STATISTICAL TESTS FOR THE STUDY OF VOWEL MERGER

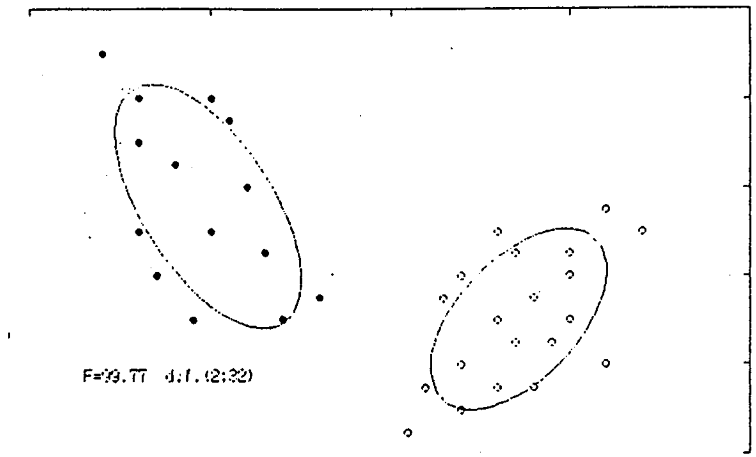
Kikuo MAEKAWA, Tottori

1. A Quantitative Study of Vowel Merger and the Need for Statistical Tests

Phonologically distinct vowels sometimes lose their contrast and merge into one, or conversely, one vowel splits into two distinct vowels. These changes, called "merger" and "split" respectively, are known to be rather common.

Historical linguists and dialect geographers have studied vowel merger and split (hereafter referred to simply as merger) through the comparison of written records from more than two different diachronic stages or from more than two geographically distant dialects. In addition to these rather traditional ways of studying merger, which had as their primary aim analyzing merger in the past, recent progress in the field of sociolinguistics has enabled researchers to observe change in progress directly. The pioneering works of W. Labov and his co-workers gave birth to the new paradigm in the late sixties, and we came to have an empirical reason to claim that, quite contrary to the past belief, it is possible to observe the actual process of merger through the analysis of tape recorded materials carefully gathered according to the rigid sampling methodology.

The use of tape recorded speech has had an additional effect on the organization of the study. Sociolinguists working in the field of phonology often adopt acoustic measurement techniques elaborated in experimental phonetics in order to analyse more exactly and perhaps more successfully the process of ongoing merger. For example in Labov et al. (1972) we see many F1-F2 (first formant versus second formant) vowel charts representing the complicated process of ongoing mergers in the United States and England. Although this quantitative technique has without doubt made a substantial contribution to the study of phonological change, it is also true that it becomes more and more difficult, as the charts increase, to know whether there is really a significant difference between two (or more) defined sets of vowels. Taking for example the artificial data in Figure 1a through c, we can say with confidence that the two vowels sets shown in Figure 1a are distinct, whereas the vowel sets in Figure 1b are not distinct at all, hence merged. But what if we compare Figure 1b with Figure 1c? Some will say that only Figure 1b is in merger, and others will claim that the vowels in Figure 1c are also merged. Thus, the quantitative nature of acoustic measurement gives rise to a new problem here: how do we know that two given sets of vowels are really distinct or not? The problem turns more serious when one is confronted with actual data, where the partial overlapping exemplified in Figure 1c occurs rather frequently. It is apparent, then, that one should be equipped with some reliable means to check the significance of the difference on statistical grounds. The present paper deals with two kinds of statistical tests; one has been rather well known in the field of mathematical statistics and the other is proposed by the present author in his 1984 paper. The combination of these two tests can resolve, if partially, the above mentioned delicate problem that seems to have been left untouched by the preceding literature in the field.



F2 (abscissa) versus F1 (ordinate). In this figure, as well as in Figure 1b,c, two vowels are indicated with open and filled circles

Figure 1a. Example of two distinct vowel sets: artificial data.

