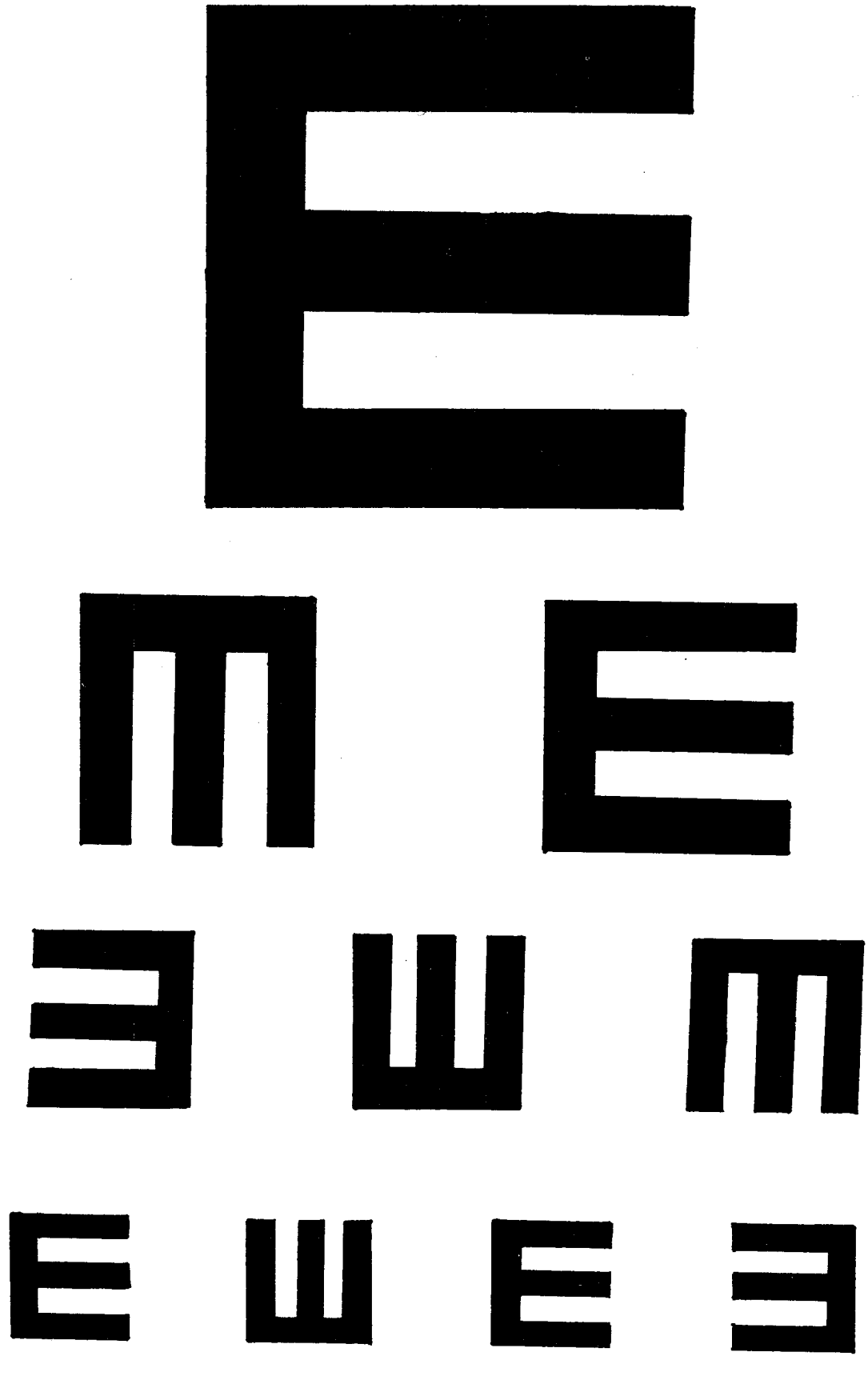


WORKING PAPERS IN PHONETICS

Number 40 UCLA Phonetics Lab April 1978

Vowels in Germanic Languages

Sandra Ferrari Disner

UCLA Working Papers in Phonetics

April 1978

University of California, Los Angeles

ABSTRACT OF THE THESIS

Vowels in Germanic Languages

by

Sandra Ferrari Disner

Master of Science in Linguistics

University of California, Los Angeles, 1978

Professor Peter Ladefoged, Chairman

This work reports on some findings--both methodological and empirical in nature--in the ongoing search for objective methods of describing vowel quality. Many of the major theories and analyses proposed in the current literature are summarized and applied where possible to data sets from different languages. The various approaches are critically evaluated, and a new statistical approach based on analysis of variance is proposed.

The study is divided into four major sections. The first section, which serves as a brief prologue, states the goals of this body of research--specifically, improved means of describing vowel quality--and attempts to illustrate how such findings might be utilized in a 'natural' explanation of a particular phonological process. Section two reviews some of the work that has been done in the past on vowel feature systems and their phonetic correlates. Section three finds that not all languages utilize these features in the same way. A mathematical procedure is

therefore used to test the hypothesis that vowels differ in statistically significant and reliable ways across languages. The advantages and drawbacks to such a procedure are discussed in detail. Section four sets forth two alternative statistical procedures, one based on factor analysis and the other on analysis of variance. The results of each are presented.

"I make this explanation for the reason that without it many readers would suppose that all these characters were trying to talk alike and not succeeding."

Mark Twain
Preface to The Adventures of Huckleberry Finn

It would be of unquestionable value to the linguist to be able to describe vowel quality precisely and systematically. Not only would this aid the linguist in formulating a descriptively adequate model of the sound patterns in a given language -- e.g., those characteristics which can be said to make English sound uniquely like English--but it might also provide insights into how these patterns emerge, and so refine the concept of rule 'naturalness'. Indeed, Chen (1973) suggests that cross-linguistic studies of phonetic detail in language be undertaken "with the view of extracting from the language-specific variations the universal norms governing the rules that operate in individual languages."

The notion of naturalness in phonology seemingly would have to be founded in phonetic facts. Until we have a firmer grip on phonetic reality, however, rule naturalness can be nothing more than an impressionistic recording of what sorts of trends predominate in languages which have been widely studied. It should be our goal, therefore, to determine the phonetic structure of a language reliably and accurately.

Let us consider the well-known phonological process of palatalization in this light. Numerous languages have in their grammars a rule which replaces velar consonants with alveo-palatal or other affricates (e.g. k → č) before front vowels. There is a clear implicational hierarchy of conditioning environments for palatalization rules: palatalization before low vowels implies palatalization before mid vowels implies palatalization before high vowels, but not vice-versa. Furthermore, Hyman (1975) notes that palatalization before both [i] and [e] is less well attested in natural languages than is palatalization before [i] alone; still, palatalization before [i] and [e] does occur "fairly frequently," according to Hyman, and cases of palatalization even before [æ] (as in French) are reported. In order to assess the relative naturalness of a phonological rule in a given language, we must not rely upon statistical frequency alone, but rather should consider the underlying factors that determine naturalness--factors which are to be sought, according to Chen, in the physio-acoustic constants of speech production and perception which are common to all linguistic systems. Let us therefore examine some of the factors which govern palatalization.

Because of the particular musculature of the tongue, there is a component of forward motion in any upward motion of the body of the tongue (Ladefoged 1964; Harshman et al. 1977). Thus, an upward motion of a certain magnitude will automatically push the tongue forward to a certain degree (unless it is pulled back by the action of additional muscles, as in [u]). In running speech, coarticulation effects tend to spread this fronting onto an adjacent consonant (Öhman 1966, 1967; Ohala 1971). The pervasiveness of coarticulation results in the presence of some degree of phonetic palatalization in nearly all languages. Moreover, in some languages this process has been phonologized--the "intrinsic" cues becoming "extrinsic" (Wang and Fillmore 1961)-- as the phonological rule of palatalization.

Ohala (1971) proposes a general model to explain the naturalness of such phonological processes, based largely on Öhman (1967). According to this model, the position of any given point on the tongue in articulating a consonant is determined in large part by the "target position" of that point for the adjacent vowel. All other things being equal, then, phonologization is likely to take place earlier in conditioning environments which are phonetically closer to the output segment. Thus, for example, the higher and farther front the conditioning vowel is, the earlier palatalization is likely to take place. It is important to note that a rule which evolves in this manner is to be considered 'natural' even if its representation in features differs from the prevailing form of the rule in other languages.

As an illustration of this point, let us consider the rules of palatalization in Fe? -Fe? and in Akan. In the former language, as in most languages, the conditioning environment for the rule is [i] alone; in the latter, both [i] and [e] condition palatalization. If we examine Table 1, we find that the [-high] vowel [e] in Akan has a mean F_1 value which is close to the range of the [+high] vowel [i] in Fe?Fe? -- indeed, closer than to its phonological counterpart [e] in the latter language.

	F_1				F_2			
	i	e	ɛ	a	i	e	ɛ	a
Akan	290	320	460		2500	2280	2100	
Fe? Fe? Bamileke	291	368		751	2334	2143		1492

Table 1. Mean formant frequencies of the front vowels of Akan and Fe? -Fe? -Bamileke. (Lindau (1975), Hombert (personal communication)).

Thus, if these figures prove to be reliable indicators of the phonetic vowel quality in Akan and Fe?-Fe?, the configuration of these rules may be considered 'natural' on the basis of the phonetic structure of each language.

Vowel Features

Because speech is a continuum, infinitely divisible over time, and is subject to an infinite number of possible variations due to coarticulation with adjacent (or even anticipated or persevering) portions of the utterance, the domain of physical phonetics is enormous. Yet an adequate linguistic description of any given language can be formulated by translating the physical input into values along a small number of linguistic parameters. These parameters, or features, are part of universal phonetic theory; all languages have sounds and all sounds are measurable in one way or another.[1] However, whether or not a language chooses to contrast sounds along a particular parameter is very much a language-particular fact, and must be made explicit in the phonology of that language. Certain of these parameters, such as Height for vowels or Place of articulation for consonants are widely employed in the languages of the world, while others, such as Pharynx width or Trill ([Rate] in Williamson's (1977) feature system) are relatively rare. All, however, are at once classificatory -- marking contrasts and similarities between sounds -- and phonetically descriptive--telling us something about the actual phonetic quality of these sounds; they further serve to define the phonological notion of natural classes and to highlight the natural sound changes and sound patterns in a language. (Lindau 1975)

There is, however, nothing in linguistic theory that specifically ensures that languages will use similar phonetic values in making corresponding phonological contrasts. Whether or not they do so is subject to empirical verification. Just such an investigation will constitute the body of this paper, along with a look at some of the difficulties inherent in translating physical phonetic data from numerous speakers into rather more abstract linguistic features, and a discussion of some of the phonological implications of the results.

Of particular interest are the phonetic features which determine vowel quality. Unlike the features used to classify consonants and glides, which, according to a consensus of current accounts, are primarily articulatory in nature, vowel features have variously been ascribed to the acoustic or the articulatory domain, or to both. For example, both Chomsky and Halle (1968) and Jakobson, Fant and Halle (1951) suggest that, in principle, simultaneous definitions from articulation, acoustics, and perception ought to be available; nevertheless

---

[1] The actual scales, of course, depend on whose systematic phonetic features are chosen. As Chomsky (1964) points out, there is "room for much discussion as to what is the actual character of the universal phonetic theory."

Chomsky and Halle utilize exclusively articulatory definitions of features, while Jakobson, Fant, and Halle utilize exclusively acoustic ones. Lindau (1975) proposes a feature system with some articulatory features, some acoustic features, and other features for which no single acoustic or articulatory parameter seems to be adequate; for example, Lip rounding and Pharynx width are described in articulatory terms, Rhotacization (Lowered F_3,F_4) and Height and Backness in auditory terms, and Tenseness rather differently as "the degree of centralization on the acoustic chart."

Lindau's revised (1977) set of vowel features takes the articulatory correlates of each feature more closely into account. Only one segmental feature, Peripheral, is described exclusively in acoustic terms, while Height and Backness, which in the earlier version were correlated with the acoustic dimensions F_1 and F_2-F_1, respectively, now are recognized as articulatory features as well. Both features are correlated with the position of the highest point of the tongue, as well as with the values of the formants, leading Lindau to observe that "the traditional highest point of the tongue is virtually as good a measurement of height and backness as the formant chart is." Similarly, Lindau's (1975) feature [lowered F_3,F_4], which is described in auditory terms by Ladefoged (1975) and Stevens and Blumstein (1975), is found also to have as a reliable articulatory correlate a constriction of the pharynx, just above the epiglottis, which appears in r-colored consonants as well as retroflex consonants and r-sounds.

Lindau's most recent set of features may indeed reflect accurately some reliable correlates of vowel quality in natural languages, but it must be kept in mind that reliability per se does not automatically imply linguistic significance for a proposed phonetic feature. An equally plausible model of speech production might propose that articulatory features describe the means to a linguistic end, rather than the end itself. That is to say, speech sounds may indeed be implemented articulatorily, but on a higher level they are specified auditorily.

Some evidence for the latter account is presented by Lindblom, Lubker and Gay (1977), who report on an experiment in which vowels were produced with jaw positions that were both fixed (through the use of bite-blocks) and unconstrained. In both circumstances, the subjects were able to produce vowels of similar auditory quality. That is, when the customary articulatory gestures were, out of necessity, bypassed by the speakers, the alternatives they adopted were nevertheless ones which achieved the same results: namely, "F(ormant) patterns within the ranges of variation of normal vowels." Such evidence might suggest that the speaker's internalized representation of vowel quality draws heavily from the acoustic domain; however, there is some

additional evidence from the Lindblom et al x-ray data which suggests that tongue positions may remain fairly similar to the normal unconstrained positions even when fairly large bite blocks are introduced. If this too is shown to be a reliable effect, it can be taken as evidence in favor of an articulatorily-based set of features. However, neither this nor the preceding evidence can be accorded much linguistic significance until we have further data on the role of the mandible in speech production.

In making a case for the higher-level function of either articulatory or auditory features, advocates of each approach must involve a certain degree of invariance across speakers in the physical phonetic description of speech. In actual fact, there is a great deal of variation present. Joos (1948) points out that the difference in absolute formant values between speakers can be enormous, commonly up to seven semitones for a single vowel uttered by speakers of different ages and sexes. And it has been claimed by Lieberman (1977), largely on the basis of data published by Ladefoged, DeClerk, Lindau and Papçun (1972) that there is a great deal of articulatory variability between speakers.[2] Nearey (1977), in summary, notes that "within-phone variation of the type discussed above appears to be no less problematic in articulatory than in acoustic terms." In the final analysis, then, the true linguistic significance of the phonetic determinants of vowel quality may have to be sought through other means.

For the present it appears to be the case that at least height and backness, the phonetic features which primarily determine vowel quality in most languages, are best described in acoustic terms. These two features have, since the time of D. Jones (1917), been described in auditory terms (despite the articulatory labels accorded to them by Jones). In Lindau's (1977) data, "the correlations between auditory measurements and the acoustic measurements are fractionally higher than the correlations between the auditory measurements and the articulatory measurements" of height and backness. Moreover, the very high correlation between auditory and articulatory measurements of height and backness which she reports does not seem

---

[2] Ladefoged (personal communication) has pointed out that he doubts that the data in Ladefoged et al (1972) cited by Lieberman does show articulatory variability between speakers. He considers that while it was shown that different speakers may use different means to achieve similar articulatory shapes (some, for example, moving the tongue within a comparatively fixed jaw, and others making more use of jaw movements), there is very little evidence that different speakers have quantifiably different shapes of the vocal tract when producing similar sounds. He notes that the Ladefoged et al (1972) paper has appeared only as a working paper and has not been submitted for publication precisely because it does not provide real evidence of this kind.

to carry over to certain other data sets (cf. Ladefoged 1967, 1975; Jacobson 1978). It will therefore be taken as at least a working hypothesis that the phonetic quality of a vowel can be expressed in terms of its acoustic characteristics.

