University of Pennsylvania Working Papers in Linguistics

Volume 7
Issue 3 *Papers from NWAV 29*

Article 6

1-1-2001

When It's Not Nice To Be Normal: What's Missing from Normalized Data?

Betsy Evans
evansbel@msu.edu

Dennis R. Preston
preston@msu.edu

When It's Not Nice To Be Normal: What's Missing from Normalized Data?

When It's Not Nice To Be Normal: What's Missing from Normalized Data?

Betsy Evans and Dennis R. Preston

It is widely accepted that normalization of vowel formant data is a necessity for cross-respondent comparison. We agree that normalization is a critical step for investigating the phonetic properties of a corpus. Nevertheless, we have found, by comparing normalized F1 scores and index scores assigned according to the relative position of a vowel within a system, that discrete differences, visible within individual systems, are lost when only normalized data are examined. Further, we found that tests for statistical significance revealed non-significant results for some variables when using the normalized data but significant results when the index scores were used. Thus we find that evidence of change and the significance of some variables are not always visible when only the normalized F1/F2 scores are examined without consideration of the individual speaker's entire system.

All linguists agree that, because of differences in vocal tract size, normalization is a necessity when comparing vowel formant data of one speaker to another. The upper diagram in Figure 1 shows the large differences due to vocal tract size between child, female, and male means (Hindle 1978). The lower diagram of Figure 1 shows the effect of the reduction of those differences through normalization procedures (Nearey 1977). Hindle (1978) concludes that Nearey (1977) normalization is equal to other normalization procedures for doing away with differences due to vocal tract size but superior to other normalization procedures for preserving differences of the sort sociolinguists are interested in (e.g., status, sex, phonetic environment, etc.)

We attempt to verify whether it is true that the sociological differences that sociolinguists are interested in are, in fact, preserved once a normalization procedure has been carried out. To achieve this, we examined data drawn from our study of Appalachian immigrant respondents and their offspring in Ypsilanti, Michigan. Our main focus in that study was to examine to what degree they had accommodated to features of the Northern Cities Shift (NCS), an ongoing sound change in urban areas in the northeastern part of the United States. For purposes here, we will give data only for F1 (height) of /æ/.

The vowel charts of Darcy's (a fifty-year-old female Ypsilanti respondent, Figure 2) normalized and unnormalized vowels show exactly what Hindle suggests; that is, the relationships among the vowels remain the same, showing in this case the reduction of the overall vowel space expected for a female respondent without compromising the character of Darcy's system. In fact, t-tests on both normalized and unnormalized means of /æ/ and /ɛ/ show that they are statistically not different for F1 (height).

