Language Variation and Change, **22** (2010), 175–190.

doi:10.1017/S0954394510000104

The effect of outliers on the perception of sound change

WILLIAM LABOV
University of Pennsylvania

MACIEJ BARANOWSKI
University of Manchester

AARON DINKIN
Swarthmore College

ABSTRACT

The role of outliers in vowel distributions is examined through an experiment that registers subjects' assessments of symmetrical and asymmetrical distributions. Subjects first scaled their subjective impressions of a series of six resynthesized tokens of the word *bad* ranging from low front to upper mid front. They were then asked to register on this scale their overall impressions of four series of five phrases including *bad:* low symmetrical, low with a very low outlier, high symmetrical, high with a very high outlier. Subjects show the capacity to integrate outliers into their overall assessments in a manner consistent with their acoustic properties. The effect of low outliers was significantly greater, reflecting the socially marked status of this form in the Philadelphia community.

Since the late 1960s, the acoustic analysis of vowel systems has played a major role in elucidating the mechanism of sound change (Labov, Yaeger, & Steiner, 1972). In every such system examined, the analyst must deal with the problem of which tokens and which measurements best represent the central tendencies of a given vowel as perceived by listeners and learners. This paper reports an experiment that helps resolve the question of how to deal with the eccentric tokens that appear as outliers to the main distribution. It further explores the question of what role such outliers play in the progress of linguistic change.

In plotting the distributions of stable vowels, we often find a tightly clustered distribution as in Figure 1.[1] This represents the measurements of the nuclei of 24 tokens of the phoneme /ow/ (*goat, so*, etc.) as spoken by a woman from Birmingham, Alabama. The bars extend outward from the mean to one standard deviation. It is evident that all 24 tokens are well within two standard deviations from the mean. But we also find the situation of Figure 2, the dispersion of 16 /ow/ tokens as spoken by a 12-year-old girl from Charleston, South Carolina, where /ow/ is undergoing extreme fronting and lowering from its original monophthongal position in the traditional Charleston dialect (Baranowski, 2007). One token of *no* is within one standard deviation from the mean for both

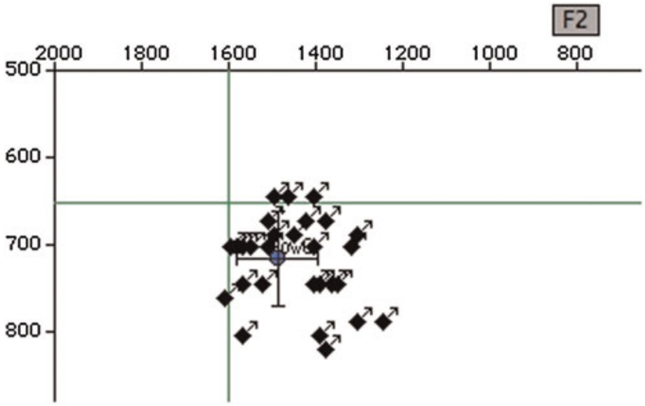

FIGURE 1. Distribution of /ow/ (not before /l/) for Thelma M., 41, Birmingham AL, TS 3, interviewed in 1995.

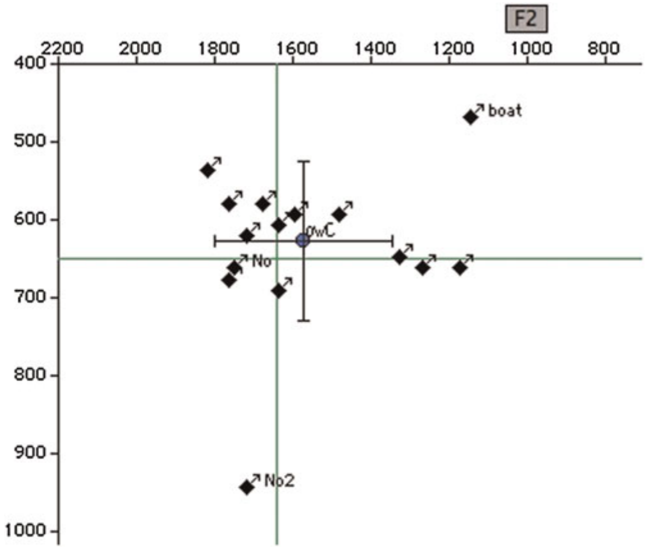

FIGURE 2. Distribution of /ow/ (not before /l/) for Nicole D., 12, Charleston SC, TS 07, interviewed in 1996.

F1 and F2, but F1 of the other, *no2*, is over two standard deviations greater than the mean. The token of *boat* is in a conservative position, about 1.5 standard deviations from the mean on both the F1 and F2 dimensions.

Outliers have been given a great deal of attention by statisticians. The *NIST/SEMATECH e-Handbook of Statistical Methods* (2010 [2003]:7.1.6) states:

> Outliers should be investigated carefully. Often they contain valuable information about the process under investigation or the data gathering and recording process.