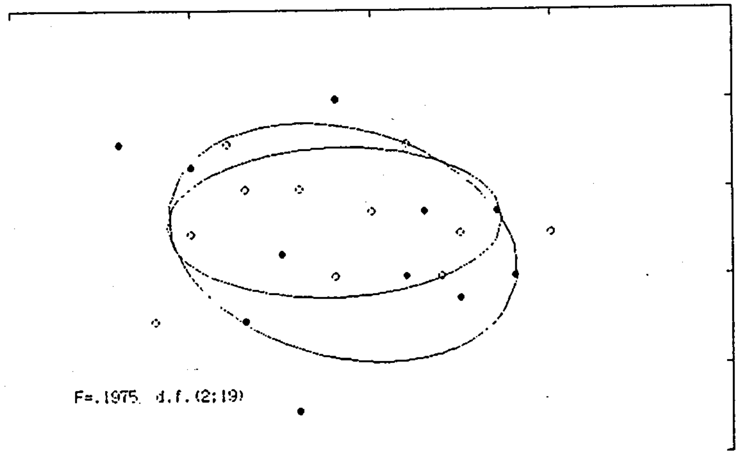


Figure 1b. Example of merged vowel sets: artificial data

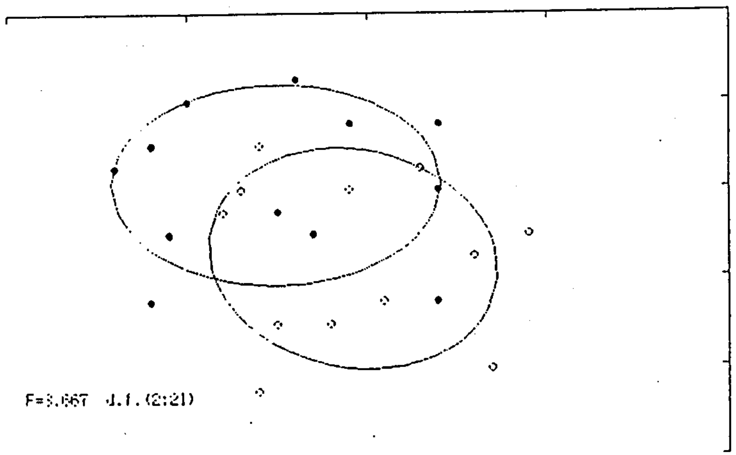


Figure 1c. Example of a doubtful merger: artificial data

2. Test Based on Bivariate Normal Distribution

2.1. The Bivariate Analog of Student's t-test

In Figures 1a through c, ellipses are drawn to visualize the distribution boundaries of given vowel sets. This kind of elliptic representation, used often in the literature of experimental phonetics or sociolinguistics, is called Equal Probability Density Contour and derived from a mathematical function called Bivariate Normal Density Function, the two-variate analog of the univariate normal probability density function (the well known bell-shaped curve). As is well known, in univariate statistics, Student's t-test is commonly used to test the null hypothesis that there is no statistically significant difference between the means of two independent samples. The test which will be explained in this section - and which is based on the bivariate normal density function - is the extension of Student's t-test to the bivariate case; the substantial difference being that this latter test deals with the significance of the difference between the two-component mean vectors rather than scalar means. What follows is the mathematical fomulation by Green (1978).

In the bivariate analog of Student's t-test, we use a statistics called Hotelling's T^2 - the bivariate counterpart of Student's t -, which is defined as

$$T_2 = \frac{m_1 m_2}{m_1 + m_2} \underline{d}'_{\bar{\underline{x}}_1 - \bar{\underline{x}}_2} \; \underline{C}^{-1} \; \underline{d}_{\bar{\underline{x}}_1 - \bar{\underline{x}}_2}$$

where m_1 and m_2 are the sizes of sample one and sample two; the vector $\underline{d}_{\bar{x}_1-\bar{x}_2}$ denotes the difference vector between the sample centroids (the subscripts $\bar{\underline{x}}_1$ and $\bar{\underline{x}}_2$ represent the two sample centroids); and $\underline{C}^{-1}$ denotes the (inverse of the) unbiased sample-based estimate of the population covariance matrix $\underline{\Sigma}$.

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$

where σ_1 and σ_2 are the standard deviation of the variables x_1 and x_2, and r is their population product moment correlation. (Note: we assume that the covariance matrices ot two samples are the same and $\underline{C}$ is estimated from the pooled sample.)

By the way, with regard to Hotelling's T^2, it is known that this statistics follows the F distribution under the following condition:

$$\frac{m-n-1}{n(m-2)} T^2 \approx F[n, m-n-1].$$

where $m = m_1 + m_2$ and n = number of dimensions (here, $n = 2$).