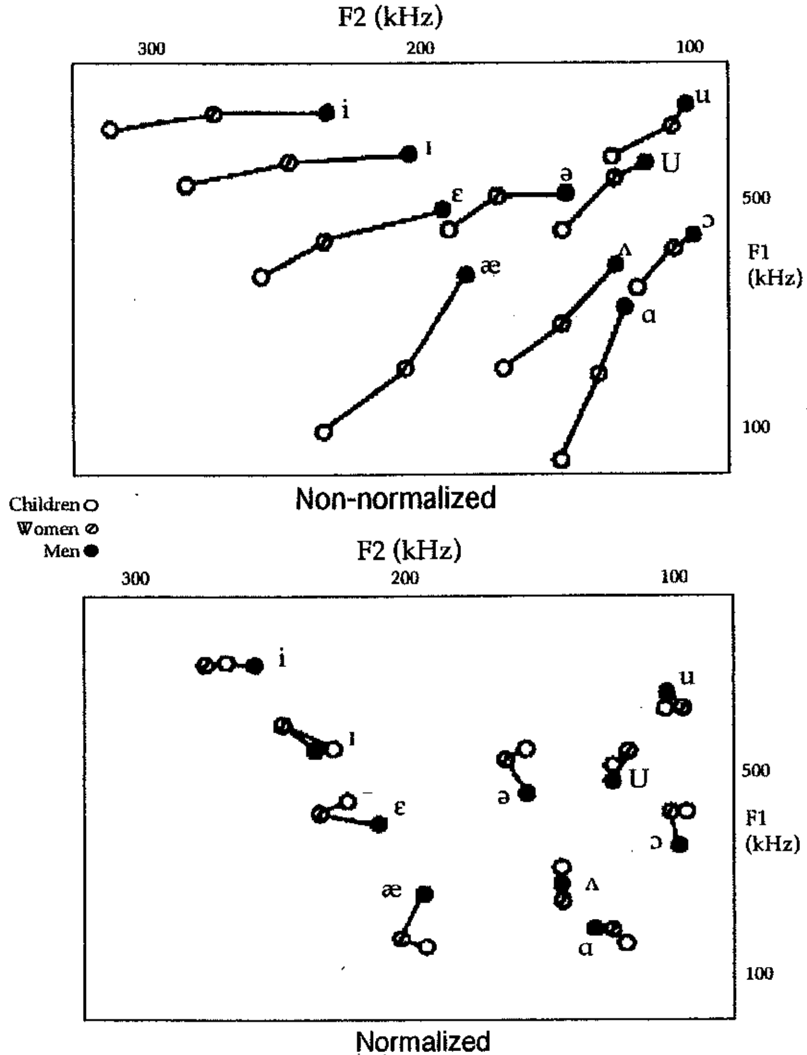

Figure 1. Normalized (Nearey 1977) and unnormalized (Hindle 1978) F1-F2 plots

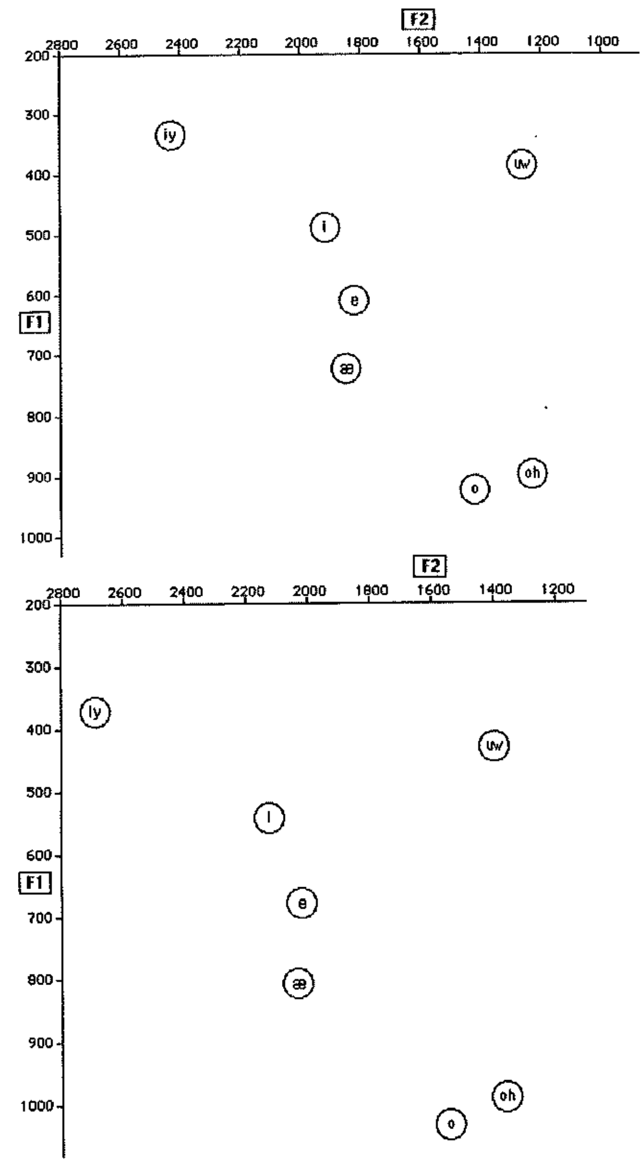

Figure 2: Normalized (above) and unnormalized (below) mean vowel plots for Darcy, a fifty-year-old female Ypsilanti respondent.

The character of an individual vowel system, whether normalized or not, can, however, be quantified using a system of index scores. For our analysis, we assigned index scores based on the relative positions among the vowels. For example, Table 1 shows that for /æ/ height, we assigned a score of 1 if the vowel was significantly lower than /ɛ/ (as revealed in a t-test). We then assigned a score of 2 when the mean F1 score for /æ/ was not significantly different from /ɛ/, a score of 3 when /æ/ was significantly higher than /ɛ/ but closer to /ɛ/ than to /I/, a score of 4 when /æ/ was closer to /I/ than to /ɛ/ but significantly different from /I/, and a score of 5 when /æ/ was not significantly different from /I/. Interestingly, these relationships were not different for normalized and unnormalized data, but, of course, the actual F1 scores were been changed by the normalization procedure.