The adoption of a set of acoustic features to describe phonetic vowel quality in natural languages is hardly a new concept. For over a century vowels have been analyzed according to their formant frequencies, and more recently with respect to the relationships which hold between these frequencies. The fact that F_1 and F_2 are "the principal determinants of vowel quality" (Joos 1948) is evidenced by the primacy in natural languages of the features of height and backness, which, as we have argued, are most directly correlated with the acoustic parameters F_1 and F_2 (actually F_2-F_1, to better approximate the auditory properties of backness). Lindau (1975) states that all known languages contrast high and low vowels, and all but a very few contrast front and back vowels as well.

Additional acoustic features, notably F_0 and F_3, have been suggested as determinants of vowel quality. F_0 may be considered to be a suprasegmental property, and as such it lies beyond the scope of this paper; furthermore, its effects on vowel quality do not appear to play a distinctive role in any natural language (Maddieson 1977). F_3, we find, plays something of a dual role in determining vowel quality. In one sense, it serves to modify the basic F_1 & F_2 pattern, as the primary acoustic cue (along with F_4) for retroflex and rhotacized vowels. It is also an important, if auxiliary, cue to rounding, particularly in signaling non-standard values of rounding, as in front rounded or back unrounded vowels.[3] In light of this, it may be useful to plot the F_3 values of the vowels along a separate, third dimension in acoustic space. However, Broad and Wakita (1977) have shown that when front and back vowels are considered separately, F_3 is fairly predictable from F_1 and F_2. There is further compelling evidence for integrating F_3 into the basic F_1 & F_2 pattern, rather than regarding it as a separate dimension, for it also tends to replace an unusually weak F_2 as the acoustic feature which is most closely correlated with the phonetic quality of high front vowels. Fant and others (Fant 1959, Nordström and Lindblom 1975) have suggested various means of integrating the F_3

---

[3] Eli Fischer-Jørgensen (personal communication) suggests that F_3 is the principal cue for front rounded vowels in Danish. The situation is not as clear for back unrounded vowel in other languages, as F3 is much weaker in these regions of vowel space.

values into an "effective second formant" in order to take into account "a gradual increase in the importance of the third formant as F_2 is raised in frequency." (Fant 1959)

Vowels in Natural Languages

Given a universal set of acoustically-based phonetic features, we might wish to study the ways in which these are utilized in the vowels of natural languages.

It has long been assumed that the phonetic quality of vowels is most insightfully expressed in terms of points along a continuum, with the latter defined according to a presumably universal set of phonetic features. In transcribing a vowel, the phonetician takes into account these same features and--after discounting such individual characteristics as head size and mean fundamental frequency, as well as certain contextual effects--assigns the vowel a phonetic symbol to identify it at the systematic phonetic level; the vowel is then plotted along the phonetic continuum. What is more interesting from a theoretical standpoint, however, is the fact that the phonetician similarly is able to assign a symbol and a single point on the phonetic chart to the representation of the overall quality of a vowel allophone in a language--that is, to the phonetic quality of the vowel common to all speakers of that language. (See Fig. 1) This suggests that in some real sense there is a level of representation, comprised of all but the speaker-particular aspects of the speech signal, in which each vowel is effectively a point in phonetic space. In fact, it is not unreasonable to postulate a systematic phonetic level with a certain degree of invariance of quality, in view of the high intelligibility of vowels across many speakers of the same language. Accordingly, the physical phonetic signals from various speakers may be viewed as mapping onto a unitary set of phonetic targets (i.e., the vowels), at least within the domain of a single language. It thus may be said that, assuming an optimal normalization, all vowels transcribed with the same phonetic symbol in the same language will occupy the same point in phonetic space. The position of the point corresponds to the quality of the vowel in that particular language.

It is not to be assumed in consequence, however, that vowels of different languages which appear to have the same phonetic quality and which are sometimes transcribed with the same phonetic symbol are necessarily the same. Two languages might well choose different combinations of phonetic features in defining the position of, say, the vowel [e] in phonetic space. Thus, the vowel transcribed as [e] in one language may have a phonetic quality that is perhaps higher or more fronted or more tense than the vowel [e] in another language.

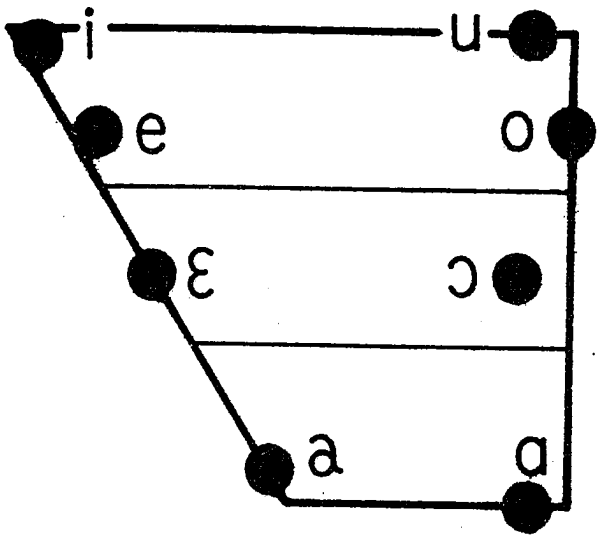


Figure 1. An auditory representation of some Danish vowels, transcribed by Uldall (1933).

In fact, the auditory impressions reported by trained phoneticians seem to lend support to the notion outlined above. Language-instruction texts and phonetic outlines report with considerable regularity such observations as:

Swedish [i] is "closer than the vowel in English seen" [4]

Danish [i] is "tenser than English [i] as in see"[5]

Norwegian [o] is "similar to Swedish [o], but rather less tense"[6]

Dutch [ɛ] is "more open than the English vowel in bed; intermediate between set and sat"[7]

Danish [ø] is "a little more close than Swedish [ø]"[8]

Norwegian [ø] "resembles German [ø], but is also less rounded"[9]

"All German vowels are tenser than their English counterparts."[10]

In light of phoneticians' impressions, then, it may be hypothesized that at least some real differences hold between similarly-transcribed vowels in different languages. In order to prove this hypothesis, however, it must be shown that such vowels differ in statistically significant and reliable ways. If such reliable differences were found, we might profitably chart the vowels in phonetic space; in doing so we would gain at least a first impression of where in the vowel space these differences occur, as well as the magnitude of these differences. This in turn would allow us to speculate on the phonological implications of such phonetic differences.

In order to test the proposed hypothesis, we must rely upon physical phonetic data from a number of speakers of various languages. However, we cannot determine the phonetic quality of a vowel directly from its formant frequency values. As we have observed, there is a great deal of phonetic variability present in the acoustic signal, and the formant frequencies which convey the

---

4. McClean (1969) p. 5
5. Walshe (1965) p. 90
6. Walshe (1965) p. 101
7. Koolhoven (1968) p. 5
8. Nielsen and Hjorth (1971) p. 14
9. Haugen (1935) p. 12
10. Moulton (1962) p. 58

quality of the vowel [æ] for one speaker may variously convey the qualities of [ɛ,ɐ,ʌ,ɑ] for other speakers (cf. Peterson and Barney 1952); by the same token, a single phonetic vowel quality may be conveyed by a wide range of formant values depending on the speaker. For example, in their study of American English vowels, Peterson and Barney (1952) observed F_1 and F_2 values for the vowel [ɪ] ranging from 280 and 2000 Hz. for one speaker to 680 and 3250 Hz. for another; Kahn (1977) replicated this study for male speakers only, with more rigorous dialect control, and still observed a considerable amount of variation, ranging from 360 and 1880 Hz. to 460 and 2125 Hz. for the vowel [ɪ].

Some of the acoustic variability in natural speech is contextually determined, and this can be minimized by embedding the vowels in a neutral frame (Kahn uses #h_#). Most of the variability, however, is determined by the physiological differences between the individual speakers of a language, and these differences have numerous manifestations. Speakers typically vary in vocal tract length (15% shorter, on the average, in women than in men (Chiba and Kajiyama 1941)), in cavity size and the ratio of pharyngeal to oral cavity (both greater in males), and in a number of more subtle configurational aspects of the vocal tract as well. Moreover, the work of Cleveland (1975) and Papçun (personal communication) has shown that a number of these physical characteristics are more properly represented as a continuum--from basso through tenor to soprano, for example--than as a binary opposition between males and females or adults and children. The number of possible gradations of personal quality is thus even greater. It is therefore unlikely that even a very few speakers of a language will produce--or will even be able to produce -- the same phonetic vowel in a measurably identical way.

Such variability on a physical phonetic level makes it difficult for the linguist to specify the phonetic quality of a vowel on the basis of its acoustic characteristics alone. Indeed, the determination of which acoustic characteristics are the relevant ones, and of the precise relationship which holds between these characteristics and the phonetic quality of a given vowel, has been and still remains the target of considerable linguistic research. Yet even untrained listeners are able -- in ways which as yet are not fully clear to the linguist -- to compensate for such differences, thus extracting the phonetic quality of the vowel from among the many speaker-particular aspects of the speech signal. To date, no mechanical procedure has been proposed which is able to classify vowels on the basis of their acoustic structure with comparable accuracy.[11]

---

11 Gerstman's (1968) normalization procedure, described below, has been shown by Nearey (1976) to give results which are

To gain an accurate idea of the acoustic variability inherent in speech, it is useful to plot the vowels of a number of different speakers of a language in a formant space with F_1 and F_2-F_1 as the axes. Such charts reveal that a representative sample of the speakers of a language will tend to produce a given vowel with formant values which are scattered throughout a certain range of acoustic space, consistent with the variance of the sample (but subject to certain universal constraints, such as the greater mean variance of low vowels). As we have observed, this variance makes cross-linguistic comparisons particularly problematic. Figure 2 illustrates this point; it shows a plot of several vowels as spoken by six randomly-selected speakers of German (indicated by points) and seven of Swedish (indicated by crosses). One cannot readily determine from this chart exactly how a given vowel in German differs from the corresponding vowel in Swedish, if at all; the speaker-related differences tend to obscure whatever overall (phonetic) differences actually obtain between the two languages. Two possible ways around this problem will be considered in this paper. One, the most frequently used in vowel studies, is to apply any one of a class of algorithms known as normalization procedures, to the data in order to remove an appreciable portion of the speaker-related variance for each vowel. These algorithms all proceed from a basic assumption of systematicity in the overall variance, and all perform a (linear or non-linear) rescaling of the data, which is then plotted on the same scale. To the extent that the normalization procedure is successful in reducing the scatter, differences in phonetic quality between the vowels can be "read off" the acoustic chart. However, there are drawbacks of a practical as well as a theoretical nature to all of the normalization procedures which have been proposed thus far, particularly with respect to their use in cross-linguistic studies. This matter will be discussed in detail below.

As an alternative approach,it is possible to tease out the overall differences by using statistical procedures such as analysis of variance or t-tests. In essence, both of these procedures compare the amount of difference between the languages to the amount of speaker-related difference within each language. If the former exceeds the latter, the lan-

---

comparable to Kahn's (1977) recognition study. ("over 97% correct" --i.e. non-overlapping--for the former; 97.1% correct identification for the latter). However, Kahn points out that much of the 2.9% error rate in his study is attributable to residual dialect heterogeneity. Furthermore, there are disadvantages of a theoretical nature associated with the Gerstman procedure which will be discussed in more detail below.

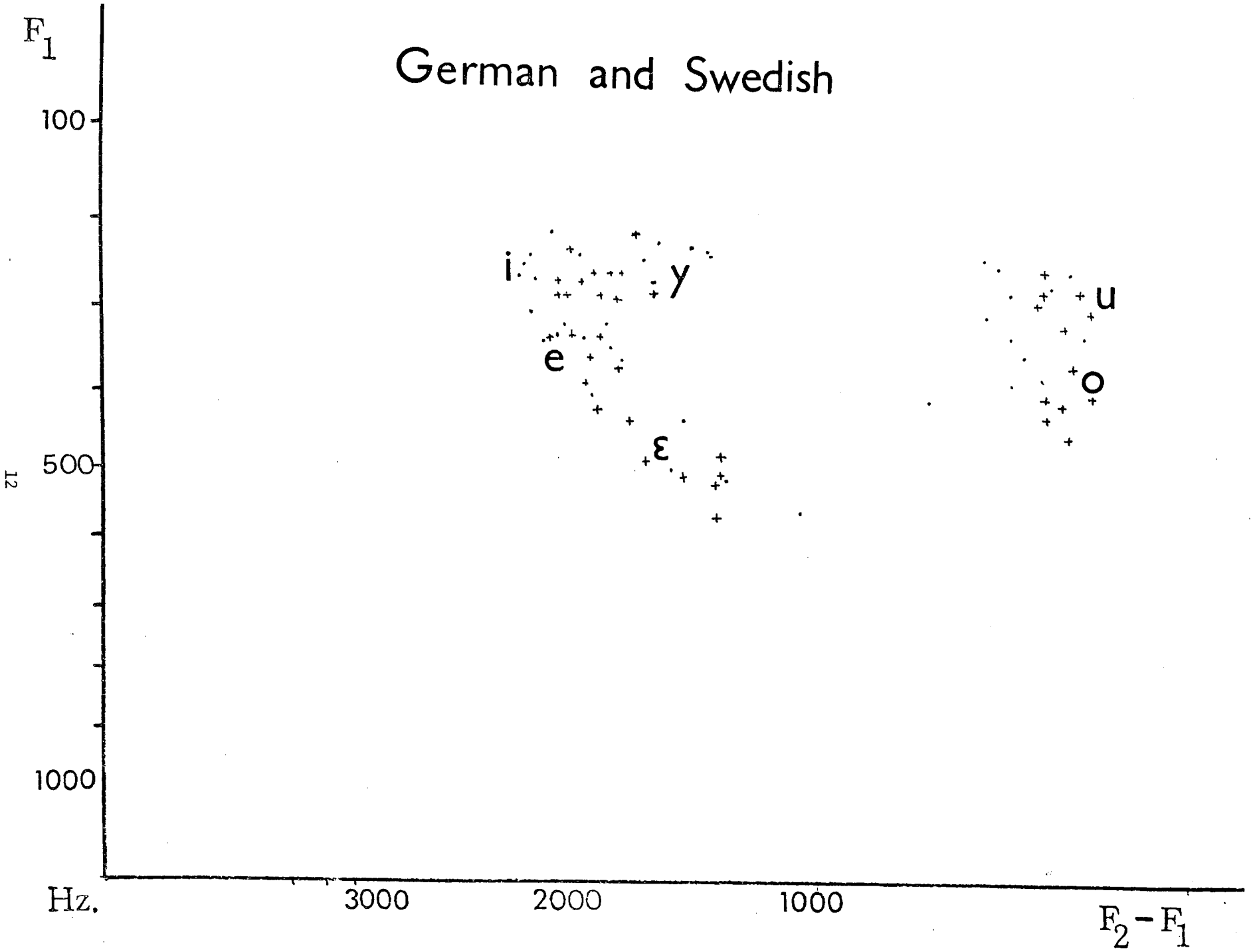

German and Swedish
F1
100
500
1000
Hz.
3000
2000
1000
F2 − F1
i
y
e
ɛ
u
o

Figure 2

guages can be said to differ reliably from each other along the parameters under consideration.

Whichever the approach, we should take care in defining the scope of our investigation if we are to conduct our search for reliable phonetic differences most profitably. If we were to find that vowels in all languages had similar phonetic quality, any language would be equally able to provide useful insights into the nature of phonetic features and how they participate in the determination of vowel quality. If, on the other hand, reliable differences were found to obtain between the phonetic qualities of comparable vowels in two genetically, typologically, and areally unrelated languages such as English and Yoruba, separate investigations would have to be carried out for different language families; still, it would not be clear whether or not languages belonging to a single family could be considered as one in this respect. The more conservative approach, and the one which makes the strongest case for reliable differences between languages in general, would be to examine related languages such as English and German. If two related languages were to show reliable differences in vowel quality, potential differences would be generalizable to languages such as Yoruba and Thai as well.