It means that the transformed T^2 statistics follows the F distribution with n degrees of freedom in the numerator and m−n−1 degrees of freedom in the denominator.

Using this property of Hotelling's T^2, we can test the hypothesis concerning the population equality of the means of two given sets of samples. In the case of our specific interest, the study of vowel merger of F1−F2 plane, this hypothesis means that the two sets of vowels are from one and the same population; and if the bivariate t-test does not reject the hypothesis then we would be able to infer from it that the two sets of vowels are in merger. (But see Concluding Remarks below for the problem concerning the mathematical definition of the linguistic notion of merger.) In the next section, we will apply this test to the data of vowel merger in a Japanese dialect.

2.3. Example of the Bivariate T-test: Stylistic Differenciation of the Syllable Merger in Izumo Dialect

As far as the vowel system is concerned, Japanese areal dialects are known to be rather uniform. Most of them have basically the five vowel system of /i/, /e/, /a/, /o/ and /u/, which is identical to the system of so-called Standard Japanese. In this respect, the Izumo (出雲) dialect spoken in the Shimane and the Tottori prefectures of Western Japan is an exceptional case, together with the so-called Tôhoku (東北) dialects (a general term for the dialects spoken in the North-Eastern district of Japan). It is known that in the Izumo dialect the close vowels /i/ and /u/ are merged into a central close vowel when preceded by the dental consonants /s/, /t/ and /z/ in the same syllable. Though a detailed phonological analysis reveals the increasingly complex nature of this phenomenon, here we would like to restrict the problem to the form stated above in order not to lose the real objective of this section.

The results of spectrographic analysis of the tape recorded materials of the Izumo dialect are presented in Figures 2 and 3. The materials are collected in three different styles: reading of a list of nonsense syllables written in syllabic letters called kana (仮名) (abbreviated as SYL), reading of meaningful short sentences written in kana and chinese-origin logographs or kanji (漢字)(STC), and a natural discourse between two native speakers of the Izumo dialect (DIS).

From Figure 2 it is clear that speaker K (male, 70 years old, retired public official) did not merge at all two vowels in SYL, and it is also clear that the merger gradually took place as the style lowered. Basically the same holds true with regard to the speaker T (male, 67 years old, farmer). But it seemed to be the case that the occurrence and the progress of the merger as a function of style shifting differed according to individuals.

The results of the bivariate t-test are shown in Tables 1 and 2. The x means that no significant difference was found at the 5% level. The triangle means that significant difference was found only at 5% level but not at 1% level. The circle means a significant difference at 1% level. The point is that the results of the bivariate t-test seemed to coincide rather well with the previous intuitive investigations by inspection. And with the statistical facts shown in Table 1 and 2, we can say with much confidence that, other things being equal, speaker T preserves more conservative features of Izumo dialect, and speaker K is more standardized in comparison to speaker T.