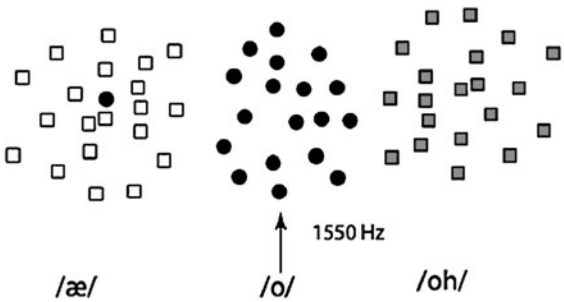

FIGURE 3. Effect of an outlier with symmetrical distribution of neighbors.

> Before considering the possible elimination of these points from the data, one should try to understand why they appeared and whether it is likely similar values will continue to appear. Of course, outliers are often bad data points.

and the StatSoft Statistics Glossary adds:

> Outliers are atypical (by definition), infrequent observations; data points which do not appear to follow the characteristic distribution of the rest of the data. These may reflect genuine properties of the underlying phenomenon (variable), or be due to measurement errors or other anomalies which should not be modeled.

It is clearly important to decide whether outliers represent noise or some inherent characteristic of a distribution. If they are only noise, our understanding of the phenomenon will be improved by deleting them. In the study of sound change, the critical question is how listeners—and language learners—deal with outliers. This question plays a central role in the general theory of chain shifts, which rests on the phenomenon of probability matching (Labov, 2010:143-4). In the stable situation of Figure 3, a phoneme /o/, as in *cot* and *Don*, has two symmetrical neighbors with margins of security that allow a field of dispersion within the typical range of speakers' accuracy. An outlier of /o/ is squarely within the distribution of /æ/, as in *cat* and *Dan*. The hypothesis is that such an outlier will not be recognized as a member of the /o/ category as frequently as those within the main distribution, but in some (perhaps a small) number of cases will be misunderstood as /æ/ and so not enter into the pool of tokens from which the language learner calculates the mean or target value of /o/ in matching the distribution of the target phoneme. This will then have a conservative effect and the mean for F2 of /o/ will remain at 1,550 Hz. But in Figure 4, /æ/ has moved away; the same outlier is less likely to be misunderstood as a member of /æ/ and the mean therefore shifts in that direction, say to 1,571 Hz. This first step in the mechanism of a drag chain sets the stage for future fronting of the /oh/ phoneme in *caught, law*, etc., as /o/ shifts to the front.

However, we do not in fact know how listeners deal with outliers. Do they pay less attention to them, even when there is no competing phoneme? Or do they pay

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

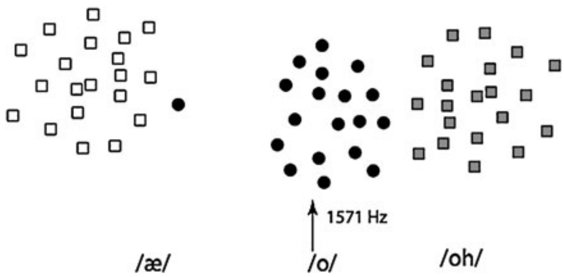

FIGURE 4. Effect of an outlier with asymmetrical distribution of neighbors.

more attention to them? Given the existence of change in progress, the outlier in the direction of change may be weighted more heavily than others in calculating the mean. The third possibility is that presumed in Figure 4: The outlier is entered into the mean calculation with the appropriate distance measure.

Outliers may be defined mathematically by Grubb's Test (NIST/SEMATECH, 2010 [2003]:1.3.5.17.) If we want to test whether the farthest token from the mean of a given distribution is to be considered an outlier, we define *G* as that token's distance from the mean, measured in units of *s*, the standard deviation.

(1) *G* can then be compared to the test statistic (2):

$$G = \frac{\max |Y_i - \overline{Y}|}{s}$$

(2)

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{\alpha/(2N),N-2}}{N-2+t^2_{\alpha/(2N),N-2}}}$$

where *t* denotes the critical value of the *t*-distribution with (*N* – 2)/2 degrees of freedom when alpha is the desired statistical significance level. If *G* is greater than the test statistic, then that token can be considered an outlier and discarded unless there is reason to do otherwise.

In the case of the /ow/ distribution from Charleston, *no2* has a *G* value on the F1 dimension of 3.08, and with an alpha value of .01, the test statistic is 2.94. We can conclude that there is only one chance in a hundred that *no2* is not an outlier. Figure 5 shows the distribution of 22 /æ/ tokens in the speech of a Buffalo woman. The token *sad* is about 1.5 standard deviations from the mean on the F2 dimension. The parameter *G* of (1) is 2.87 and with an alpha value of .05, the test statistic of (2) is 2.84. The .05 value establishes that there is 1 chance in 20 that *sad* is not an outlier.

This paper reports an experiment to resolve the question of how listeners deal with phonetic outliers. It consists of two steps. First, the subjects establish their

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

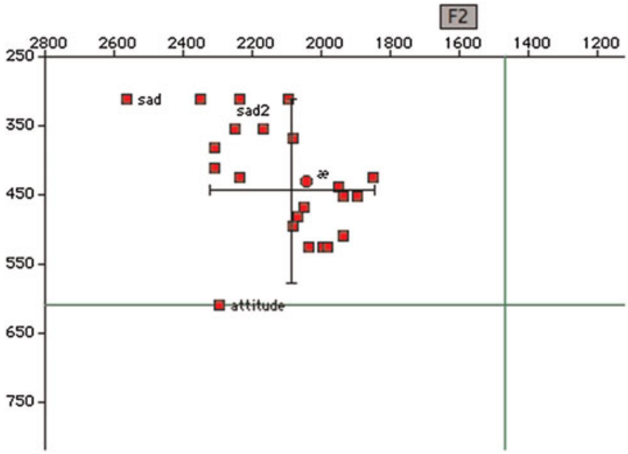

FIGURE 5. /æ/ outlier *sad* in vowel system of Jeannette S., 56, Buffalo, NY, TS 347, interviewed in 1995.