As a final note, the importance of obtaining a truly representative sample of the speakers of a language cannot be overemphasized. Systematic biases may arise when speakers with extreme vocal tract sizes are included in the sample, particularly when the sample is fairly small. These biases are problematic both for cross-linguistic studies and for studies within a single language.

Speakers with relatively large resonating cavities, for example, will tend to have lower formants than other speakers; when plotted, their vowels appear to be higher and more back than expected. A sample which includes a greater-than-average number of such speakers is thus unsuitable for use in cross-linguistic studies, where differences in absolute formant frequency between vowels assume theoretical importance. Care should be taken to ensure that the differences between languages are truly phonetic, rather than due to the personal qualities of the particular speakers.

Moreover, the skewing of the data which arises from such non-representative sampling of the population might not be distributed evenly throughout the vowel space. A speaker with very low formants may have his low and mid vowels (those with higher F_1) displaced to a relatively large degree from the corresponding vowels of an average speaker,

but his high vowels (low F_1) displaced proportionately less. This is to be expected, in light of the fact that [i] and [u] lie closest to the uppermost boundary of phonetic space (lowest F_1) and are therefore less likely to have their means affected by extremely low F_1 values than by extremely high ones.

Figure 3 illustrates this point. Ladefoged has said that he himself has "larger resonating cavities, and hence lower formant frequencies, than the average male speaker." (Ladefoged 1975), and this is reflected overall in the chart by comparing his vowels (dashed line) to the vowels of an average sample of American males (Peterson and Barney, 1952) (solid line). Yet upon closer examination it can be seen that, at least with respect to height, the two sets of vowels differ a good deal less in their high vowels [i] [u] than in any of the others. Ladefoged rules out any explanation in terms of dialect for differences of this magnitude. Rather, the asymmetries appear to be due to anatomical factors.

The acoustic effects of varying vocal tract sizes can be demonstrated by means of a line analogue speech synthesizer (Rice 1971). Schematic vocal tract shapes approximating the vowels [i u e o ɛ ɔ ə] were generated, and for each vowel the length of the vocal tract was varied systematically from 14.4 cm to 18.0 cm.

For each vowel, formant frequency values corresponding to each increment in tract size were calculated, and then plotted in F_1 and (F_2-F_1) space. (See Figure 4) As is evidenced by the chart, a set of idealized vocal tract sizes will produce a set of formant values with the least F_1 slope in the range of the high vowels, and progressively greater slope for the lower vowels. This lends support to an anatomical interpretation of the difference between Ladefoged and the Peterson and Barney data above.

We may thus conclude that averaging over a physically non-representative sample of the speakers will not only tend to produce vowel points which are systematically displaced from their true phonetic values, but it will also tend to introduce asymmetries into the vowel system. The latter is as problematic for within-language studies as it is for those across languages. Still it is possible, at least in principle, that speakers adjust their articulations to compensate for this -- i.e., to make their F_1 and F_2 values equal to those of other speakers, regardless of where their tongue has to go to do this.

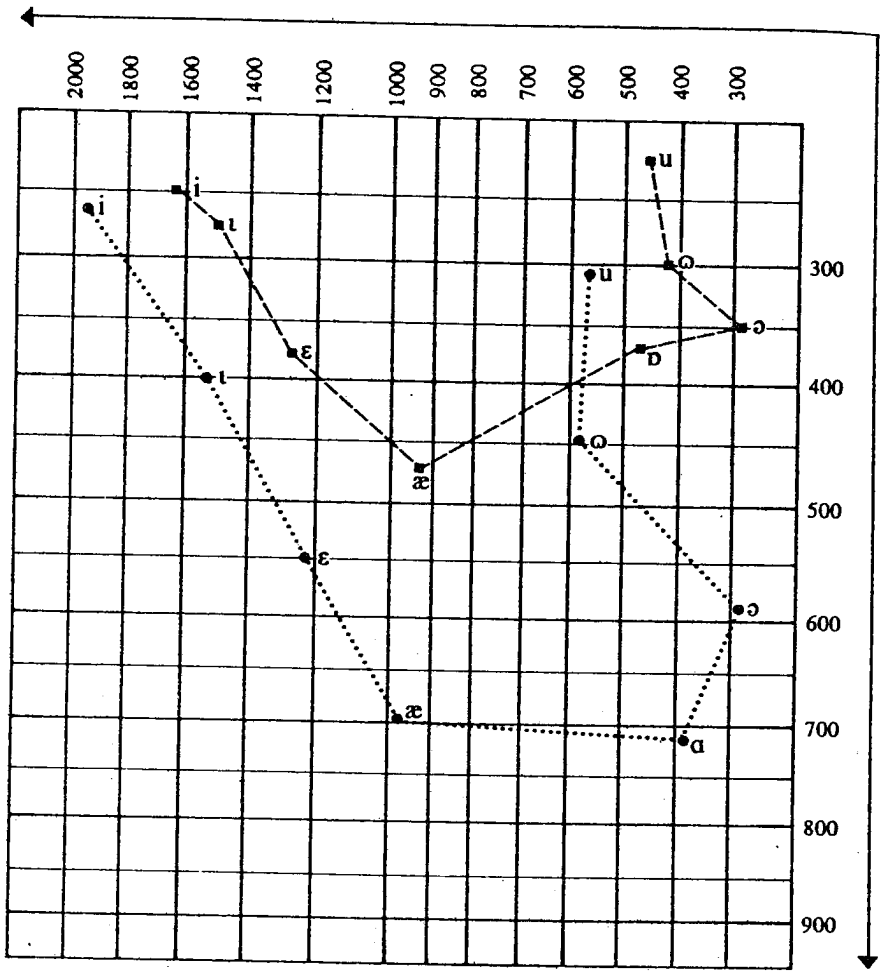


Fig. 3 Plotted formant frequency values of eight English vowels as spoken by a speaker with large resonating cavities (dashed line) and a mean sample of adult male speakers (dotted line).

(from Ladefoged, 1975)

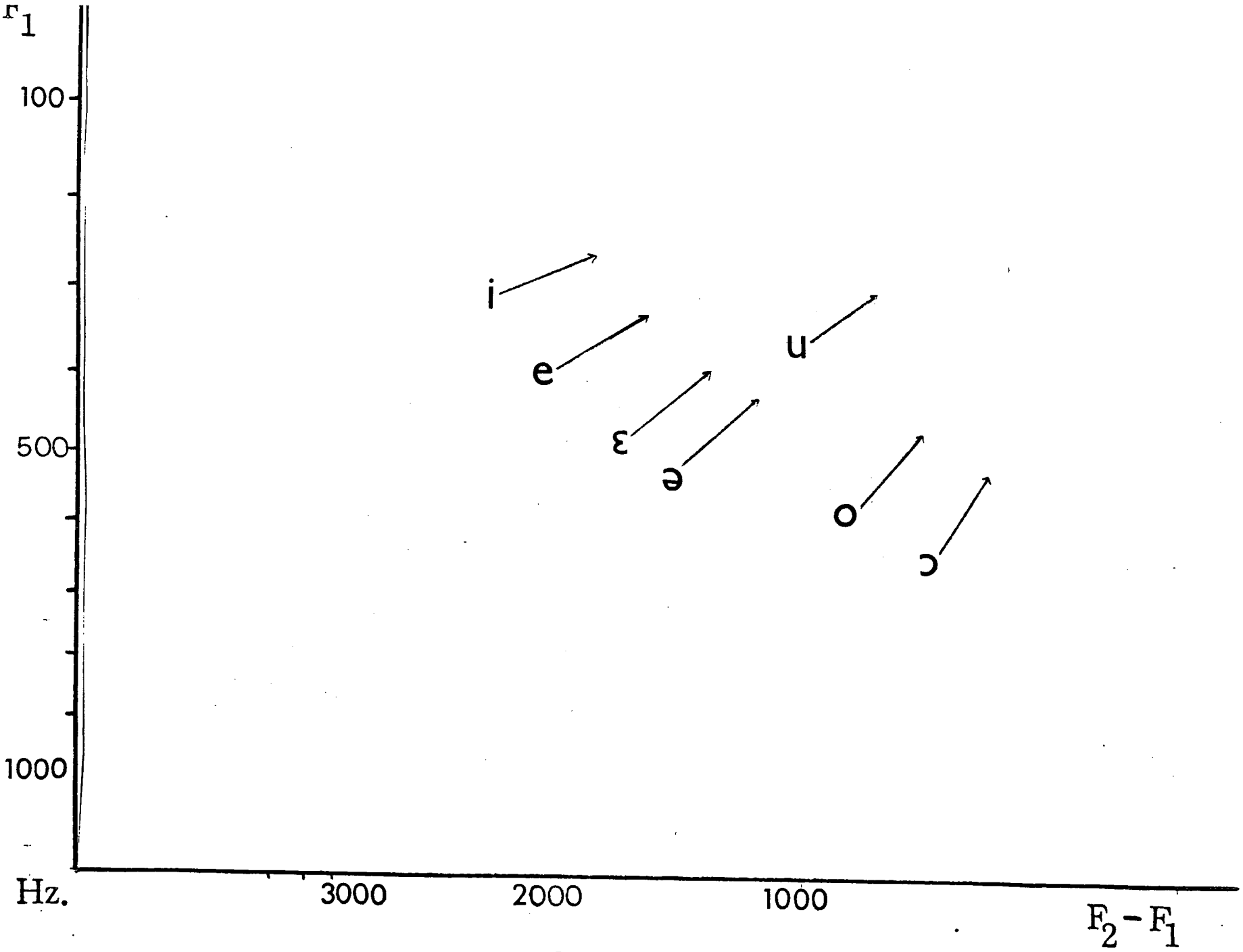


Figure 4

The present study utilizes data from male speakers only, as some of the data sets which were analyzed spectrographically were limited to the vowels of male speakers[12]; thus, in order to avoid cross-linguistic bias, all data from female speakers was excluded from the remaining data sets. However, as we have seen, a subtle bias of a different sort is introduced when the data is limited to male speakers only. Ideally, studies should be based on large samples of speakers, both male and female.

Procedure

Data.

In order to test for reliable differences in vowel quality, a cross-linguistic study was undertaken. The languages chosen were six Germanic languages: Dutch, English, German, Norwegian, Swedish, and Danish. The English data was further broken down into 'Standard American English' and the dialect spoken in Southern California. Formant frequency data (F_1, F_2, and F_3 only) for the vowels of at least six male speakers of each language was selected from published accounts; all the measurements had been made of vowels in similar contexts, except for the Swedish data, in which the vowels had been pronounced in isolation.

To begin with the most extensive data set, formant frequency values for the 12 monophthongal vowels of Dutch, as spoken by 50 males with standard pronunciation, are presented by Pols, Tromp, and Plomp (1973). These vowels are [i ɪ e ɛ a ɑ ɔ o u ʏ ø œ].

The vowels were pronounced in the context hVt, forming monosyllabic Dutch words. For each vowel, a steady-state portion of the waveform was selected out and then analyzed by means of a wave analyzer.

Peterson and Barney (1952) present a formant analysis of ten American English vowels [i ɪ ɛ æ ʌ ɑ ɔ ʊ u ɚ] (diphthongs excluded) spoken by (among others) 33 adult male subjects. All the English vowels in this study were pronounced in the context hVd, forming the words heed, hid, head, had, hod, hawed, hood, who'd, Hud, heard. Measurements were made with a sound spectrograph.

A criticism that has been raised (Nordström and Lindblom,

---

12 Spectrographic analysis is more difficult to perform on female voices, as the harmonics are typically a good deal further apart than are those of male voices, by virtue of the higher typical F_0.

1975) with regard to the Peterson and Barney data is the relatively low degree of dialect control among the speakers involved in the study. The speakers represented a broad regional sampling of the U.S. and a few even happened to speak English only as a second language. Kahn (1977) recorded a similar set of vowels that were more rigidly controlled for dialect. His subjects were 7 male college students born in Southern California; the corpus was similar to that recorded by Peterson and Barney, with the exception of hawed [hɔd], as [ɔ] tends to be merged with [ɑ] among Californians. The California dialect data also included the initial vocalic portions of hayed and hoed ([e] and [o]). Measurements of a steady-state portion of each vowel were made using a linear prediction technique.

The nine long vowels of Swedish [u o ɑ ɛ e i y ʉ ø] as pronounced in isolation by 24 male college students, have been analyzed by Fant, Henningsson, and Stålhammar (1969). Formant frequency measurements were made by means of a sound spectrograph at a point near the vowel onset.

The nine long vowels of Norwegian [i e æ ɑ ɔ u ʉ y ø] as spoken by 10 male subjects were analyzed in the context C V: C (+dental +dental) by Gamnes (1965). Presumably due to limitations inherent in the formant-extraction technique used by Gamnes, F_2 and F_3 for the vowels [u] and [ɔ] were not generally available. Only in the case of speakers with higher formant values (2 females and 7 children) were the F_2 values resolvable.

The ten long vowel phonemes of Danish [i e ɛ a ɔ o u y ø œ] in addition to [ɑ], the allophone of /a/ in context before /r/[13] were studied by Fischer-Jørgensen (1972). Seven male speakers of standard Danish were included in the study, of whom five reflected Copenhagen standard pronunciation while the remaining two retained a few characteristics, mainly intonational, of other, previously-learned dialects. Formant frequencies were measured by means of a sound spectrograph. The vowels were pronounced in VC or hVC contexts, with the consonants chosen to minimize the effects of the formant transitions. Specifically, front vowels and low back vowels, which have relatively high F_2 values, were followed by dental consonants (comparatively high locus of F_2), while back non-low vowels, which have relatively low F_2 values, were followed by labial consonants (comparatively low locus of F_2).

---

13 Other vocalic allophones in pre-r position showed less variability.

The German data—formant frequency measurements for the eight long, monophthongal vowels [i e ɛ y ø ɑ o u] - is the product of two separate investigations, one conducted by Eli Fischer-Jørgensen and the other by Hans Peter Jørgensen; both investigations are reported in Jørgensen (1969). Three of the six speakers, (those studied by Fischer-Jørgensen) pronounced the vowels in various lexical contexts, with a mean of nine tokens per vowel. The remaining three (studied by Jørgensen) pronounced the same vowels in a more rigorously controlled set of test words; these were all disyllables of the form $\left\{\begin{matrix} h \\ \# \end{matrix}\right\}$VCx. As with the Danish data, the consonantal contexts were varied in order to minimize the formant transition effects. All vowels were then pronounced once again in contexts before velar consonants. The vowels in both studies were analyzed by means of a sound spectrograph,and average values were reported.

In order to ascertain whether these two methods (broad or constrained set of contexts) are sufficiently comparable for purposes of the present investigation, a series of F-tests was conducted. Out of the 16 formant values tested (F_1 and F_2 for eight vowels), the F-scores for 13 ranged from 0 to 1.90 ($p > .25$); these showed the two methods to be similar to a statistically significant degree. F-scores for the remaining three (F_1 for [a], F_2 for [u ,y]) had probability levels at or near $p > .10$. None of the values approached probability levels which would show them to be significantly different ($p < .05$). Accordingly, both sets of German data were used in the present study.

Formant frequency plots.

All the vowels of each language were plotted in a formant space with F_1 along the vertical axis and the differences between F_2 and F_1 along the horizontal axis. Distances along the axes were made proportional to the mel scale in order to approximate better the perceived distances in phonetic space. The areas occupied by each vowel were then delimited by means of an ellipse program described by Davis (1977). For each cluster of vowel points, an ellipse with radii of two standard deviations was drawn along axes oriented along the principal components. Thus, the ellipses enclose approximately 95% of the population along each axis.