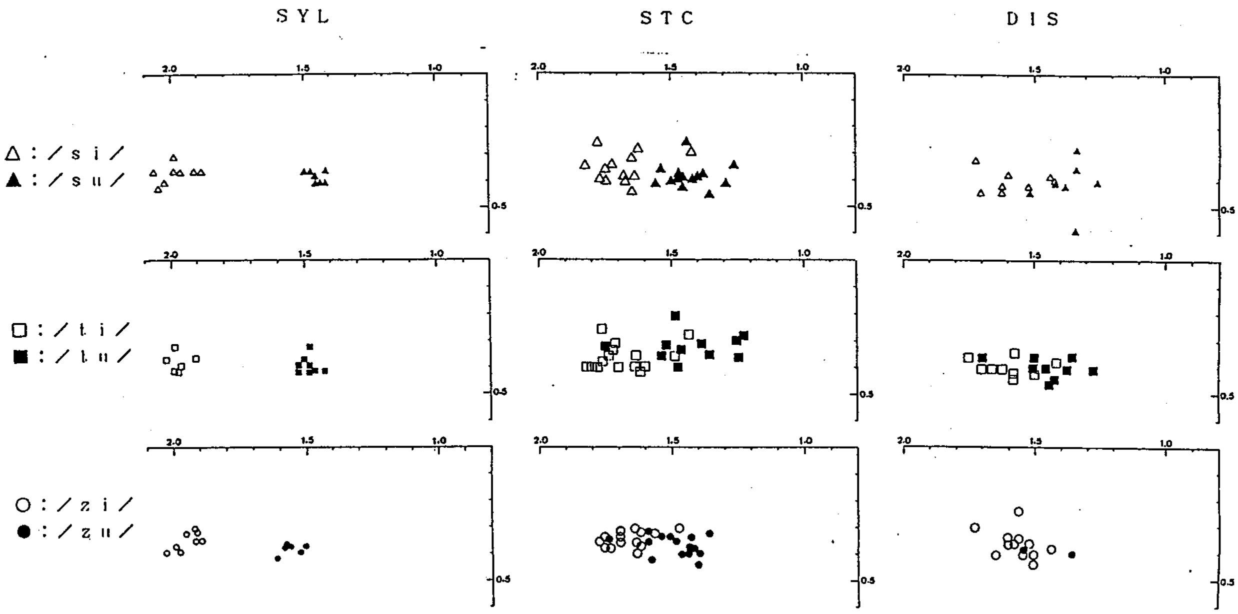


Figure 2. The stylistic differenciation of
three syllable pairs in Izumo dialect
(Speaker K. Male 70 years old.)
F2 (abscissa) and F1 (ordinate) in KHz

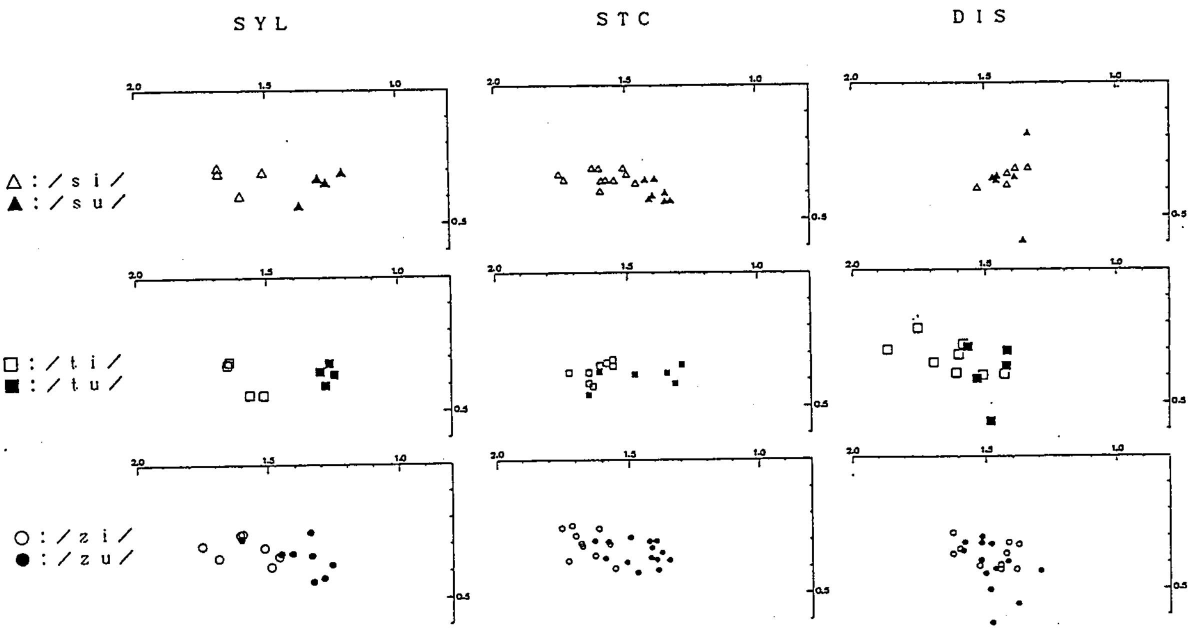


Figure 3. The stylistic differenciation of
three syllable pairs in Izumo dialect
(Speaker T. Male 67 years old.)
F2 (abscissa) and F1 (ordinate) in KHz