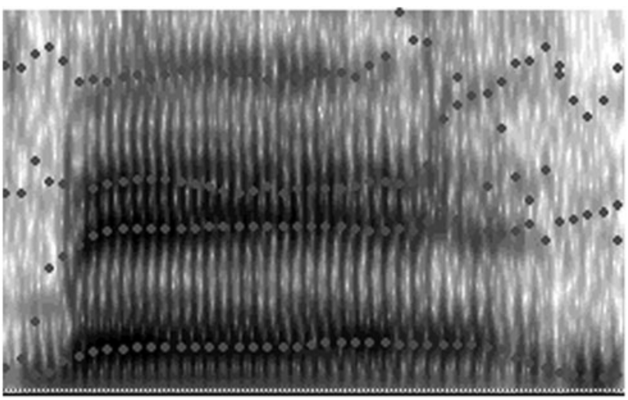

FIGURE 6. Spectrogram of *bad* as originally spoken by Melinda D., 28, South Philadelphia.

own impressionistic ordering of a resynthesized series of vowels. Second, they register their overall impressions of a sequence of phrases with and without outliers on the scale that they have themselves established.

STEP 1: IMPRESSIONISTIC ORDERING OF A RESYNTHESIZED SERIES

We began with the word *bad* originally pronounced by a young South Philadelphian woman as a lower mid ingliding vowel. The spectrogram and linear predictive coding coefficients are shown in Figure 6.

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

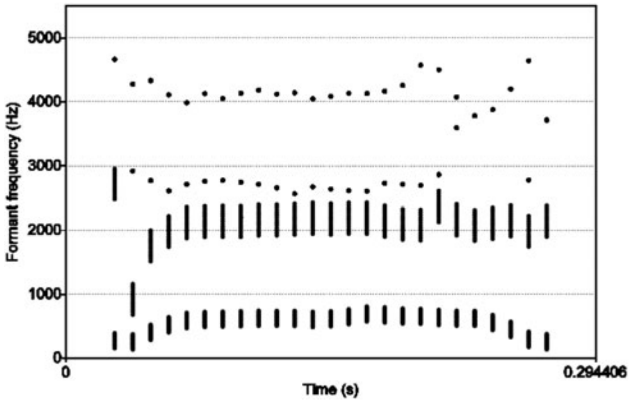

FIGURE 7. Modifications of F1 and F2 of Figure 6 shown as vertical bars.

TABLE 1. *Eleven resynthesized versions of* Bad *with six selected for Step 1*

	F1	F2
Bad1	802	1,729
	791	1,813
Bad2	780	1,921
	752	1,953
Bad3	686	2,050
	609	2,135
Bad4	594	2,212
	556	2,318
Bad5	480	2,353
	421	2,490
Bad6	392	2,534

This was modified in F1 and F2 to produce a series of 11 tokens along the axis of change common to the speech community, from low front to peripheral upper mid. The range of modifications of the first two formants are shown in Figure 7, and the F1 and F2 measurements of the resulting stimuli are given in Table 1. The series comprised the original token of “bad” with five tokens higher and fronter and five tokens lower and backer. F1 ranged from 802 Hz to 392 Hz, and F2 from 1,729 Hz to 2,534 Hz.

Spatial and auditory sorting of the stimuli

The first step was designed to obtain subjects’ perception of the relative ordering and auditory distances of the resynthesized series and to provide the framework for their judgments of central tendencies in Step 2. A total of 55 subjects were recruited by advertisement in the University of Pennsylvania community. The great majority of subjects were undergraduates; their geographic and gender

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

TABLE 2. *Social characteristics of the subjects*

Gender	*n*	Geographic region (4–13 yrs)	*n*
Male	19	Philadelphia–New York City	21
Female	36	Midland, West	12
		South	8
		North	7
		Outside of U.S.	7

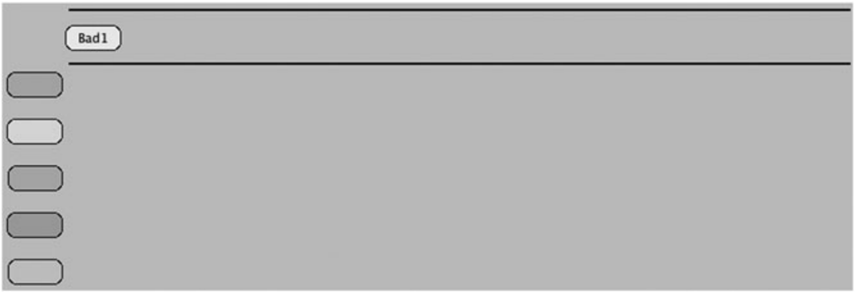

FIGURE 8. Initial configuration for sorting by spatial and auditory distance in Step 1.

distribution are shown in Table 2. None had received special training in linguistics or phonetics.