Results

Figure 5 (ellipses alone; actual vowel points omitted) shows the relative placement of vowels transcribed as [e] in the six different languages.

It seems evident at first glance that there is a difference among at least some of the languages with respect to their placement in acoustic space of this particular vowel. This and similar charts for other vowels may be taken as evidence

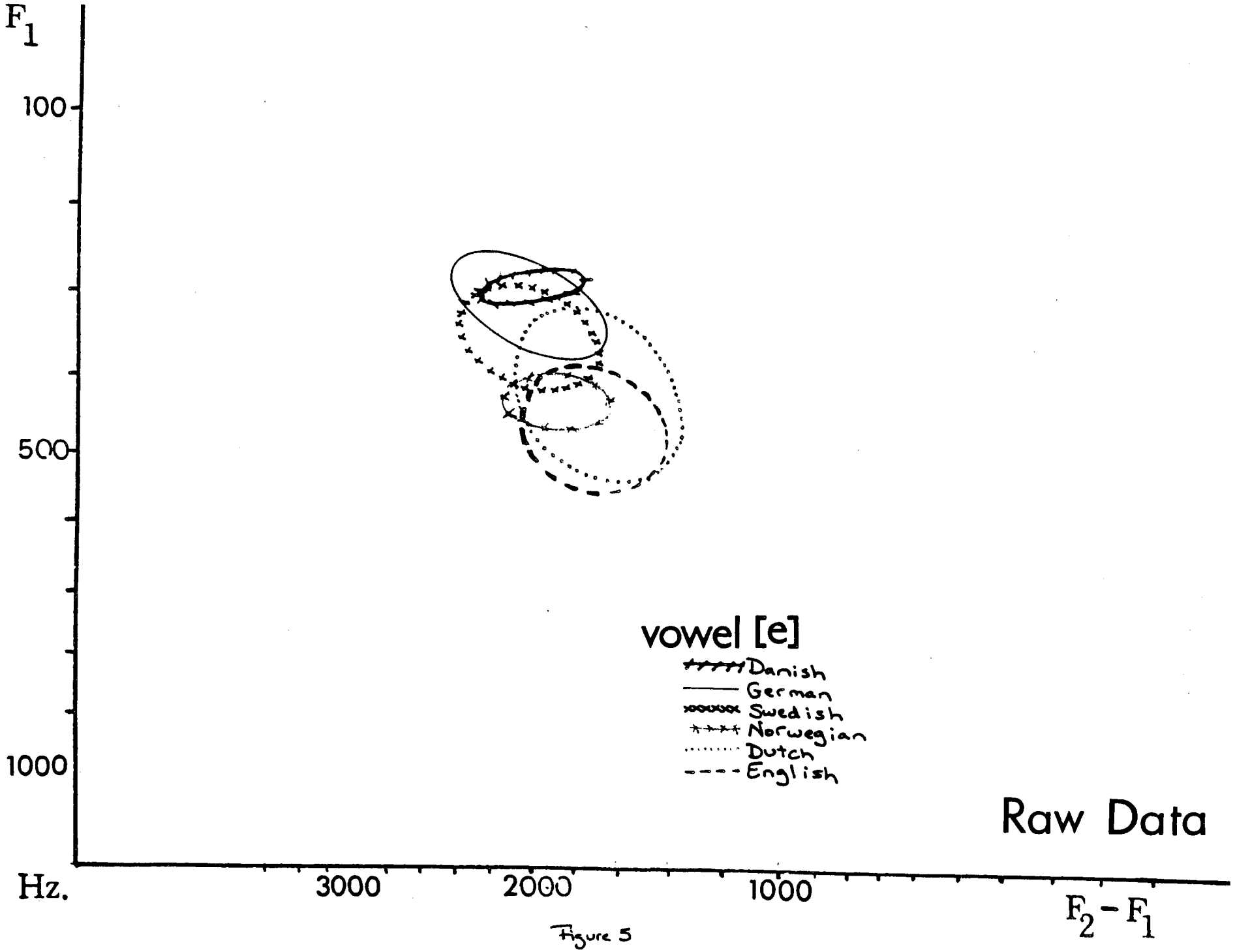

F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
vowel [e]
Danish
German
Swedish
Norwegian
Dutch
English
Raw Data

Figure 5

that vowels transcribed with the same symbol do not necessarily have identical phonetic quality. Rather, there seem to be at least some real differences between "corresponding" vowels in different languages, confirming phoneticians' judgments of the sort cited above (p. 7). The plotted raw values alone--without the benefit of any sort of normalization--reveal the [e] sounds of at least some languages to be distinct populations. The hypothesis propounded above--that differences exist and that they are significant and reliable--is thus confirmed. It is quite evident, for example, that the Danish [e] is closer than the corresponding English vowel [e]. This is exactly as suggested by Bredsdorff (1958) and Uldall (1933).

Similarly, but less strikingly, the vowel symbolized as [e] in Danish is closer than Swedish [e], as is noted by Nielsen and Hjorth (1971). Koolhoven (1968) notes in addition that Dutch [e] is closer than English [e] as in say. This last case is not so readily verifiable, in view of the rather wide variance present in the data sets of both languages. Speaker-related differences within the languages here tend to obscure whatever overall differences may obtain between English and Dutch.

Discussion

We may at this point wish to go beyond the original hypothesis and--having determined that there are some differences between languages -- attempt to state where these differences occur and how great they are. Before we can do so, we must be able to specify as precisely as possible the phonetic quality of each vowel. As we have mentioned, the most common approach to this problem has been to use one of a number of different normalization procedures; these will be described below. To our knowledge, however, there has been no attempt in the literature to evaluate the different procedures on the basis of how well they succeed in removing the speaker-particular variance inherent in any given data set. Moreover, there has been no consideration of the fact that, while a particular normalization procedure may be successful in reducing the variance within one particular language, it may be less successful -- even counterintuitive -- for other languages, and hence will be relatively unsuitable for cross-linguistic studies such as the present one.[14] Our present investigation

[14] The converse of this situation may be of interest to the linguist working within the confines of a single language. Specifically, a normalization which is less than optimal from a cross-linguistic viewpoint may nevertheless be the most successful in removing the variance from the data of a particular language. In certain cases this fact is directly related to the phonological structure of the languge (e.g., to

of the differences which obtain between languages will thus have the added benefit of providing a heuristic for choosing normalization procedures which best suit the data at hand and the scope of the study.

Normalization

It is plausible to assume that, in making a perceptual judgment of a vowel in his language, the listener takes as input the acoustic signal and--with the appropriate compensation for the speaker's individual characteristics--maps it onto a specific point in phonetic space. The trained phonetician can be said to have learned, or developed, a strategy which generalizes this mapping process; he or she can, with a good deal of reliability, remove the speaker-particular aspects of any speech signal and plot it directly on a standard chart of vowel qualities.

It should be possible, at least in theory, to approximate the phonetician's skill by means of a mathematical transform which is applied to the raw data. The value of such a function may be measured in terms of its success in reducing the variance in a data set. It would be especially valuable if it reduced the variance between vowels that could be classified as being phonetically the same while maintaining the variance between vowels that were classified as different. An algorithm of this sort may be judged <u>adequate</u> if it succeeds in reducing the variance in the plotted values of a set of vowels without distorting the relationships which hold between them. Figure 6, for example, shows a plot of the regions in formant space occupied by Peterson and Barney's (untransformed) data from 33 American male speakers. Note the fairly large degree of scatter, and the areas of ambiguity (shaded) in which phonetically distinct vowels were produced with similar sets of first and second formant frequencies.[15]

---

the presence or absence of asymmetries such as a set of front rounded vowels, unmatched by back unrounded ones) and thus the "better" normalization may to some extent be determined offhand. Additional research is necessary, however, to advance our understanding of how the phonology of a language interacts with the mechanics of normalization.

15 It is not crucial to the understanding of spoken vowels that these areas of ambiguity be eliminated, as factors other than the formant values (e.g. length, diphthongization, etc.) may be the cues utilized by the speaker in resolving the ambiguity. Even a relatively successful normalization procedure may allow some overlap if all it has to do is provide a partial basis for classifying vowels into phonemes.

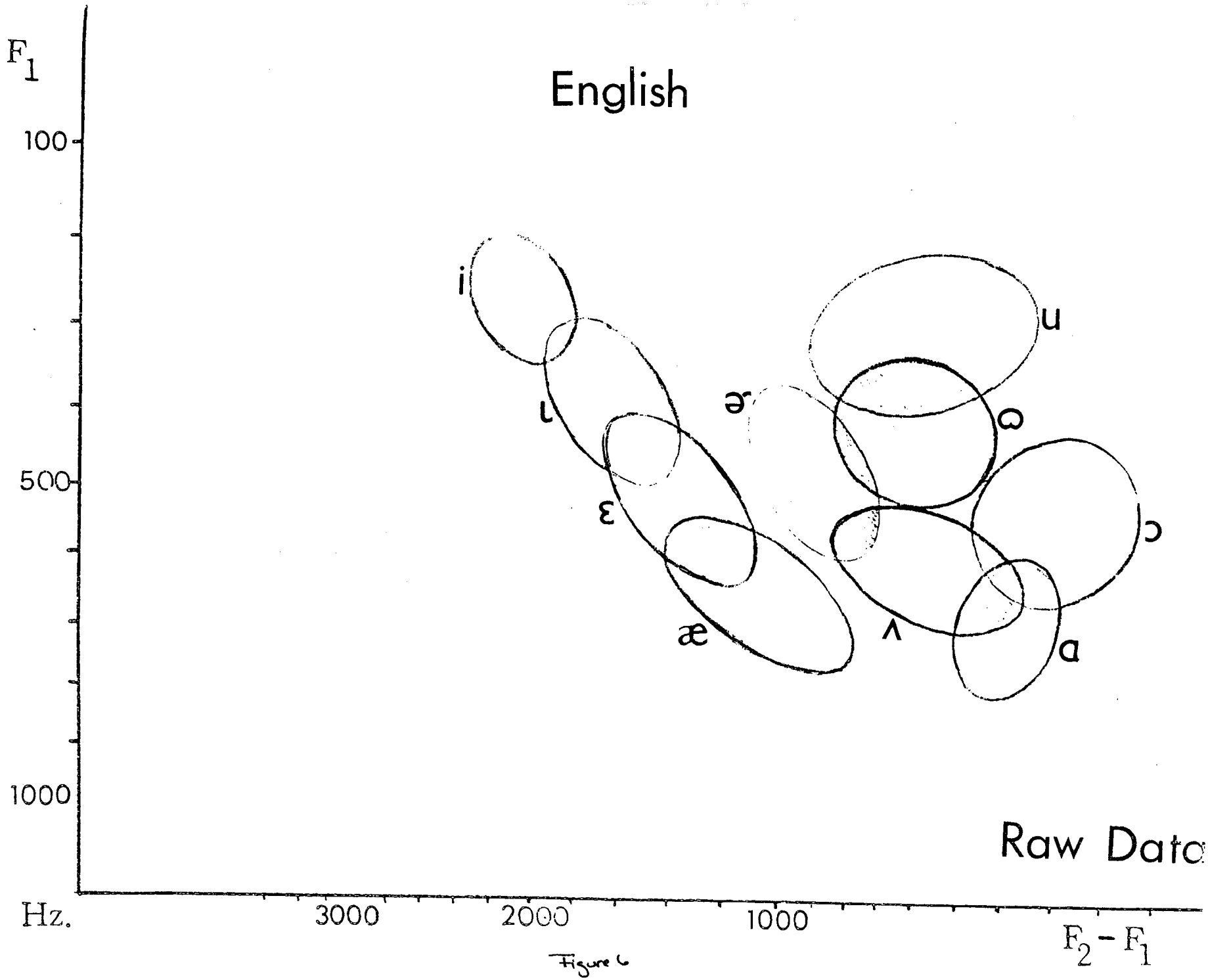

English
F_1
100
500
1000
Hz.
3000
2000
1000
$F_2 - F_1$
i
ɪ
ɛ
æ
ə
ʌ
u
ʊ
ɔ
ɑ
Raw Data