Table 1

Results of the Bivariate T-Test for Speaker K of Izumo

Syllables	SYL	STC	DIS
/si/:/su/	○	○	△
/ti/:/tu/	○	○	○
/zi/:/zu/	○	○	×

Table 2

Results of the Bivariate T-Test for Speaker T of Izumo

Syllables	SYL	STC	DIS
/si/:/su/	○	○	×
/ti/:/tu/	○	○	×
/zi/:/zu/	○	○	×

2.3. Practical Limit of the Bivariate T-test

Although the bivariate t-test is useful in the study of merger, this test has some practical limits. The problem that can be encountered most frequently by linguists is the problem caused by individual differences in vocal tract size and shape. As is well known, the locations of the linguistically "same" vowels on the F1-F2 plane differ to a great extent if uttered by different individuals, since the size and the shape of the vocal tract varies on the individual basis, and removing this linguistically meaningless pertubation is not at all an easy job. Even though many normalization procedures have been proposed and some proved to be effective for specific purposes (see Disner 1980 for the concise overview of normalization procedures), it is not an exaggeration to say that there is no omnipotent way of normalization. In the examples presented above, the test is used in order to test the differences existing between two vowel sets spoken by the same speaker and thus caused no problem. But when one wants to apply this test to material uttered by more than two individuals, it is impossible to bypass the problem of normalization.

3. Test Based on Geometrical Relations of Vowels

3.1. Basic Conception

An alternative way of testing merger was proposed in Maekawa (1984). The characteristic of this test is that it is based on the mutual geometrical relations of vowels defined on the F1-F2 plane rather than based on the Bivariate Normal Density Function.

In Figure 4 the sustained five vowels of Tokyo dialect, which approximates the so-called Standard Japanese spoken by 18 male speakers (mostly university students) are plotted and the token pairs of /i/:/e/ and /o/:/u/ uttered by the same speaker are connected by a line. Although each set of vowels shows dispersion caused partly by the individual differences, it is evident that the geometrical relations expressed by those lines are rather identical: the lines belonging to a given pair go side by side in nearly the same direction from /i/ to /e/ or from /u/ to /o/. To put it differently, in Tokyo dialect where no merger happens, all members of a given vowel pair are "parallel" as far as their geometrical relations on the F1-F2 plane are concerned. Figure 5 is the plot of the five vowels uttered by 14 elderly male speakers of the Tsugaru dialect (a typical Tôhoku dialect spoken in the Aomori prefecture of North-Eastern Japan). In Figure 5, as well as in Figure 6, the five vowels are pronounced in the same phonetical context /_ki /. In this context, five meaningful words can be obtained: /iki/ (息: 'breath'), /eki/ (駅: 'railway station'), /aki/ (秋 : 'autumn'), /oki/ (沖 : 'offing') and /uki/ (浮き: 'float'). Figure 6 is the same plot of five vowels uttered by 26 elderly male speakers of the Izumo dialect. The comparison of Figures