The six tokens labeled *Bad1* through *Bad6* as in Table 1 were associated with unlabeled (colored) buttons on a computer screen as shown in Figure 8. Subjects were then asked to order these tokens by dragging the six buttons between the two parallel lines displayed, with the instructions given in (3).

(3) Instructions for Step 1.
On this page you'll see a column of six colored buttons. When you click on each one you will hear a different pronunciation of the word "bad." They are all spoken by Melinda, a young woman from South Philadelphia, at different times in her life.
The yellow button "Bad1" is fixed. We'd like you to drag each of the other buttons to some position between the two lines, so that the ones that sound the most different are the furthest apart, and the ones that sound most alike are closer together.
Please spread the buttons out so they don't overlap. If two buttons sound very much alike, put one next to the other, not on top.
When you've got the best arrangement, click "Done".

Subjects performed well on Step 1 in both the order and the linear spacing of the five tokens *Bad2– Bad6*. Of 55 subjects, 47 arranged them in the same order as the formant measurements. The eight others each showed only one reversal of order; six involved the order of *Bad5* and *Bad6*.

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

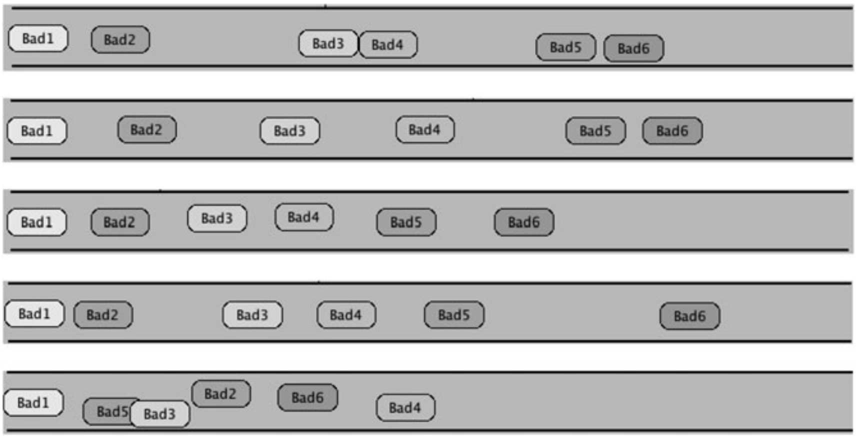

FIGURE 9. Spatial arrangements of *Bad1–Bad6* buttons by the first five subjects in Step 1.

The degree of linearity of the spacing is shown by *r*-correlations of the perceptual scale values with the linear regression line: .99 for the 47 subjects with no reversals of order and equally high for all 55 subjects combined. Figure 9 shows the final arrangements for subjects 1–5 of Step 1. It can be seen that all but one are arranged from left to right in an order corresponding to Table 1. Some tokens were heard by some subjects as closer than others, whereas other subjects spaced the six tokens quite evenly. Most importantly, it can be seen that subjects differed in the proportion of available space they utilized, an important consideration for the use of these arrays in Step 2.

The mean positions of the six *Bad* buttons placed by the subjects were fitted to the physical F1 measurements by the following procedure. The experimental program recorded the horizontal locus of each button in pixels. Each such mean location was then adjusted to match the mean location of button *Bad1* and other values scaled accordingly.[2] The results are shown in Figure 10. The two series fit closely and are ordered in exactly the same way at the same distances. The physical measurements fit a linear trend line with an r^2 of .98; the locations fit a linear trend line with an r^2 of .99. As Table 1 shows, our efforts to achieve an equal series of perceptual steps resulted in a minimal difference of only 22 Hz in F1 between *Bad1* and *Bad2*. The mean scaling values of the phonetically naïve subjects were accordingly closer to linearity than the F1 values. Figure 11 shows a comparable result in fitting the average locations to the F2 measurements.

There were no significant differences by gender in responses to Step 1, either in ordering of the original items or in the width of the total space occupied. The geographic origin of subjects was examined. It was thought that subjects from the Philadelphia and New York City regions might show different responses from others, because the raising and fronting of the vowel of *bad* is a socially stratified variable in those cities with the strong reactions in subjective reaction tests characteristic of stigmatized variables (Labov, 1966, 2001). Such

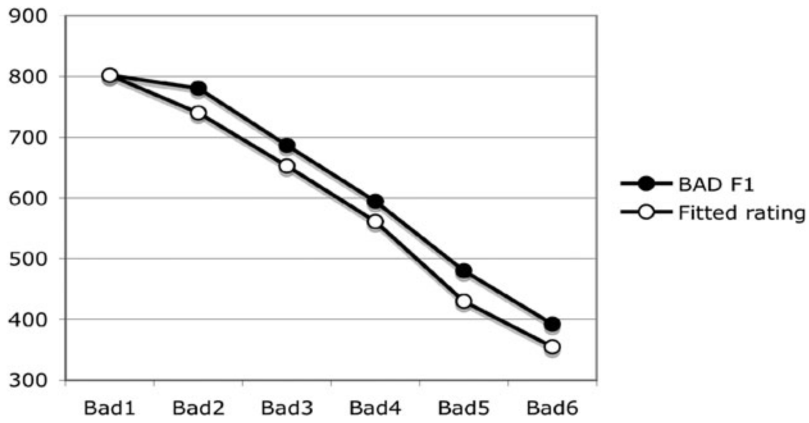

FIGURE 10. Mean values of *Bad1–Bad6* ordering fitted to F1 measurements in Step 1.