Figure 6

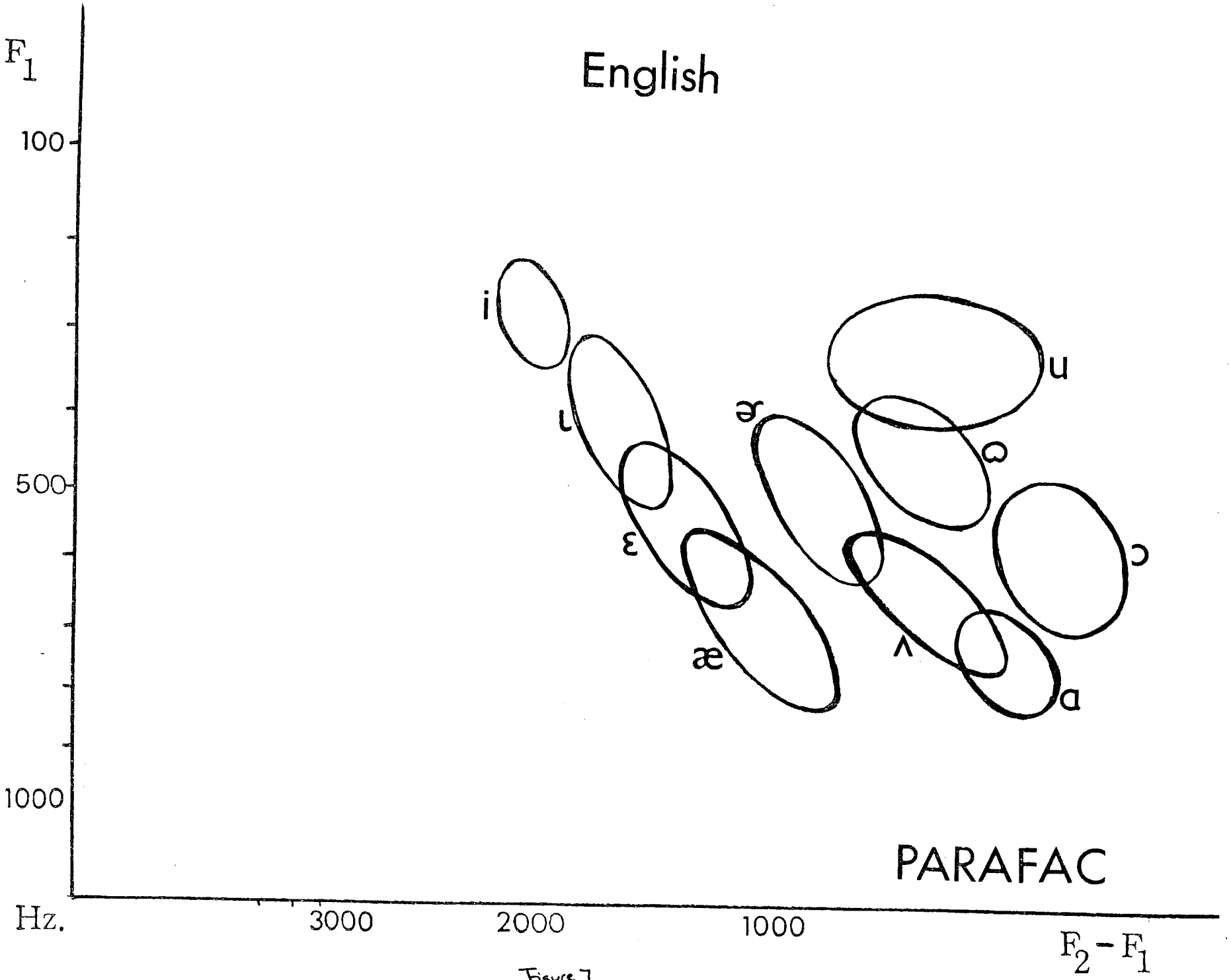


Figure 7

F. I English

Figure 1

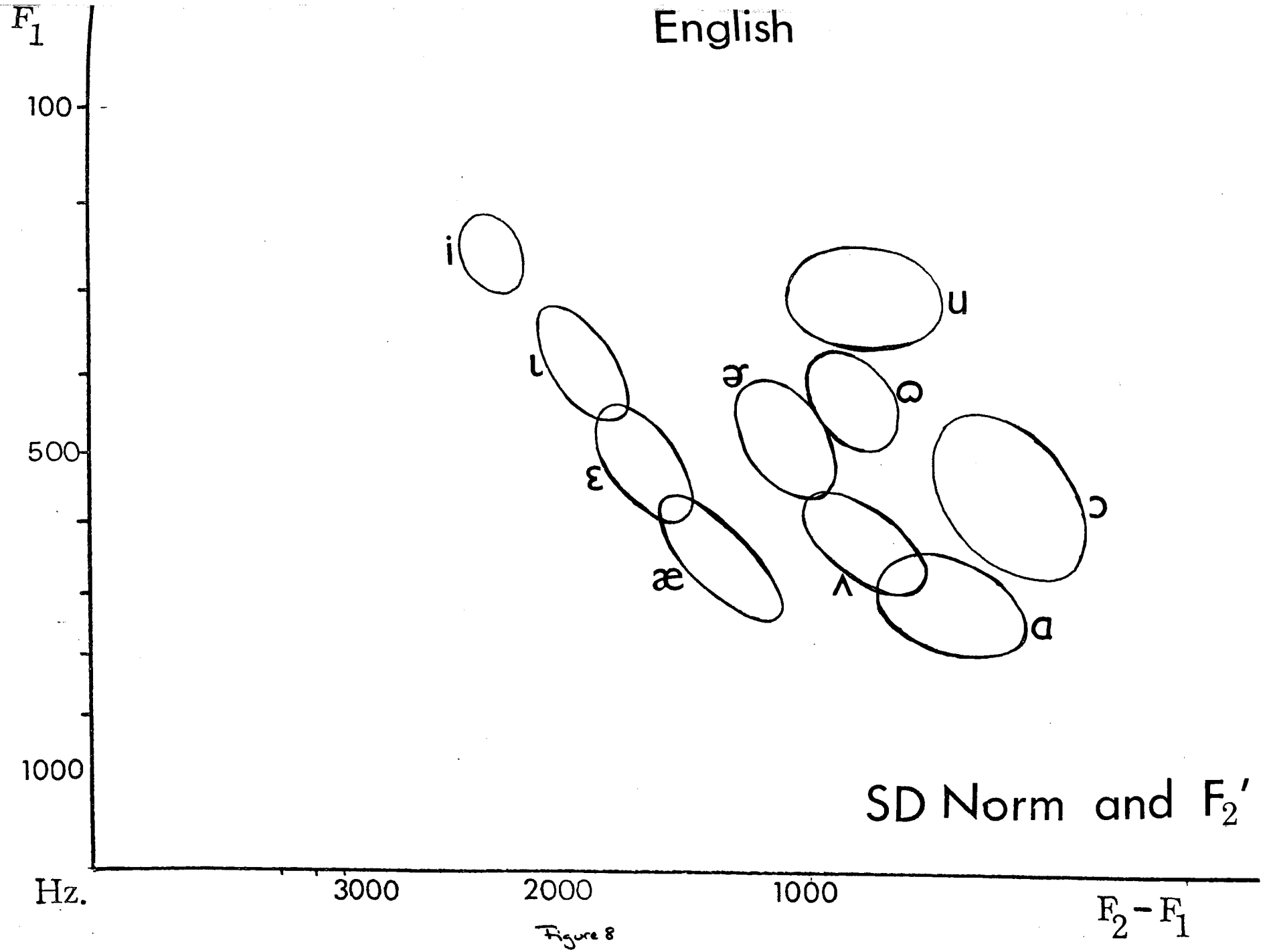


Figure 8

Now let us consider this same data after it has been subjected to two different normalization procedures that will be described in detail further on. Figure 7 shows a plot of the data, subjected to a normalization procedure based on factor analysis, as suggested by Harshman (1970). Here the scatter is reduced to varying degrees for all the vowels. A different normalization procedure (Lobanov, 1971) produces the plot in Figure 8 when applied to the data. As we can see at a glance, this reduces the scatter even further; in accordance with the above definition, this may be viewed as a more highly valued normalization, at least with respect to English.

At this point, one should briefly comment on the need for a precisely-defined evaluation metric for normalizations, and on possible alternatives to impressionistic evaluations of the type just illustrated.

Of all the possible adequate normalization procedures, there may be said to be one which is optimal. This is defined as one which eliminates the scatter to an absolute degree, reducing each vowel in the language to a single point without altering the distance between different vowels. This last qualification is necessary to obviate "normalization" procedures such as reducing each formant frequency by dividing it by, say, 1,000,000; this certainly reduces the variance of each vowel, but at the expense of also reducing the variance between vowels. An optimal normalization is, in a sense, the phonetician's aim in characterizing the vowels of a language (cf. Fig. 1), and accordingly it is an empirical problem to find a mathematical procedure which makes explicit such an aim.

A normalization procedure which approaches this optimal level more nearly than another will be judged as the better of the two, again assuming that there is no major distortion of the relationships within the vowel system. The ellipse program described above (Davis 1977) proves to be of value in making precise judgments along these lines; it seems a valid heuristic to associate the goodness of a normalization with the size of the vowel ellipses it produces when applied to any given data set, provided that the ellipses are drawn at some constant number of standard deviations from the mean of each group of points, and provided further that the sizes of the ellipses are not shrunk by simply reducing the values of all the data points. Future work should attempt to verify the latter by means of discriminant analysis.

The results of the various normalization procedures, then, can be quantified by measuring the areas of the ellipses they produce, using the formula:

$$\pi \left(\frac{a\ b}{4}\right)$$

where a and b are the axes. We will assume that a normalization procedure which produces vowel ellipses with the fewest total units is the better one. In the remainder of this discussion, all the relevant normalization procedures will be evaluated in this fashion (subject, of course, to caveats of a phonological nature).

Normalization Procedures

At this point we might consider whether it is appropriate to use any procedure in which an algorithm is applied to individual points so as to reduce the area considered to be representative of each vowel.

Averaging.

The most commonly employed of all data reduction techniques is a simple averaging of the formant values over all speakers. For each vowel, the means are calculated for each formant and then plotted on a vowel chart; each vowel is thus represented as a single point. This approach is not without its merits, and is an efficient way of reducing a large amount of acoustic data to a linguistically useful form. It is used by Lindau and Wood (1977) for Yoruba, Akan, and other Kwa languages; by Petursson (1976) for Icelandic; by Heike (1964) for German; by Halle (1959) for Russian; and by many others.

However, the averaging technique fails to show the degree of speaker-dependent variance which is present in the data; this information is crucial in determining whether two sounds in different languages are significantly different or not. As will be discussed in greater detail further on, differences of equal magnitude in mean formant frequency may prove to be significant or non-significant depending entirely on the magnitude of the speaker-dependent variance within each language. Therefore, any data reduction technique which is to be adequate for detailed cross-linguistic comparisons must accurately reflect the relative amounts of variance in each language.

In contrast to the averaging technique, which obscures the speaker-related variance in a language, more sophisticated normalization procedures attempt to reduce the variance systematically. These algorithms all proceed from the single, basic assumption that the pattern of each speaker's vowels is systematically related to the pattern of the vowels in the language. Of the numerous normalization procedures which have been proposed, we will necessarily limit our investigation to those which utilize the values of F_1, F_2, and F_3 only, as this is all the information which is provided in all the published data sets. Normalizations such as Wakita's (1977) uniform vocal tract scaling procedure and Bernstein's (1977) normalization procedure must therefore remain outside the present discussion, as the former requires bandwidth data and the latter requires formant amplitudes. Scaling techniques based on the frequencies of the higher formants (Ladefoged 1975) are also beyond the scope of this study.

Nearey (1977) provides some measure of the groundwork for the present study by evaluating several related normalization procedures, Gerstman (1968), Lobanov (1971), and two scaling techniques of his own (Nearey 1977), on the basis of their "resolving power", i.e., their ability:

> "to allow the separation of normalized formant values into distinct groups corresponding to phonetic categories."

Note, however, that this is not quite the same criterion as proposed above, since it (a) requires that the groups obtained by non-overlapping, and (b) imposes no requirement that relative distances between vowels be preserved.

Gerstman Normalization.

Nearey points out that the Gerstman (1968) procedure

> "obtained an identification rate greater than any of the procedures described above."

Gerstman's normalization procedure has been set forth in several forms, utilizing different sets of phonetic parameters. In its original form, it fixes the maximum and minimum F_1 and F_2 values in each speaker's vowel system at fixed, artibrary levels: 0 and 999 Hz. A revised form (as used at the UCLA Phonetics Lab) uses values that are more typical of speech: a minimum of 250 Hz. and a maximum of 750 Hz. for F_1, and a minimum of 850 Hz. and a maximum of 2250 Hz. for F_2. In each case, all other F_1 and F_2 values are then scaled within these ranges. In other forms of this normalization procedure there are as many as 10 separate constants: the minima and maxima for F_1, F_2, F_3, (F_1+F_2), and (F_2-F_1).

However successful it may be in normalizing the vowels of a single language, the Gerstman procedure is unsuited to cross-linguistic comparisons. By fixing the maxima and minima for the formants across languages, it makes the untenable claim that all languages have [i] of identical phonetic quality (or at least [i] of identical F_2, in those languages in which another high vowel -- e.g. [u,y,ɯ] -- tends to have a lower F_1 than [i]) and that other vowels at the extreme ends of the vowel chart are also likely to have identical degrees of height or rounding across languages. Moreover, vowel systems with maxima or minima at frequencies which differ significantly from the arbitrary values (e.g. German, lacking [ɑ], has maximum F_1 in the region of [a]) will be distorted to some degree in consequence.

SDNorm.

The next most highly valued normalization according to Nearey's criterion is the normalization proposed by Lobanov (1971). A version of the normalization, which is less likely to be affected by sampling errors--called SDNorm for"standard deviation normalization" -- was developed independently at UCLA, and will be the procedure referred to in the balance of this discussion. Unlike the Gerstman procedure, which fixes a vowel system at its endpoints, the SDNorm procedure fixes it at the center. Relatively few phonemic contrasts are made in this region, and the vowels of

most languages are distributed more or less symmetrically around this point.

Specifically, SDNorm fixes the mean formant values for all the vowels in a given system at 500 Hz. for F_1, 1500 Hz. for F_2, and 2500 Hz. for F_3 (i.e., an idealized schwa), and the standard deviation for all vowels at arbitrarily-chosen values of 150 Hz., 500 Hz., and 300 Hz. for F_1, F_2, and F_3, respectively. Formant values for all of a speaker's vowels are rescaled together according to the formulæ.

$$F_1 = \frac{F_{1ij} - \overline{F}_{1i}}{SD_{1i}} \cdot 150 + 500$$

$$F_2 = \frac{\overline{F}_{2ij} - F_{2i}}{SD_{2i}} \cdot 500 + 1500$$

$$F_3 = \frac{F_{3ij} - \overline{F}_{3i}}{SD_{3i}} \cdot 300 + 2500$$

where subscript i designates the speaker and j, the vowel; $\overline{F}_n$ the mean value for formant n, and SD_n the standard deviation for formant n.

The F_2 and F_3 values so derived may profitably be integrated by means of Fant's (1966) F_2' formula. This adds oral cavity information such as rounding and retroflexion to the schema.

It will be recalled from the previous discussion (figures 7 and 8) that SDNorm very effectively reduces the variance in a single language's vowels. (See also Appendix, table I) By extension, this procedure should be of value in cross-linguistic studies as well. It might be expected that such a reduction in the variance would increase the likelihood of discovering phonetic differences in the data. However, even though the SDNorm procedure reduces the variance within each language to a considerable extent, it also draws the ellipses into closer proximity to one another, and thus has the effect of obscuring, rather than highlighting, some of the differences between languages that were evident in the raw data. Compare, for example, the raw data in Fig. 5 (reproduced here as Fig. 9) with the SDNormalized data in Fig. 10. It will, in fact, become evident that SDNorm has the effect of reducing the distance between some -- though by no means all -- phonetically distinct vowels.

One should therefore not be too hasty in adopting the SDNorm procedure as presently formulated. In fact, the merging of the

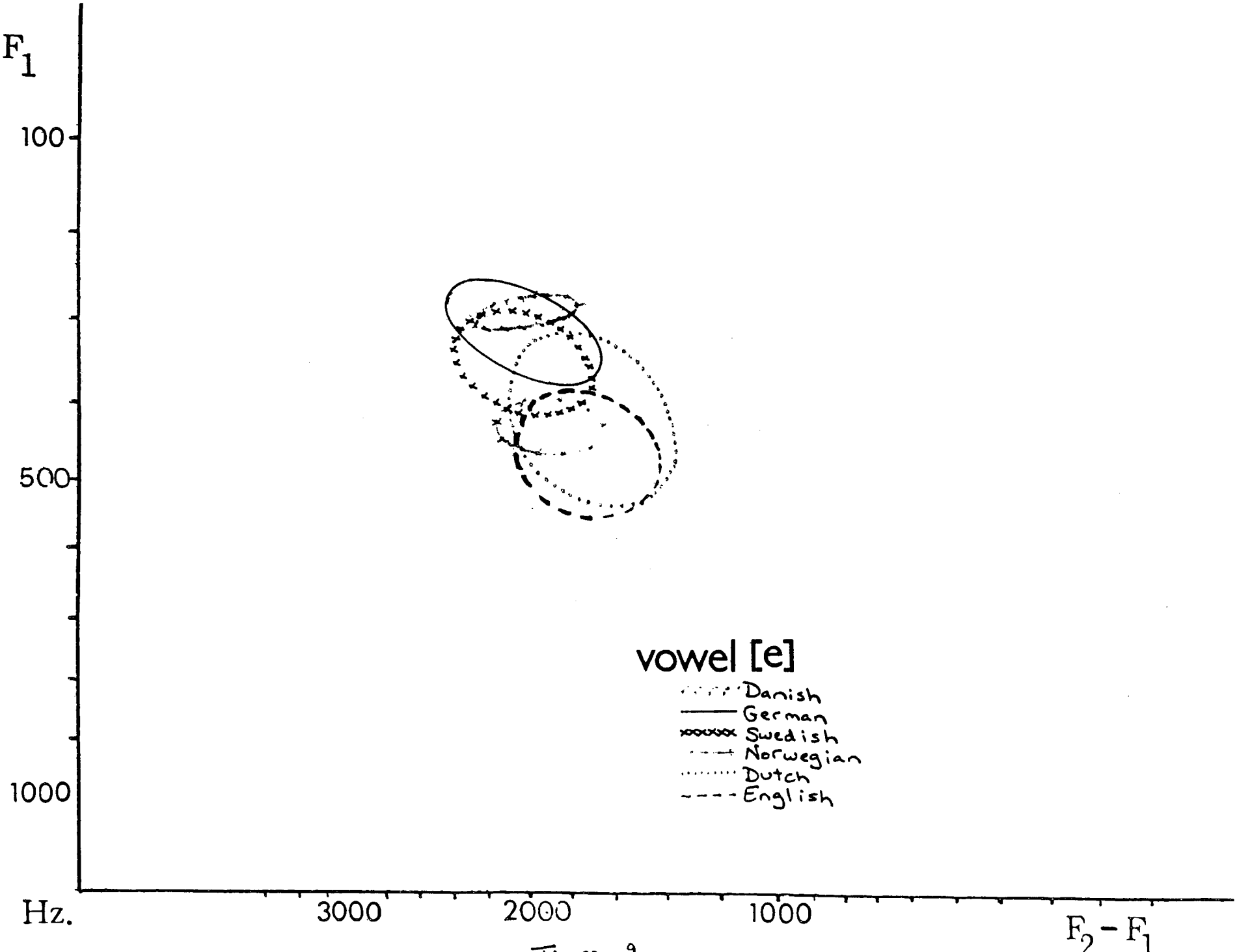

F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
vowel [e]
Danish
German
Swedish
Norwegian
Dutch
English