Figure 4.
Five Vowels
in Tokyo

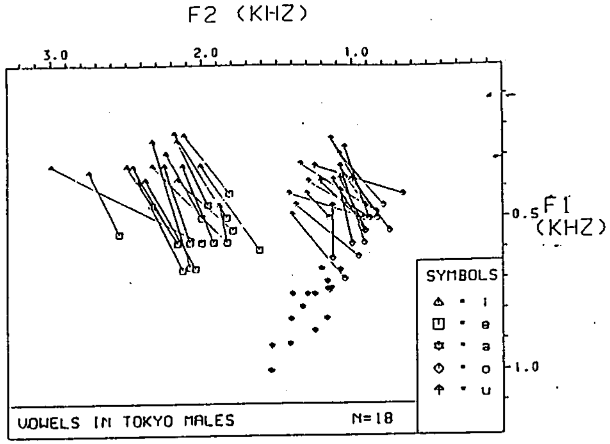


Figure 5.
Five vowels
in Tsugaru

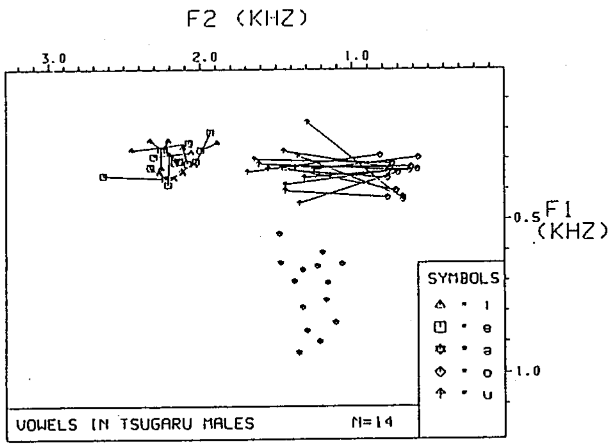
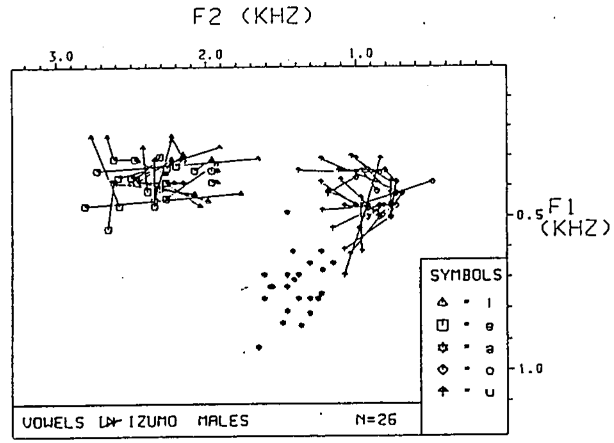


Figure 6.
Five vowels
in Izumo



Figures 4 — 6. Five vowels in Tokyo, Tsugaru and Izumo

5 and 6 with Figure 4 makes it clear that the geometrical relations between /i/:/e/ in both the Tsugaru and the Izumo dialects and those of /o/:/u/ in the Izumo dialect are not "parallel", hence quite different from the relations in Tokyo, whereas the relations between /o/:/u/ in the Tsugaru dialect are "parallel" and hence similar to those in Tokyo.The test introduced by Maekawa (1984) is based on the following assumption. If two vowels, say *j* and *k*, are phonologically distinct, that is if the participants of the same speech community distinguish two vowels in roughly the same way, the lines connecting the pairs of *j* and *k* on the F1-F2 plane hold the "parallel" relation even though they are not exactly parallel in the mathematical sense. On the other hand, if *j* and *k* are completely merged and are phonologically the same vowel, the lines connecting them will disperse in all directions. In this latter case, the directions of lines are supposed to be uniformly distributed.

3.2. Mathematical Formulation

Let two vowels j and k uttered by speaker i be expressed as

$$\underline{V}_{ij} = [f_{1ij}\ f_{2ij}]'$$

$$\underline{V}_{ik} = [f_{1ik}\ f_{2ik}]'$$

where f_1 and f_2 are the scalar values of first and second formant frequencies. The subscript i distinguishes the speaker, and subscripts j and k distinguish the vowels (j,k = /i/,/e/,/a/,...). The vector representation of the lines connecting two tokens of vowels j and k can be defined by the difference vector $\underline{D}_{ijk}$

$$\underline{D}_{ijk} = \underline{V}_{ij} - \underline{V}_{ik} = [\ d_{1ijk}\quad d_{2ijk}\ ]'$$

where $d_{1ijk} = f_{1ij} - f_{1ik}$ and $d_{2ijk} = f_{2ij} - f_{2ik}$.

The information contained in $\underline{D}_{ijk}$ consists of two parts: (a) the information concerning the directional relation of two vowel tokens in the F1-F2 plane, and (b) the geometrical distance between $\underline{V}_{ij}$ and $\underline{V}_{ik}$.

In order to remove information (b), which is irrelevant to the test, we normalize the length of the difference vector by the following transformation.

$$\underline{D}^{*}_{ijk} = \underline{D}_{ijk}/|\underline{D}_{ijk}|,$$

where $|\underline{D}_{ijk}|$ is the norm of $\underline{D}_{ijk}$. Next, let θ_{ijk} be the angle given in radian defined between $\underline{D}_{ijk}^{*}$ (or $\underline{D}_{ijk}$) and the unit vector $\underline{e}_1 = [1 \quad 0]'$. Then the null hypothesis of our statistical test that there is no phonological distinction between j and k can be formulated as

$$\theta \approx U(0,2\pi].$$

This formula means that θ (we refer to the θ_{ijk} simply as θ) is distributed uniformly in the half open interval of (0,2π]. Stated in more empirical terms, this means that the probability of $\theta = x$ ($0 < x \le 2\pi$) is identical for every x in this interval. The probability density function of U is written as