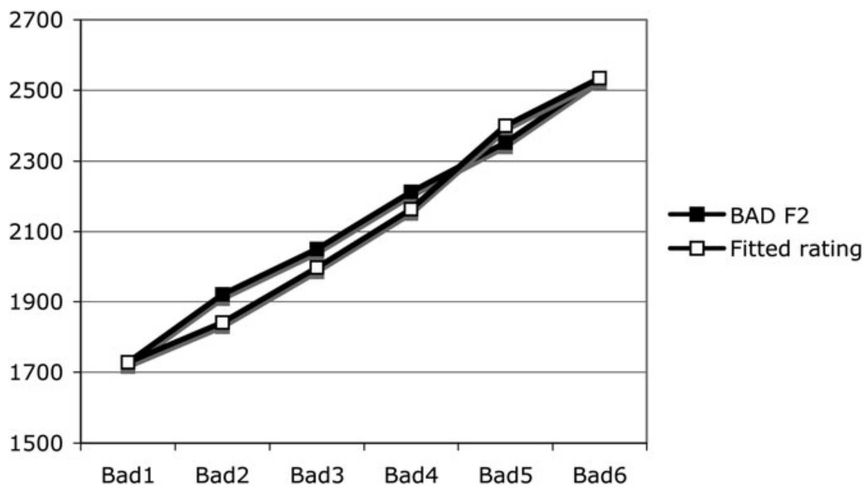

FIGURE 11. Mean values of *Bad1–Bad6* ordering fitted to F2 measurements in Step 1.

an effect did appear. Table 3 shows the number of adjacent pairs (*Bad1*/*Bad2*, *Bad2*/*Bad3*, etc.) ordered consistently with the physical dimensions (in terms of F1 and F2) by region; five such pairs constitute maximum agreement with F1/F2 measurements. Of the eight subjects who failed to show such consistent order, five were from the Mid-Atlantic region (including New York City).[3] In all five cases, it was the order of *Bad5* and *Bad6* that was reversed. This would indicate that this degree of raising and fronting is identified as a distinct category ("tense") by the Mid-Atlantic subjects, who do not distinguish within the category as consistently as those who treat the whole series as a phonetic continuum. Analysis of variance applied to the number of ordered pairs shows

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

TABLE 3. *Number of adjacent pairs ordered to match F1 and F2 by geographic origin*

Region	Four	Five
Mid-Atlantic	5	16
North	1	6
South	2	6
West	0	3
Midland	0	9
Outside of U.S.	0	7

one main effect for Mid-Atlantic origin ($F = 5.3$, $p = .025$). Such a result is consistent with previous findings that the Mid-Atlantic and New York City both show a split of the /æ/ word class into two distinct lexically and grammatically defined categories.

Aside from this Mid-Atlantic tendency to group the highest stimuli as a single category, all speakers were able to use their psychoacoustic abilities to order the stimuli along a single dimension in ways that correspond closely with physical reality.

STEP 2: PERCEPTION OF CENTRAL TENDENCIES WITH AND WITHOUT OUTLIERS

Step 2 was designed to measure subjects' perception of the central tendency of a distribution of vowels heard in context, comparing symmetrical with asymmetrical distributions to detect the effect of outliers. Subjects heard groups of five short phrases all ending in the word *bad*. The carrier phrases were recorded by the South Philadelphia woman who spoke the initial token of *Bad1*, along with five combinations of the altered and resynthesized *bad* tokens. The five phrases heard were:

He's really bad
It's too bad
Pretty bad
That's bad
He's really bad

Table 4 shows the design of the experiment with four different distributions of the /æ/ vowels in the five phrases. Columns 2–5 identify these four as "Low symmetrical," "Low outlier," "High symmetrical," and "High outlier," and the numbers in the column identify the type of /æ/ nucleus heard in the word *bad* for each set. The term *symmetrical* refers to the distribution of the five phrases: three with the central value, one that is a single step higher, and one that is a single step lower. The low symmetrical distribution is centered on three phrases embodying *Bad3*, with one phrase a single step higher and another a single step

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

TABLE 4. *Stimuli for Step 2 (values in Hertz)*

	Low symmetrical	Low outlier	High symmetrical	High outlier	Vowel tokens	F1	F2
F1 + 200, F2 – 400					*Bad1*	802	1,729
F1 + 160, F2 – 320		1				791	1,813
F1 + 120, F2 – 240					*Bad2*	780	1,921
F1 + 80, F2 – 160	1					752	1,953
F1 + 40, F2 – 80	3	3			*Bad3*	686	2,050
F1, F2	1	1	1	1		609	2,135
F1 – 340, F2 + 80			3	3	*Bad4*	594	2,212
F1 – 80, F2 + 160			1			556	2,318
F1 – 120, F2 + 240					*Bad5*	480	2,353
F1 – 180, F2 + 320				1		421	2,490
F1 – 200, F2 + 400					*Bad6*	392	2,534

lower. In the low outlier distribution, the lower phrase is shifted three units lower on the 11-step scale, with an F1 of 791 Hz as compared to the mode of 686 Hz, and F2 of 1,813 Hz as compared to a modal value of 2,050 Hz. The high symmetrical distribution has a mode at *Bad4*, with one phrase higher and one lower on the 11-step scale; the high outlier pattern shifts the higher value three units to an F1 of 421 Hz and F2 of 2,490 Hz.