Figure 9

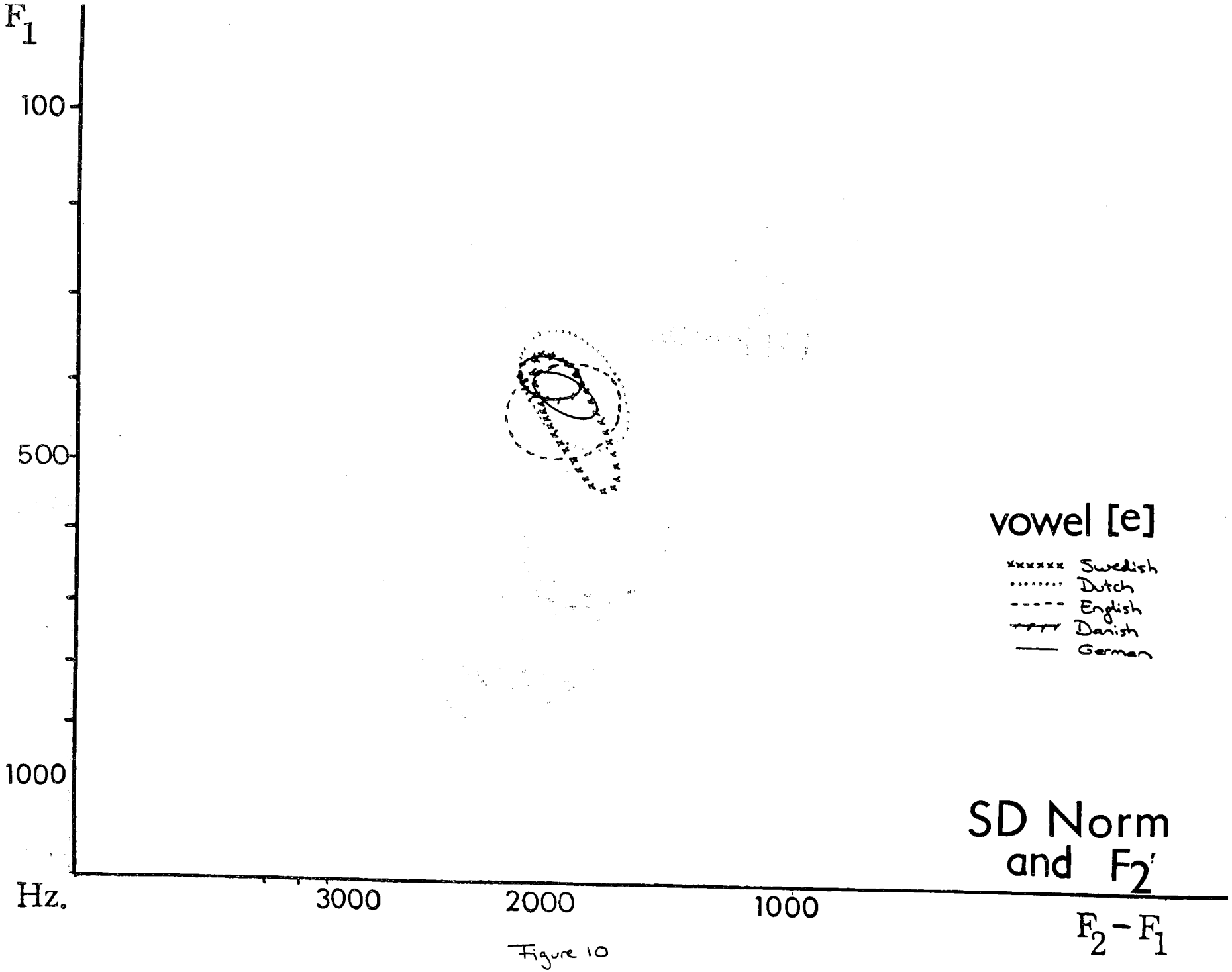

F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
vowel [e]
Swedish
Dutch
English
Danish
German
SD Norm
and F2'

Figure 10

vowels in phonetic space is at least partly attributable to some unwarranted assumptions that one must make in applying the SDNorm procedure across languages. One of these is that all vowel systems are basically symmetrical around a mean. The Germanic data at hand serves to illustrate some problems in this notion.

In normalizing the vowels of a number of speakers of a single language, the SDNorm procedure effectively superimposes a number of vowel systems in phonetic space. As it is a plausible assumption that each speaker's mean will bear a fixed relationship to the vowels of the language, this is a valid procedure. However,it does not follow automatically that a number of languages which are normalized in this fashion can readily be compared. Before superimposing two languages in vowel space -- i.e. plotting their SDNormalized vowels on the same chart-- it should first be ascertained whether each language's mean bears a comparable relationship to the vowels in the study. If they are not comparable, the normalization will, as indicated, introduce spurious trends in the data. For example, the vowel [ɪ] seems to have very similar phonetic quality in Dutch and in English, as shown by a plot of the raw data (Fig 11).

However, the SDNormalized version of the same data (Fig. 12) reveals a (spurious) difference in phonetic quality, which is not noted among the auditory impressions of the phoneticians cited in this study. This phenomenon may be seen as resulting directly from the SDNorm procedure. Fig. 13 illustrates this point.

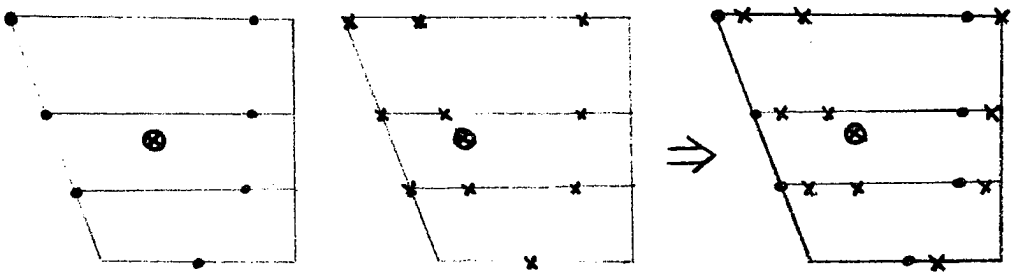


Fig. 13. Superimposition of different vowel systems (⊗ = mean)

In a language without a set of front rounded vowels, as shown at the left, the mean lies roughly midway between the F_2 values of the front and the back vowels. In a language with these same vowels and, in addition, a set of front rounded vowels, the mean lies further front, due to the preponderance of front vowels. If we were to superimpose these means, however, as is shown at the right of Fig. 13, we would find that vowels of identical quality now occupy different regions of the vowel space. To the

Raw Data

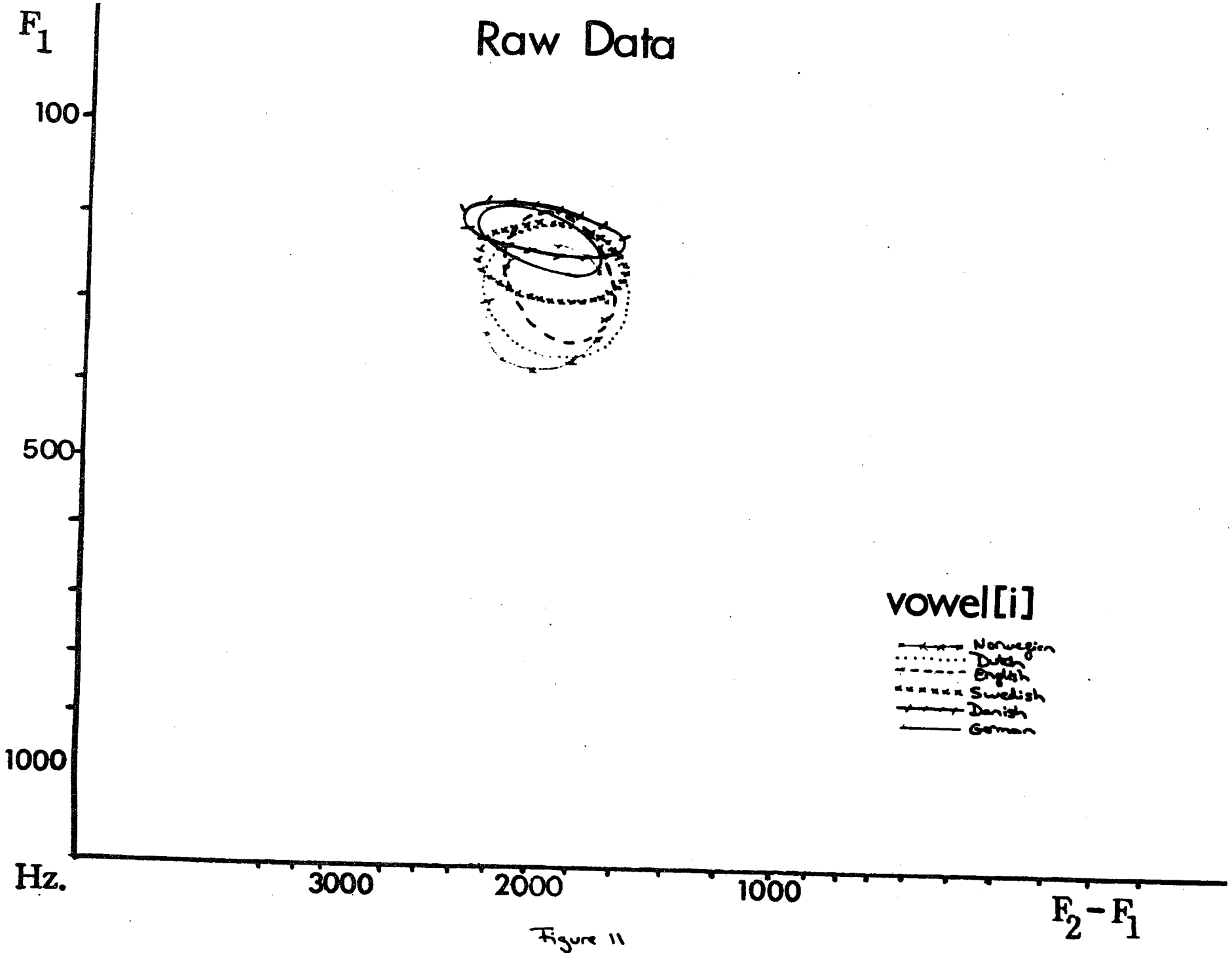


Figure 11

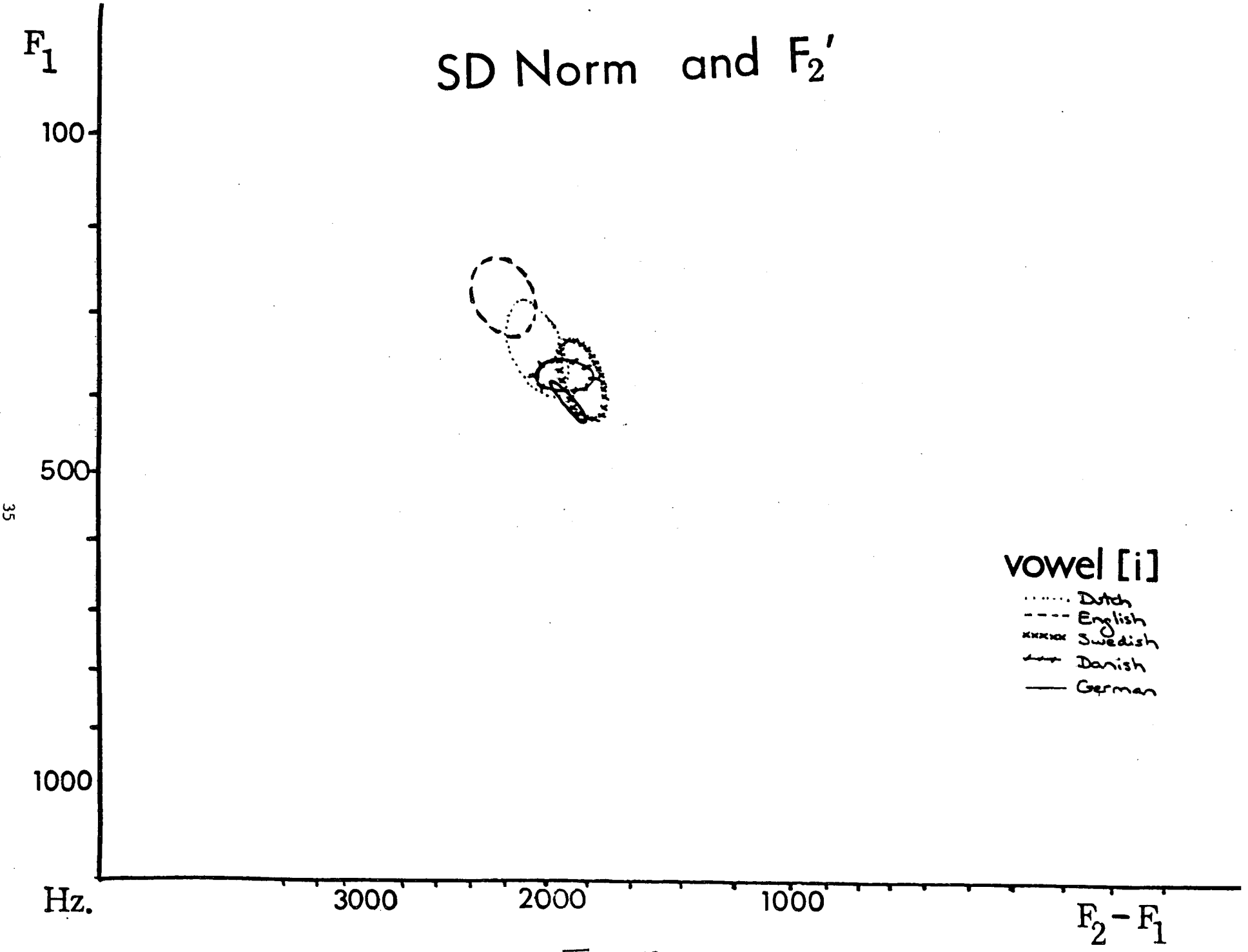

SD Norm and F_2'
F_1
100
500
1000
Hz.
3000
2000
1000
$F_2 - F_1$
vowel [i]
Dutch
English
Swedish
Danish
German