$$f(x) = \Pr\{\theta = x\} = (2\pi)^{-1}.$$

And the expectation and the variance of $f(x)$ are

$$E(x) = \int_0^{2\pi} x f(x)\,dx = \pi .$$

$$V(x) = \int_0^{2\pi} \{x - E(x)\}^2 f(x)\,dx = \pi^2/3 = 3.29.$$

And by the central limiting theorem, the quantity Z defined below is distributed as the standard normal variable N(0,1) when N --> ∞,

$$Z = \frac{\sqrt{N}(\bar{\theta} - \pi)}{S}$$

where S and $\bar{\theta}$ are the sample standard deviation and sample mean of θ. We can test whether our hypothesis defined above is congruent with the data by using the value of Z.

3.3 Example of Test Based on Geometrical Relation: Vowel Merger of Izumo and Tsugaru

The Z values calculated for every possible pair of two vowels of the three dialects discussed in 3.1 are shown in Tables 3, 4 and 5. The values of N.V. (Normalized Variance) shown in these tables are obtained by dividing the sample variance by the theoretical value of $(1/3)\pi^2$ calculated in 3.2. A brief look at the tables is enough to see that Z is small only for the pairs of /i/:/e/ of Tsugaru and Izumo and /o/:/u/ of Izumo. The other 27 values of Z are by far greater. And by consulting the table of the standard normal distribution N(0,1), we see that (a) the null hypothesis can not be rejected at the 5% level as far as the /i/:/e/ of Tsugaru is concerned, (b) as for the /i/:/e/ and /o/:/u/ of Izumo, the hypothesis is rejected at the 1% level, (c) the other 27 pairs are needless to say rejected at the 1% level.

Table 3

Distribution of Θ.
Tokyo dialect

Pairs	$\bar{\theta}$	S	N.V.	Z
/i/:/e/	5.42	0.22	0.01	44.9
/e/:/a/	5.03	0.11	0.01	72.8
/a/:/o/	4.08	0.14	0.01	27.8
/o/:/u/	2.20	0.35	0.04	-11.4
/u/:/i/	1.61	0.06	0.00	-108.0
/i/:/a/	5.15	0.09	0.00	94.7
/i/:/o/	4.86	0.05	0.00	145.0
/e/:/o/	4.68	0.06	0.00	108.0
/e/:/u/	4.45	0.15	0.01	37.0
/u/:/a/	6.48	0.36	0.04	39.3

Table 4

Distribution of Θ
Tsugaru dialect

Pairs	$\bar{\theta}$	S	N.V.	Z
/i/:/e/	2.92	1.83	1.02	-0.45
/e/:/a/	5.15	0.12	0.00	62.6
/a/:/o/	4.12	0.19	0.01	19.3
/o/:/u/	1.61	0.13	0.00	-44.1
/u/:/i/	1.62	0.11	0.00	-51.8
/i/:/a/	5.16	0.10	0.00	75.5
/i/:/o/	4.75	0.03	0.00	201.0
/e/:/o/	4.75	0.04	0.00	151.0
/e/:/u/	4.76	0.12	0.00	50.5
/u/:/a/	5.95	0.47	0.07	22.4

Table 5

Distribution of Θ
Izumo dialect

Pairs	θ	S	N.V.	Z
/i/:/e/	2.10	1.71	0.88	-3.11
/e/:/a/	5.05	0.12	0.00	81.1
/a/:/o/	4.26	0.12	0.00	47.5
/o/:/u/	2.20	1.23	0.46	-3.9
/u/:/i/	1.65	0.11	0.00	-69.1
/i/:/a/	5.20	0.20	0.01	52.5
/i/:/o/	4.79	0.06	0.00	140.0
/e/:/o/	4.76	0.05	0.00	165.0
/e/:/u/	4.75	0.08	0.00	103.0
/u/:/a/	0.86	0.43	0.06	-27.1

This result leads us to the interpretation that a true merger happens only in the pair /i/:/e/ of the Tsugaru dialect, since, as far as this pair is concerned, the mutual geometrical location between the two vowels on the F1-F2 plane is determined at *random* (it is impossible to predict their geometrical relation), and it seems that the relation between the two vowels is equivalent to the relation between the two tokens of a "same" vowel.