The rightmost two columns give the F1 and F2 measurements of the original series of 11 resynthesized forms of *bad* and the column labeled "Vowel tokens" identifies the 6 items that were used in Step 1 and form the basis of the impressionistic scale established by each subjects.

It should be noted that the original stimuli for Step 1 were created to give approximately equal steps in impressions of vowel height and fronting, and Step 1 confirmed the success of this design. The difference in the F1 values of the low outlier and low symmetrical token is 39 Hz, but the difference in the F1 values of the high outlier and its symmetrical counterpart is much larger, 135 Hz.

Figure 12 shows the screen configuration at the beginning of Step 2, with instructions that are both heard and read by the subjects. The colored buttons between the two horizontal lines are fixed at the final position that this subject registered on Step 1.

The first speaker, referred to as Melinda A, is heard as the low symmetrical series of Table 4. In response, subjects adjust the position of the triangular marker to register their overall impression of the /æ/ realization in relation to the

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

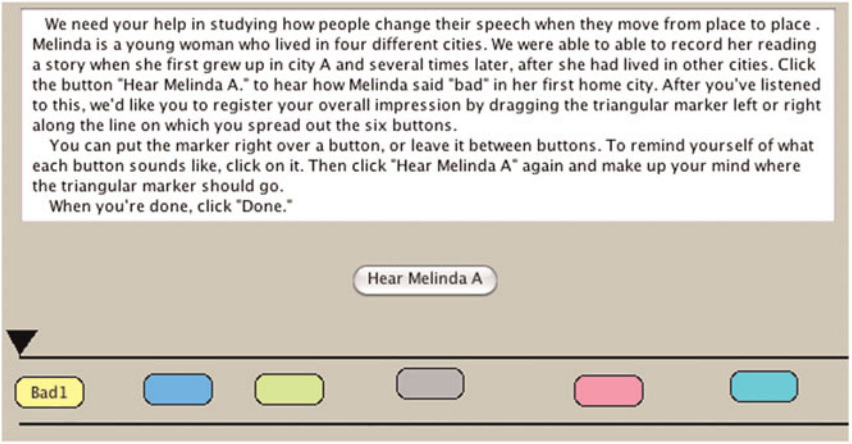

FIGURE 12. Initial instructions for Step 2.

six *Bad* tokens. When subjects indicate that they have arrived at a final decision by hitting *Done*, the screen displays the instructions for a second series:

> Now we'd like you to listen to the way Melinda said "bad" after she had moved to city B and lived there for four years. Click on the button "Hear Melinda B" to hear the phrases with "bad". Then move the marker to where it should go to register your overall impression as to how she spoke then. When you're done, click "Done."

Melinda B is the high outlier series of Table 4. It is followed by similar instructions for Melinda C, the low outlier distribution, and finally by Melinda D, the high symmetrical series.

Subjects generally proved to be sensitive to the effect of the outliers in Melinda B and C. Mean ratings of both series were significantly different from the corresponding symmetrical series (for the low outlier, $p < .0001$; for the high outlier, $p = .002$).

Because the total range used by each subject was different, the mean final positions of the markers were normalized by the following procedure. Each subjective rating in the form of the horizontal location of the triangular marker (in pixels) was transformed to the proportion of the distance from the base position on the left (the fixed position of *Bad1*) to the horizontal location of the rightmost button (*Bad6* for most speakers). Each physical mean was transformed to the proportion occupied of the total range of F1 (or F2) of the 11-step F1/F2 range.

Figure 13a compares these proportional values for the physical measurements of F1 (solid lines) with the subjective ratings (dotted lines). The left shows the symmetrical series, and the right shows the series with outliers. The slope of the solid lines shows the mean difference produced by the outlier: a small distance for the low outlier and a somewhat larger one for the high outlier. The subjective

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

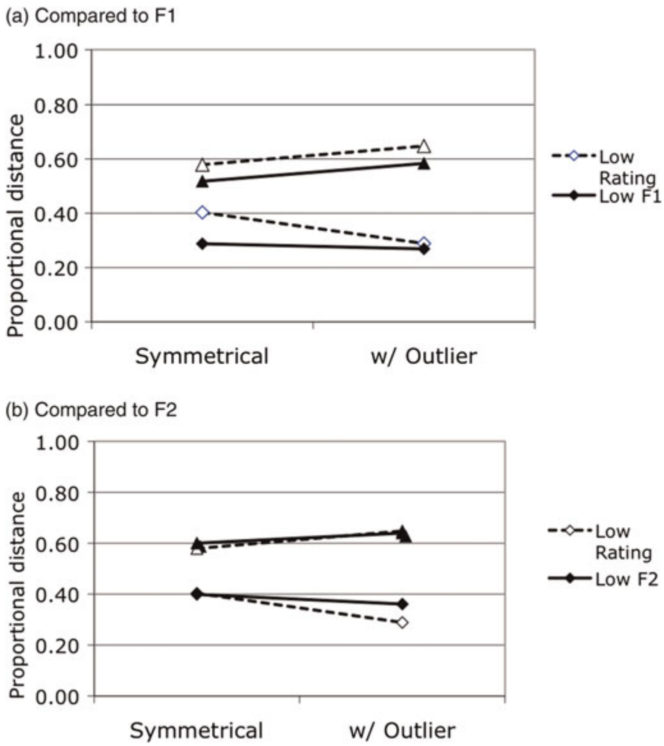

FIGURE 13. Proportional distances of subject ratings of symmetrical and asymmetrical series compared to proportional distance of physical measurements. In each figure, upper two lines represent high stimuli, and lower two lines represent low stimuli.