Figure 12

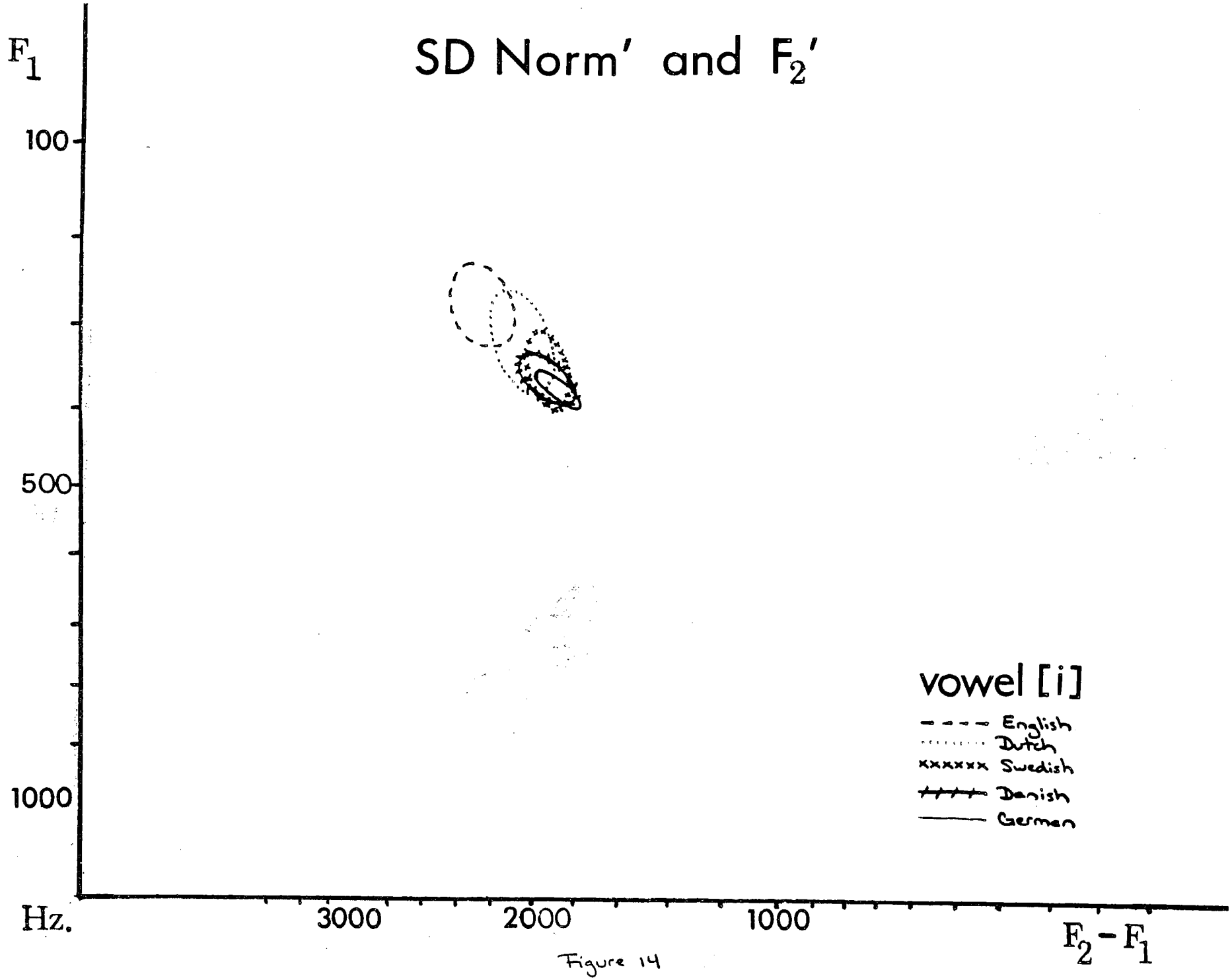

SD Norm' and F2'
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
vowel [i]
English
Dutch
Swedish
Danish
German

Figure 14

extent that differences of this sort are revealed to be artifacts of the particular normalization procedure utilized, they must be eliminated.

SDNorm'.

In an attempt to make use of the SDNorm procedure in spite of such drawbacks, it was first resolved to apply SDNorm to only those vowel systems -- or portions of vowel systems -- which are symmetrical around a central point. According to this version of SDNorm, called SDNorm', only the major vowel classes common to all the languages in the study are submitted to the normalization procedure: the remaining vowels (e.g. the front rounded vowels) are necessarily left with their original variance. However, so long as the relationships between the vowels continue to hold, this is not an untenable approach to normalization.

When we apply SDNorm' to the data, we find that the scatter is reduced even further; compare Fig. 14, for example, to figures 11 and 12. This observation is confirmed by a calculation of the areas of these ellipses (see Appendix, table I).

However, as we shall see, even this modified version of SDNorm can introduce spurious trends in a cross-linguistic comparison. As in the previously discussed example from languages with and without front rounded vowels, these effects may be attributed to certain implicit assumptions which one must make in comparing languages on the basis of their normalized values. And again as in the previous example, these effects only come into play when the normalization is to be the basis for comparisons across languages. When studying the relationships between the vowels of a single language, SDNorm presents no such drawbacks.

When making cross-linguistic comparisons, the two basic steps in an SDNorm procedure are to set the means of all the vowel systems to the same point in the vowel space and, similarly, to set all the standard deviations from those means to a fixed arbitrary value. This effectively superimposes a number of vowel systems in the same region of vowel space. In using the same fixed values to normalize data sets from several different languages, and in making a principled comparison of the languages on the basis of these normalized figures, it is implicitly assumed that all the languages have comparable means and comparable standard deviations. Otherwise, it would be impossible to draw valid conclusions from the observed patterns.

To the extent that languages violate the basic assumption of fixed mean and standard deviation, SDNorm and its variant,

SDNorm', must be regarded as unsuitable for cross-linguistic comparisons.

German, for example, seems to violate this assumption on the phonetic level. Unlike the cited example involving front rounded vowels -- with an asymmetry which is evident on the phonological level -- this and other, subtler phonetic effects cannot be compensated for a priori. Because the SDNorm procedure assumes a fixed standard deviation from the mean -- that is, a more or less standard degree of peripherality -- for all vowel systems, it cannot reflect Moulton's (1962) auditory impression that "all German vowels are tenser than their English counterparts." This may be illustrated with plots of the vowels of German and the corresponding vowels of California English, both in their raw form (Fig. 15) and after SDNorm' (Fig. 16). The raw data (Fig. 15) seems to bear out Moulton's observation: with the exception of [i], which in this dialect of English is higher but no farther front than the corresponding German vowel, all the German vowels are indeed "tenser than their English counterparts." In contrast, the SDNormalized figures are, at best, inconclusive. The high vowels of German here appear to be generally more lax than their English counterparts; only the F_2 value of German [u] gives even a slight suggestion of the auditory tenseness. California English has an exceptionally high F_2 value for [u], probably because the vowel is progressively losing its rounding. (In comparison with the California English data, the Peterson-Barney data — not used in this example because it lacks [e] and [o] — starts out with lower F_2 values for [u] and, after SDNorm', ends up with higher values.) With respect to the mid vowels [e o ɛ], which also should be tenser in German than in English, the quality appears to be virtually identical in the two languages.

There is other evidence which indicates that the means of various languages are no more comparable than are the standard deviations. One commonly encounters generalizations of the form:

> French front vowels are more fronted than Germanic front vowels.
>
> Danish is spoken higher in the mouth than other Scandinavian languages.

The latter observation can be verified with the data at hand.

Even without the benefit of normalization, the plotted formant frequencies of the Danish long vowels (Fig. 17) are sufficiently high to lead to the conclusion that "almost all the Danish vowels are placed in the upper third of Jones's [cardinal vowel] diagram." (Fischer-Jørgensen, 1972). It has for example,

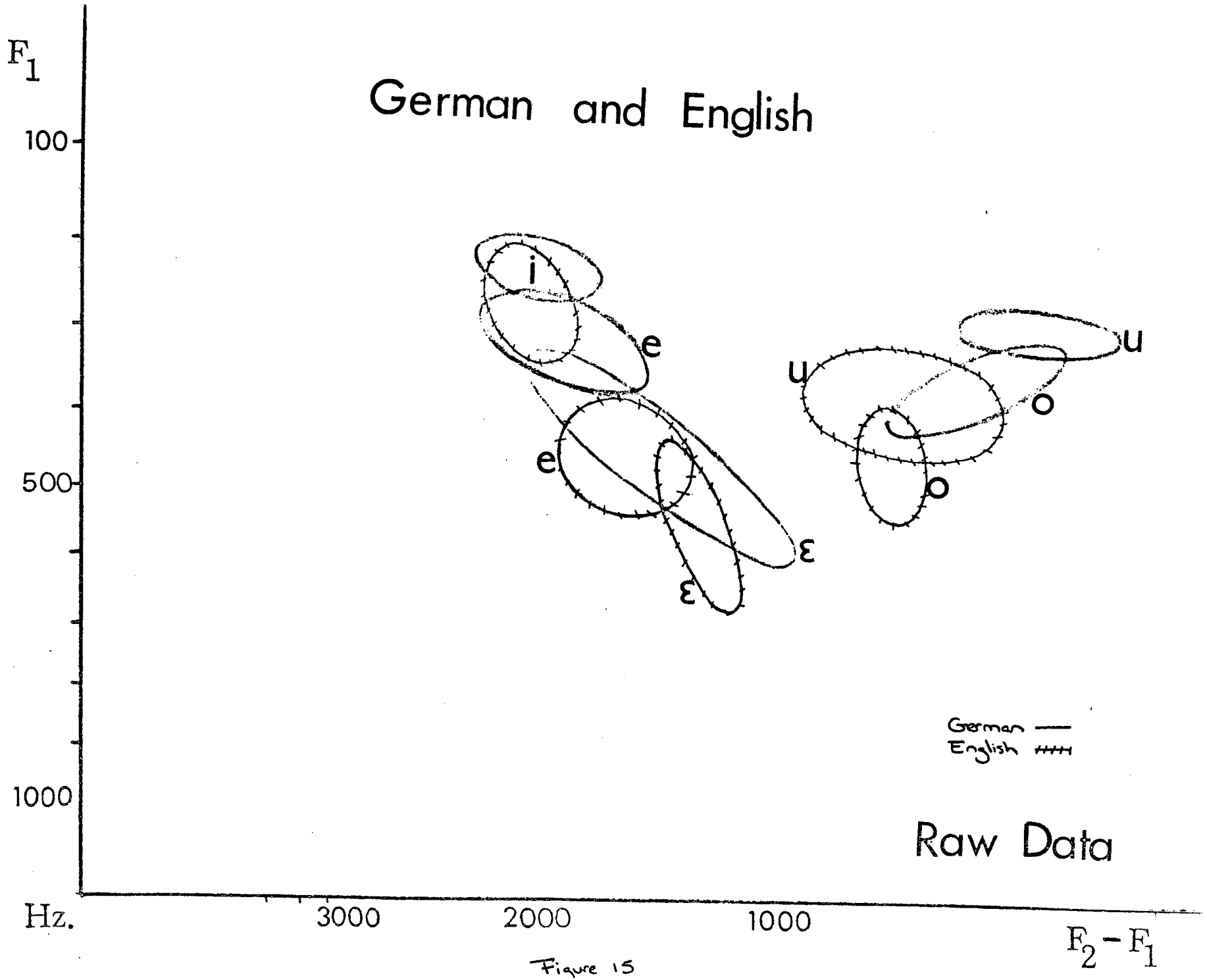

German and English
F_1
100
500
1000
Hz.
3000
2000
1000
$F_2 - F_1$
i
e
e
ɛ
ɛ
u
u
o
o
German ——
English ++++
Raw Data

Figure 15

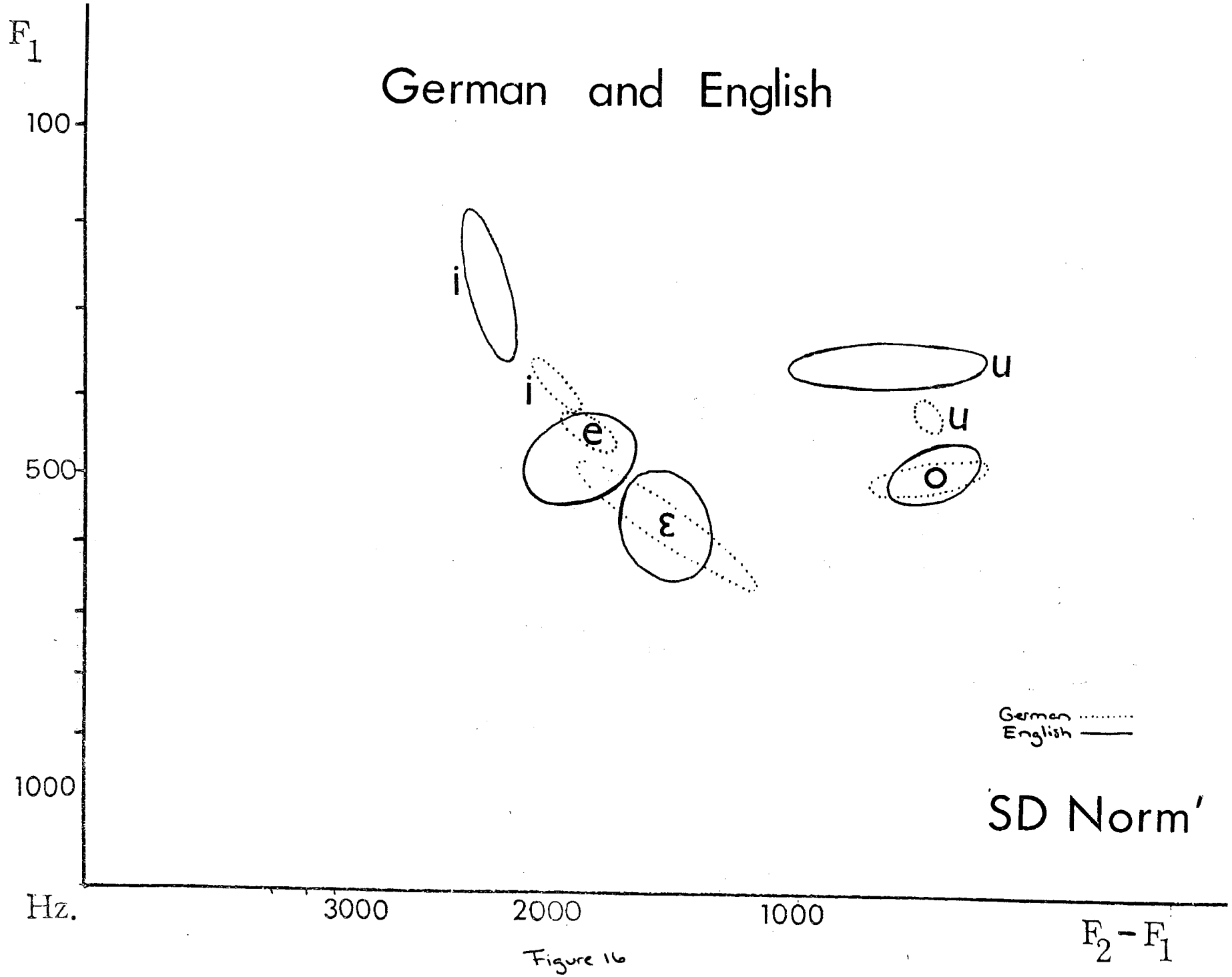

German and English
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
i
i
e
ε
u
u
o
German .....
English ——
SD Norm'