4. Concluding Remarks

As shown in the last section, the test based on the geometrical relation between vowels seems to be a useful tool in the study of merger. This rather simple test can successfully handle data uttered by more than two individuals without the normalization problem being insurmountable. However, this test has its own problems, as did the bivariate t-test. First of all, this test can be applied only to the pairs of vowels, and to form a pair we have to rely upon external (linguistic) criteria. For example in the case reported in this paper the criterion is that two vowels contained in a pair are pronounced by the same speaker and different pairs are pronounced by different speakers. Alternatively, the test can be applied also to so-called minimal pairs; in this latter case the pairs consist of two vowels pronounced in the same phonetic context, and different pairs are pronounced by the same speaker. Anyhow, the need for these external criteria restricts the application of this test to specific kinds of data sets.

The second problem concerns the way we define the merger mathematically and hence it is more fundamental. It is worth noting that the null hypothesis that the bivariate population means of /i/ and /e/ in Izumo are equal is not rejected both at 5% and 1% levels when we test the data used in 3.3 by means of the bivariate test. The same result is obtained in the case of /o/ and /u/ of Izumo. This suggests the possibility that the two statistical tests discussed in this paper are testing different things. From the mathematical point of view there is nothing surprising about this, simply because the null hypotheses used in the two tests are totally different; one is concerned with the Bivariate Normal Density Function, and the other is concerned with the uniform distribution. From a practical point of view, however, it may be necessary to add some additional explanations on this matter.

In order to help readers visualize the problem, let us use the artificial data shown in Figure 1c again. If we test this data by means of the bivariate t-test, we obtain an $F = 3.667$ with 2 and 21 degrees of freedom. This value means that the null hypothesis can be rejected at the 5% but not at the 1% level. With this result, any linguist would doubt the existence of a phonological distinction between the two populations. But this interpretation would change drastically if it is supposed that the 24 samples shown in Figure 1c have the pairwise relations as in Figure 7. In this latter case it is clear that the two vowels are distinguishable in roughly the same acoustical manner, and in reality the test on the geometrical relation applied to the data of Figure 7 tells us that the Z value obtained (Z = 26.84) is so high that we have to reject the null hypothesis.

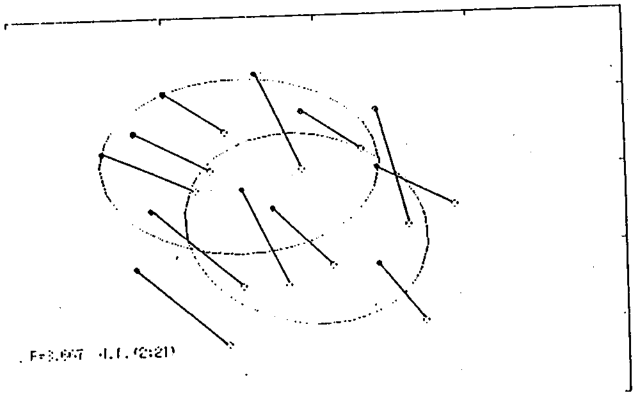


Figure 7. Pairwise representation of Figure 1c

Even though the example used above is based on artificial data, it wouldn't too much miss the point to suppose that nearly the same thing happened in the case of /i/:/e/ and /o/:/u/ in Izumo. In case this assumption is valid, we face with a new possibility that acoustical analysis of the vowels provides us with much more phonologically significant information than we have used previously. With this possibility in mind, the author would like to conclude that the simultaneous application of the two tests discussed in this paper will make a certain contribution to the linguistic study of the nature of vowel merger and its ongoing process.

References

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

Green Paul E.(1978), *Analysing multivariate data.* The Dryden Press, Hinsdale, Illinois.

Labov, W., Yaeger, M., Steiner, R. (1972), *A quantitative study of sound change in progress.* Report on National Science Foundation contract NSF-GS-3287. University of Pennsylvania.

Maekawa, Kikuo (1984), Boin no gôitsu to kondô no riron - Tsugaru, Izumo hôgen wo rei to site - (A theory of vowel confusion: based on the data from two Japanese dialects, Tsugaru and Izumo). *Keiryo Kokugogaku* 14, 4, 149-162.