reaction to the high outlier is parallel to the physical difference, indicating that listeners are successfully averaging the impact of the high outlier in their estimate of the overall impression of vowel quality. The lower half of the diagram shows that the low outlier also has an effect on listeners' impression, but even more so. The lower dotted line—the subjective ratings for the low outlier—falls with a slope greater than the solid line—the F1 measurement for the low outlier. Thus, the mean subjective reaction indicates that the low outlier has a greater impact on the mean reactions of listeners than the high outlier does. Figure 13b shows that this result holds for both F1 and F2 measures.

The shifts of ratings for both low and high outliers are both significant (low: $t = -6$, $p < .0001$; high: $t = -2.89$, $p = .0055$). It is also evident that the effect of the low outlier is somewhat stronger. The greater subjective effect of the low outlier might be attributed to the fact that it represents a correction of the vernacular form in Philadelphia; subjective reaction tests have shown that such a correction elevates ratings on the job scale of social evaluation (Labov, 2001:ch. 6). However, this result holds for all subjects, the 21 from the Mid-Atlantic States and those raised in other dialect areas.

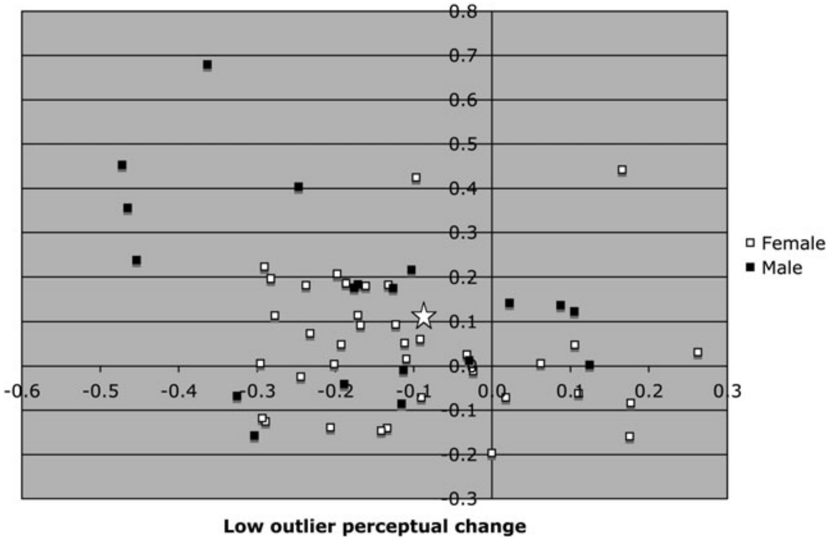

FIGURE 14. Scattergram of proportional shifts to outliers in Step 2 by gender. Star = Cartesian distance on F1/F2 plane between symmetrical mean and outlier mean divided by range; high = .11, low = .07.

Although the mean values of ratings show the regular relations of Figure 13, individual subjects were not wholly consistent in the rankings of stimuli in Step 2. Figure 14 is a scattergram of the proportional shifts of the 56 subjects. The horizontal axis shows the proportional shift created by the low outlier, calculated as the difference between the perceptual rating of the low outlier group and the low symmetrical group divided by the range of perceptual ratings involved. The vertical axis is the corresponding measure for effect of the high outlier. The star shows the proportionate shift of the outlier groups in physical terms: .07 for F1 and .11 for F2. It is clear that the majority of subjects respond to the outliers in a way that is not far from the physical difference: That is, they are capable of arriving at a perceptual summary that is not far from the mean of the series. This behavior is general to the population of 56 subjects, not significantly different for subjects from different dialect regions, and not significantly different for a subgroup of seven nonnative speakers.

There are four speakers who differ from the main body of subjects, who are in effect outliers to this study of outliers. They are the four black squares in the upper left of Figure 14, with a much greater shift for both high and low outliers. The special sensitivity of these subjects to the outliers indicate that they may be treating outliers as indicative of the target position of a phoneme, rather than stray accidents of production. This is certainly not characteristic of the majority of the subjects.

