signed-rank test statistic T was zero (i.e., for all simulated sample data sets the overlap calculated on the smaller number of dimensions was greater than the overlap calculated on the larger number of dimensions), hence the calculated p value was also zero.

REFERENCES

- Brubaker, R. S., & Altshuler, M. W. (1959). "Vowel overlap as a function of fundamental frequency and dialect," *J. Acoust. Soc. Am.* **31**, 1362–1365.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York).
- Hay, J., Warren, P., and Drager, K. (2006). "Factors influencing speech perception in the context of a merger-in-progress," *J. Phonetics* **34**, 458–484.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Morrison, G. S. (2004). "An acoustic and statistical analysis of Spanish mid-vowel allophones," *Estudios de Fonética Experimental* **13**, 11–37.
- Nearey, T. M. (1978). *Phonetic Feature Systems for Vowels* (Indiana University Linguistics Club, Bloomington IN).
- Nearey, T. M., and Assmann, P. F. (1986). "Modeling the role of vowel inherent spectral change in vowel identification," *J. Acoust. Soc. Am.* **80**, 1297–1308.
- Nearey, T. M., and Assmann, P. F. (2007 in press). "Probabilistic 'sliding-template' models for indirect vowel normalization," in *Experimental Approaches to Phonology*, edited by M. J. Solé, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford).
- 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**, 2334–2350.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). "Infant-directed speech supports phonetic category learning in English and Japanese," *Cognition* **103**, 147–162.

TABLE I. SOAM values (Ω_{SOAM}) and a-posterior-probability-based overlap metric values (Ω_{app}) calculated on two $[xy]$ and three $[xyz]$ dimensions. Distributions A and B were spherical with equal variances. The mean vector of distribution A was fixed at $[0\ 0\ 0]$, and the mean vector of B was roved.

B relative to A	B means $[x\ y\ z]$	Ω_{SOAM}		Ω_{app}	
		2-D	3-D	2-D	3-D
$B_1 = A$	$[0\ 0\ 0]$	1.00	1.00	1.00	1.00
B_2 1σ shift on x axis	$[1\ 0\ 0]$	0.69	0.63	0.80	0.80
B_3 2σ shift on x axis	$[4\ 0\ 0]$	0.00	0.00	0.07	0.07
B_4 3σ shift on x axis	$[9\ 0\ 0]$	0.00	0.00	0.00	0.00
B_5 $\sqrt{2}\sigma$ shift on x axis	$[2\ 0\ 0]$	0.39	0.31	0.45	0.45
B_6 $\sqrt{2}\sigma$ shift on xy plane	$[\sqrt{2}\ \sqrt{2}\ 0]$	0.39	0.31	0.45	0.45
B_7 $\sqrt{2}\sigma$ shift in xyz space	$[\sqrt{4/3}\ \sqrt{4/3}\ \sqrt{4/3}]$	0.50	0.31	0.57	0.45
B_8 $\sqrt{2}\sigma$ shift on xy plane + $\sqrt{2}\sigma$ shift on z axis	$[\sqrt{2}\ \sqrt{2}\ 2]$	0.39	0.11	0.45	0.23
B_9 1σ shift on xy plane + 2σ shift on z axis	$[\sqrt{1/2}\ \sqrt{1/2}\ 4]$	0.69	0.00	0.80	0.06

TABLE II. SOAM values (Ω_{SOAM}) and a-posterior-probability-based overlap metric values (Ω_{app}) for Midwestern US English /æ/ and /ɛ/ productions in the Hillenbrand et al. (1995) data. Overlaps calculated separately for men and women. The p -values indicate the statistical significance of the unsigned difference between the men’s and women’s Ω_{app} values calculated via randomization tests.

dimensions	Ω_{SOAM}		Ω_{app}		
	men	women	men	women	p
1. duration	0.00	0.15	0.07	0.12	0.278
2. steady-state F1 & F2	0.88	0.56	0.87	0.63	0.041
3. steady-state F1 & F2 + duration	0.00	0.01	0.06	0.09	0.518
4. initial F1 & F2 + $\Delta F1$ & $\Delta F2$	0.23	0.05	0.32	0.17	0.080
5. initial F1 & F2 + $\Delta F1$ & $\Delta F2$ + duration	0.00	0.00	0.04	0.02	0.565

FIGURE CAPTIONS

FIG. 1. Relationship between the a-posterior-probability-based overlap metric (solid line), 2-D SOAM (dashed line), 3-D SOAM (dotted line), and a unidimensional separation of a pair of equal-variance spherical distributions (separation measured in standard-deviation or d' units).

FIG. 2. (Color online) Plots of the two-dimensional distributions (normalized steady-state F1 & F2) and three-dimensional distributions (normalized steady-state F1 & F2 + normalized duration) of /æ/ and /ɛ/. Plots contain five-thousand tokens of each vowel generated on the basis of the mean and covariance matrices calculated from the Hillenbrand et al. (1995) data. (a) 2-D men's data. (b) 2-D women's data. (c) 3-D men's data. (d) 3-D women's data.