In spite of the efficacy of the normalization procedure for preserving the character of an individual's system, a comparison of Darcy's F1 score for /æ/ with another speaker to compare the phonetic height of /æ/ demonstrates a problem. Table 2 for example, shows that Darcy's F1 for /æ/ (either normalized or unnormalized) is lower than David's, an older male speaker in the same study. This would lead to the erroneous conclusion that David's /æ/ is higher than Darcy's. A t-test between Darcy's F1 scores for /æ/ and /ɛ/ (again, whether on normalized on unnormalized data) reveals no significant difference (an index score of 2). David, however, who has a higher absolute score even in the normalized comparison, does have a significantly lower low-front vowel and is assigned an index score of 1. If the F1 for /æ/ is considered along with its relative position in the individual's system, a comparison allowed by index scores, it is evident that Darcy's /æ/ is higher than David's.

Index for (æ)	
t-test result	Rating
Significantly lower than /ɛ/	1
Not different from /ɛ/	2
Significantly different from /ɛ/ and /I/ but closer to /ɛ/	3
Significantly different from /ɛ/ and /I/ but closer to /I/	4
Not significantly different from /I/	5

Table 1: Index scores assigned to the variants of (æ)

	F1 unnormalized/ normalized	Index score based on unnormalized/normalized data
Darcy	810/725	2/2
David	627/665	1/1

Table 2: Normalized, unnormalized, and index scores for two respondents (F1 for /æ/)

Table 2 shows a comparison of these index scores and formant measurements.

We believe that such relative positions when converted to index scores (rather than absolute acoustic values) are more revealing of the relative positions which are important in language variation and change studies. In addition, when normalized scores are pooled in statistical studies, seeking a correlation between vowel height and such categories as sex, age, status, and network relations, we find that normalization does, in fact, destroy some of the distinctiveness[1] which the relative positions of the vowels retain when expressed as index scores. To show how this may be so, we compared the results of correlations between various social facts and our normalized scores as opposed to index scores. See Table 3.

Social category	Type of test	Result for normalized data	Result for index scores
Sex	t-test	p = .52	p = .02
Actual Age	Pearson correlation	r = -10 p= .605	r = -.43 p < .03
Network	Pearson correlation	r = -.11 P = .58	r = -.58 P = .001

Table 3: Significant and non-significant results for normalized versus index scores for F1 (æ) and selected social characteristics

We found a significant difference among index scores but not among the raw normalized F1 data for the category of sex. A t-test on the F1 index scores for men and women showed a significant difference (p=.02), but the same test on the normalized F1 scores for men and women was not significant (p=.52). When the category of age was examined, a Pearson-product moment correlation showed a negative correlation for index score

[1] Disner (1980) notes that normalization may "... obscure or even reverse the linguistic trends which, in fact, are present in the raw data" (257).

and actual age (age in years) ($r=-.43$, $p<.03$) while the same test on actual age and normalized F1 scores was not significant ($r=-.10$, $p=.605$). Network scores, calculated on the basis of Milroy's (1980) procedure, showed a negative correlation (again using a Pearson test) between the network score and the F1 index score ($r = -58$, $p = .001$), but the same test using normalized F1 scores showed no significant difference ($r= -.11$, $p=.58$).

Although there were no differences in tests performed on normalized and index scores for status (actual score [Pearson] and groups [t-test]) and age groups (ANOVA), we are nevertheless suspicious that normalization will not accomplish what Hindle suggests—a minimization of differences based on vocal tract size *and* a retention of differences which will reflect social categories.

We are especially impressed by the fact that actual age and index scores show a significant correlation, since we are sure most would agree that greater distinctiveness is preserved in actual age rather than age groups, but even actual age does not correlate with the normalized data while it does with index scores.

Although we agree that comparison between one set of F-score acoustic measurements and another can only be carried out with normalization, we believe the subsequent use of "raw normalized" formant scores may mask some of the distinctiveness actually present when comparing subgroups for social characteristics. We did not consider linguistic environmental characteristics here but believe that those characteristics should also be explored in this manner.

We would agree, however, though our own data show no index score differences between normalized and unnormalized data, that such index scores are also better based on normalized data. The best of all possible worlds, therefore, involves both: the use of normalized data, which can effectively show the purely acoustic differences between the realization of a variable across speakers and the use of index scores, which more accurately reflect the relational, or systematic phonetic differences among individuals and are, therefore, better suited to correlation with other factors, for they appear not to mask such relational differences.

References

Disner, Sandra Ferrari. 1980. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America* 67:253-261.

Hindle, Donald. 1978. Approaches to vowel normalization in the study of natural speech. In David Sankoff (ed.), *Language variation: Models and methods*, 161-71. New York: Academic Press.

Nearey, T. M. 1977. Phonetic Feature Systems for Vowels, Doctoral dissertation, University of Connecticut, Storrs.

Department of Linguistics and Languages
A-614 Wells Hall
Michigan State University
East Lansing, MI 48824-1027
evansbe1@msu.edu
preston@msu.edu