Figure 15 superimposes the perceptual judgments of the symmetrical and asymmetrical series on the original six-member *Bad* scale for a single one of

https://doi.org/10.1017/S0954394510000104 Published online by Cambridge University Press

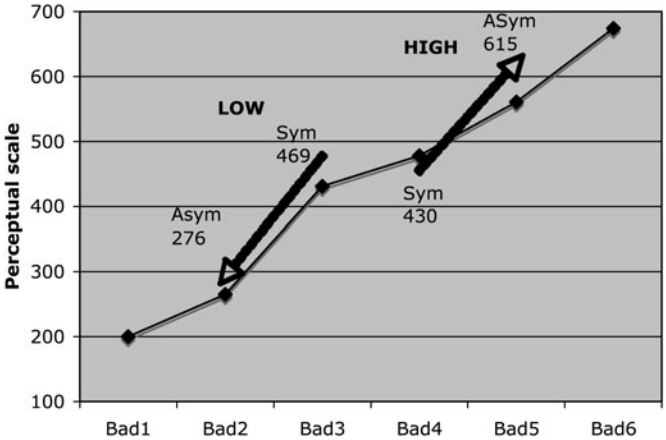

FIGURE 15. Perceptual scale for six *Bad* tokens with outlier shifts for subject JC, Miami.

these four outlying subjects. The solid line connects his original placement of the tokens in Step 1 of the experiment. (Note that the low symmetrical series was centered on *Bad3* and the high symmetrical series on *Bad4*). It is evident that the effect of the single low outlier (just below *Bad2*) on this speaker shifts his overall judgment to *Bad2*, and the single high outlier (one notch above *Bad5*) produces an overall effect equal to his impression of *Bad5*. In other words, the outlier has more effect than the members of the main distribution do.

One cannot avoid noting that all four of these exceptional subjects are male,[4] and that there is no corresponding set of outliers on the lower right corner of Figure 14, male or female. These four speakers are responsible for a significant regression coefficient for males for the high outliers ($p < .03$) and almost significant for the low outliers ($p < .065$).

DISCUSSION

All of these subjects show the ability to integrate the information from outlying tokens with others in the calculation of the central tendency of a vowel. A minority give the outlier a disproportionate importance, in some cases coming close to identifying the mean with the outlier as the "true" target of the speaker. There is no tendency to ignore the outlier or treat it as a mistake, which would lead to a shift of the estimated center in the opposite direction for the asymmetrical series.

It remains to be seen if the small group of speakers who pay particular attention to outliers have a disproportionate effect on linguistic change in progress, either in the initial advance or in the late stages of correction. But for both high and low outliers, the overall results show that outliers cannot be discarded in determining the mean values that correspond to the social perception of a speaker's characteristic target.

This experiment bears directly on the proposals made for the mechanism of sound changes and, in particular, the role of estimates of the central tendency of heard distributions. The adult subjects we are dealing with here have already formed their phonetic targets, but research on language change across the lifespan indicates that they are not immune from change in later life (Sankoff, 2005). The capacity to sum and average the physical properties of the productions of another speaker in interaction is fundamental to an understanding of how maximal dispersion within a vowel subsystem is achieved (Labov, 1994: ch. 20). The results of Step 2 show that speakers control this competence to a fine degree.

The same averaging of tokens is an inherent part of exemplar theory (Pierrehumbert, 2002). Our subjects were exposed to a small cloud (5) of exemplars of /æ/ and were asked to react in a way that suggests target formation. The behavior of the four males discussed in Figure 14 suggests that they have selected a single exemplar, the outlier, to represent the cloud as a whole, and it is possible that many of the other subjects have selected a single exemplar as well. On the other hand, subjects may be forming a more abstract sound image that is not based on any one episodic memory. The data from the experiment is consistent with either possibility.

NOTES

1. These and following examples are normalized vowel plots based on interviews of subjects for the *Atlas of North American English* (Labov, Ash, & Boberg, 2006, henceforth *ANAE*]. See chapters 2–6 for notation and methodology.

2. To compare the physical measurements to the subjects' adjustments of the sliding scale in pixels, F1 was subtracted from 1,002; F2 was added to 1,369.

3. Although New York City differs from the Mid-Atlantic cities of Philadelphia, Wilmington, and Baltimore in many features of its vowel system, it is identified with them in regard to the split of /æ/ into tense and lax categories (*ANAE*, 2006:ch. 13, 17).

4. The four do not seem to share other common social characteristics.

REFERENCES

Baranowski, Maciej. (2007). *Phonological variation and change in the dialect of Charleston, South Carolina*. Publication of the American Dialect Society 92. Durham, NC: Duke University Press.

Labov, William. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

______. (1994). *Principles of linguistic change. Volume I: Internal factors*. Oxford: Basil Blackwell.

______. (2001). *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell.

______. (2010). *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Oxford: Blackwell.

Labov, William, Ash, Sharon, & Boberg, Charles. (2006). *Atlas of North American English: Phonology and sound change*. Berlin: Mouton/de Gruyter.

Labov, William, Yaeger, Malcah, & Steiner, Richard. (1972). A quantitative study of sound change in progress. Philadelphia: U.S. Regional Survey.

National Institute of Standards and Technology (NIST) and International SEMATECH. (2010 [2003]). *NIST/SEMATECH* e-*Handbook of Statistical Methods*. Available at: http://www.itl.nist.gov/div898/handbook/.

Pierrehumbert, Janet. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory phonology*. Vol. 3. Berlin: Mouton/de Gruyter. 101–140.

Sankoff, Gillian. (2005). Cross-sectional and longitudinal studies in sociolinguistics. In U. Ammon, N. Dittmar, K. J. Mattheier, & P. Trudgill (eds.), *An international handbook of the science of language and society*. Vol. 2. Berlin: de Gruyter. 1003–1013.

StatSoft. Statistics glossary. Available at: http://www.statsoft.com/textbook/glosfra.html.