Figure 16

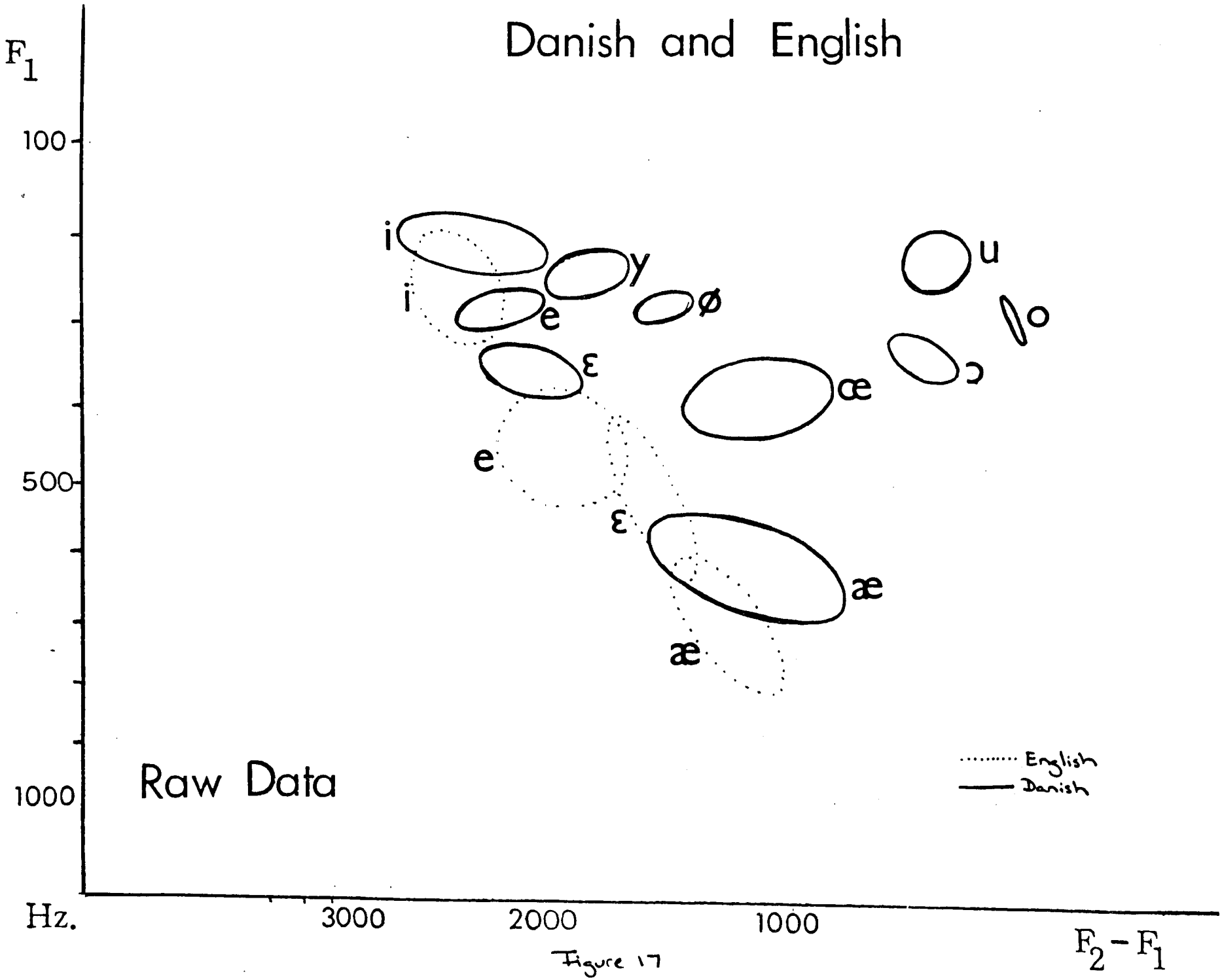

Danish and English
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
i
y
u
ɪ
e
ø
o
ɛ
œ
ɔ
e
ɛ
æ
æ
Raw Data
English
Danish

Figure 17

been noted that:

Danish [i] is "tenser than the English [i] in see." (Walshe)
Danish [e] is "more close than Swedish [e]" (Nielsen and Hjorth)
or "almost as high as English [ɪ] in pin." (Bredsdorff)
Danish [ɛ] is "more close than Swedish [ɛ]" (Nielsen and Hjorth)

It may thus be concluded that the phonetic mean for Danish vowels (leaving aside the front rounded vowels, for the sake of clarity) is considerably displaced from the center of the vowel space. Such a distribution may be shown schematically as in Fig. 18a, as compared to a more symmetrical vowel system such as Swedish or English (with the appropriate additions) as in Fig. 18b. If we were to normalize the data with SDNorm' and plot all the vowels on the same chart, we would once again encounter problems of superimposition, that is, we would find Danish [i] to be less close than English or Swedish [i], which is contrary to phoneticians' auditory judgments.

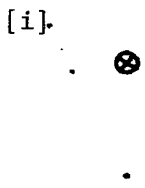
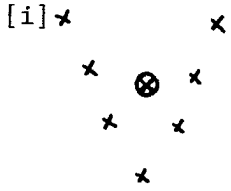
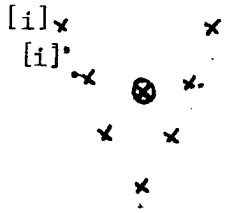


Fig. 18
a. Vowel system as in Danish (front rounded vowels excluded).
⊗ = mean.



b. System with vowels distributed normally around mean.
⊗ = mean.



c. Superimposition of the two systems above, with means coinciding.

Lobanov Normalization.

At this stage it seems as though the only way to rescue the SDNorm procedure for use in cross-linguistic studies is to have it first calculate the grand mean and standard deviation for all speakers of a language, and then to use these figures rather than the arbitrary ones (mean: 500, 1500, and 2500 Hz.; standard deviation: 150, 500, and 300 Hz., for F_1, F_2, and F_3, respectively) in the SDNorm procedure. This was suggested by Lobanov (1971) in his study of the acoustic characteristics of Russian vowels; however, it should be noted that Lobanov makes no claims as to the suitability of this procedure for cross-linguistic studies.

If this version of SD Normalization, which we shall call Lobanov normalization, is to be applied cross-linguistically, special care should be taken to ensure an unbiased sampling of the population. As the mean and standard deviation of each normalized data set reflect the overall positioning in phonetic space of the vowels in the sample, rather than some arbitrary figure, it is essential that this positioning be determined by the phonetic quality of the vowels, rather than by the anatomical traits of the speakers in the sample. We would wish to avoid selecting, for example, a group of particularly slight Dutch speakers for purposes of comparison with a more representative sample of English speakers; this would falsely indicate that the means for the two languages were phonetically distinct.

Quite apart from such sampling problems, the Lobanov procedure is subject to some of the same system-related drawbacks that the other versions of the basic SDNorm procedure are. For example, even though the Lobanov procedure does not assume that all languages have comparable means or comparable standard deviations, it nevertheless must assume a certain symmetry of distribution on the phonetic level, which ensures that the mean formant values will lie at or near the center of the vowel system.

As we have seen from the plot of the vowels of Danish, (Fig. 17), this symmetry of distribution does not always obtain in languages. It will be argued that this asymmetry in the Danish vowel system renders even Lobanov normalization unsuitable for cross-linguistic investigations.

The plotted results of the Lobanov normalization show the distribution of the vowel [i] in five languages to be as in Fig. 19. Note that, although four of the five languages bear relationships to one another which are remarkably similar to their relationships in the raw data (Fig. 11), Danish [i] is placed considerably lower in phonetic space (=higher F_1 values) than it was originally.

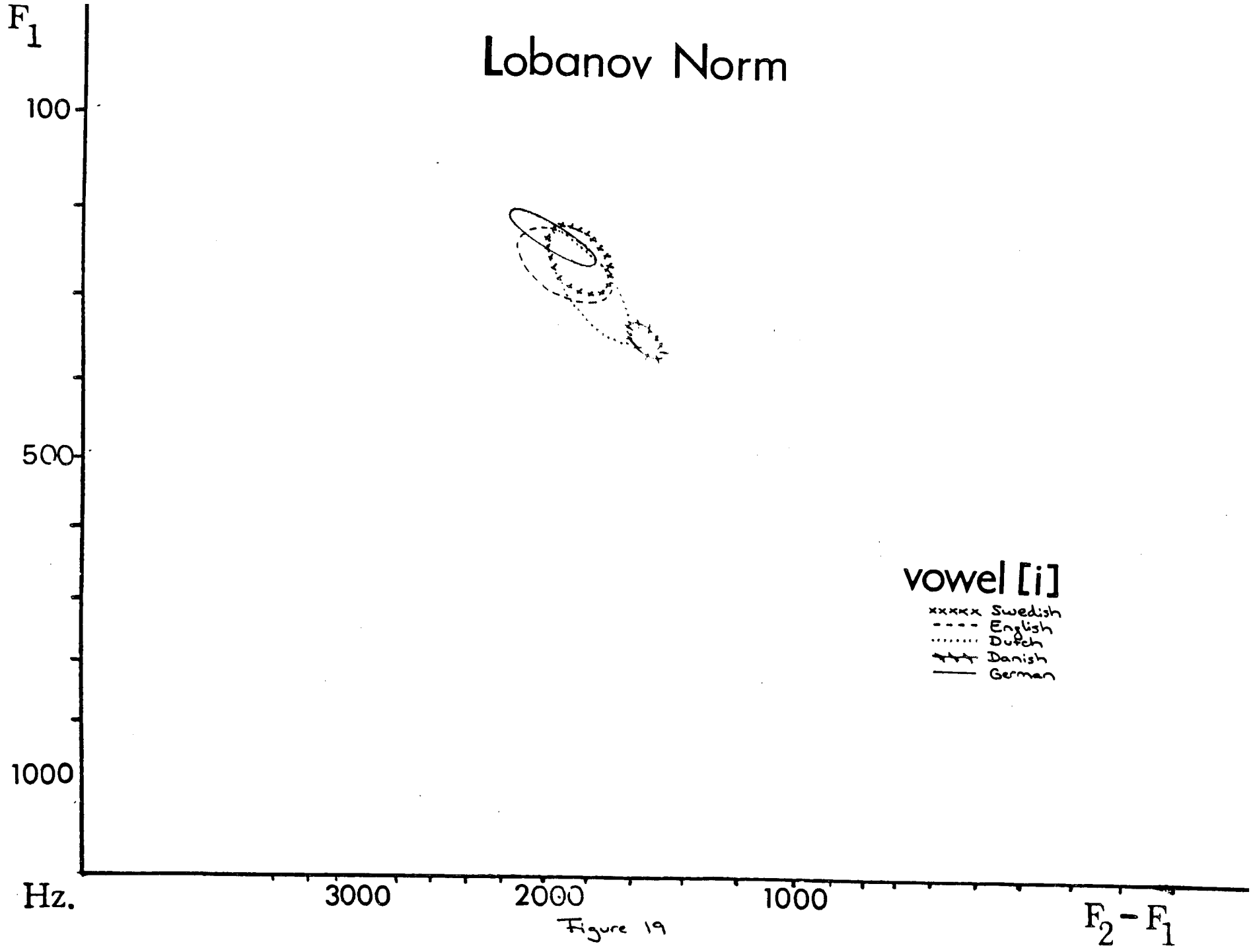
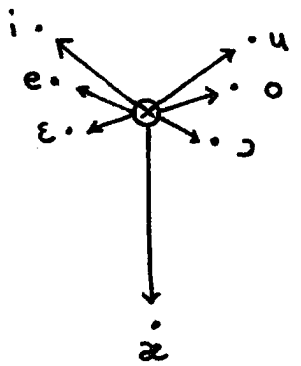

Lobanov Norm
F1
100
500
1000
Hz.
3000
2000
1000
F2 − F1
vowel [i]
xxxxx Swedish
---- English
..... Dutch
+++ Danish
—— German

Figure 19

This phenomenon has a rather simple explanation. Unlike the other languages in the study, whose means occupy different points in vowel space, but which always fall at or near the center of their respective vowel quadrilaterals, Danish has a mean which is strongly influenced by the phonetic height of six of its seven vowels. [16] Schematically, the system is:



Any characterization of data in terms of its standard deviation assumes a normal distribution. Yet it is evident from this diagram that the vowels of Danish are not normally distributed around the mean; such distributions should lead to distortions in the resultant data. This is just what we observed in the case of the Danish vowel [i].

PARAFAC

In view of the drawbacks, of both a practical and a theoretical nature, to the normalization procedures examined thus far, we may wish to consider factor analysis as a means of determining the relationships between phonetic vowel quality and observed formant frequencies.

Harshman (1970) proposed PARAFAC, a three-mode factor analysis model which can be applied to the data as a means of seeking the "true underlying" influences which determine these relationships. The PARAFAC model assumes that there are certainly underlying dimensions of vowel quality which define the vowels of a given language, although "different persons will not use each of the underlying dimensions to exactly the same degree in their vowel production." (Harshman p.48) Thus, a speaker who has larger cavities, or who speaks with a greater degree of lip rounding than the norm, may be characterized as utilizing certain aspects of vowel quality to a greater degree than do most other speakers across all vowels than do most other speakers of the language. If and when reliable dimensions of

[16] Again we are excluding from consideration the front rounded vowels, which have no role in SD Normalizations of this sort.

vowel quality are discovered, it should be possible at least in theory to remove a good deal of the speaker-particular variance and thereby achieve a normalization of the data.

The PARAFAC procedure seeks factors which best fit the deviations in the data for each language, with the formants, vowels, and speakers serving as the variables. A number of factors may be extracted, and it is an empirical problem to seek the minimum number of factors which optimally describes the data. Such an optimal point may be said to be reached when additional factors are found to improve the fit only slightly. Previous work (Harshman 1970) has shown that, for purposes of the present investigation, a two-factor approach meets this criterion.

PARAFAC determines the best way of predicting the formants, F_1 and F_2. In a two-factor analysis, the deviations from the mean F_1 that occur in vowel i as spoken by speaker j are taken to be the sum of two products, each involving a constant (a loading) proportional to the particular speaker and another constant (loading) proportional to the particular vowel. Put more formally, the PARAFAC program minimizes the error in an expression such as (1):

$$(1) \quad d_{ij} = v_{1k}s_{1j} + v_{2_i}s_{2_j} + e$$

where d_{ij} is the observed deviation from the mean first formant frequency for vowel i as spoken by subject j, v_{1_i} is one of two constants applying to all instances of vowel i, s_{1_j} is one of two constants applying to all vowels spoken by subject j, v_{2_i} and s_{2_j} are the other constants for that vowel and that speaker, and e is an error term. A similar expression is applicable for the second formant. The output of the PARAFAC procedure is thus a set of loadings for vowels and speakers which best predicts the data for each formant. These factor loadings, it is claimed, reflect the underlying dimensions of vowel quality in a given language.

The procedure employed in deriving a set of PARAFAC-normalized formant frequencies from the PARAFAC output is to take the product of the loadings on vowels by loadings on formants by mean loadings on speakers and to add back in the previously subtracted means.

This procedure was followed for each of the data sets in the present study, and the resultant formant frequency values were converted to F_1 and F_2' and as before, plotted in F_1 vs. $(F_2'-F_1)$ space.

A plot of the PARAFAC-normalized vowels of English, based on the data set in Peterson and Barney (1952), reveals a distribuion as in Figure 7 (reproduced here as Fig. 20). As can be seen in a comparison of these normalized values with the raw data plotted in Fig. 6 (reproduced here as Fig. 21) the PARAFAC normalization procedure reduces the size of the ellipses without shrinking the overall vowel space. (As has been noted, it would be a trivial normalization if this were not to hold true.) However, it is also evident from the graphs, as well as from the relative areas of the ellipses produced by the various normalizations (Appendix, table I), that the PARAFAC normalization reduces the scatter comparatively little. The dimensions of the PARAFAC ellipses more closely resemble those of the raw data than those of the more successful normalization procedures, SDNorm and SDNorm'.

Yet in spite of PARAFAC's inability to reduce the areas of the ellipses to any great degree, it will be claimed that the PARAFAC procedure is more valuable than SDNorm or SDNorm' for purposes of cross-linguistic comparison.

As is evident from Fig. 22 , the PARAFAC procedure normalizes the problematic Danish data in a way which is generally in accord with the auditory judgments of trained phoneticians: as will be recalled from the previous discussion, Danish [i] is said to be higher and more tense than English [i]. Similar differences hold between English [i] and the [i] vowels of Swedish and German. Swedish [i] for example, has been described variously as "closer than the vowel in English seen" (McClean) and "much tenser than English [i] in see." (Walshe). In further corroboration of these observations, McClean points out that for Swedish [i] the tongue is "so close to the hard palate that it ends with a fricative sound [j]." German [i] (Moulton) is also described as "more tense" than English [i]. All of these differences are evident in the PARAFAC-normalized data.

If we examine figures 12,14, and 19,the results of the various SD normalizations, however, we find them to be in direct opposition to the auditory judgments cited above. In nearly all cases, it is the English data which appears to be higher and more tense than the corresponding data in other languages. As we have argued above, all these SDNorm procedures are rather suspect on auditory/ acoustic grounds.

Figs. 23-24 show that PARAFAC similarly preserves the auditory judgments for the vowel [e].

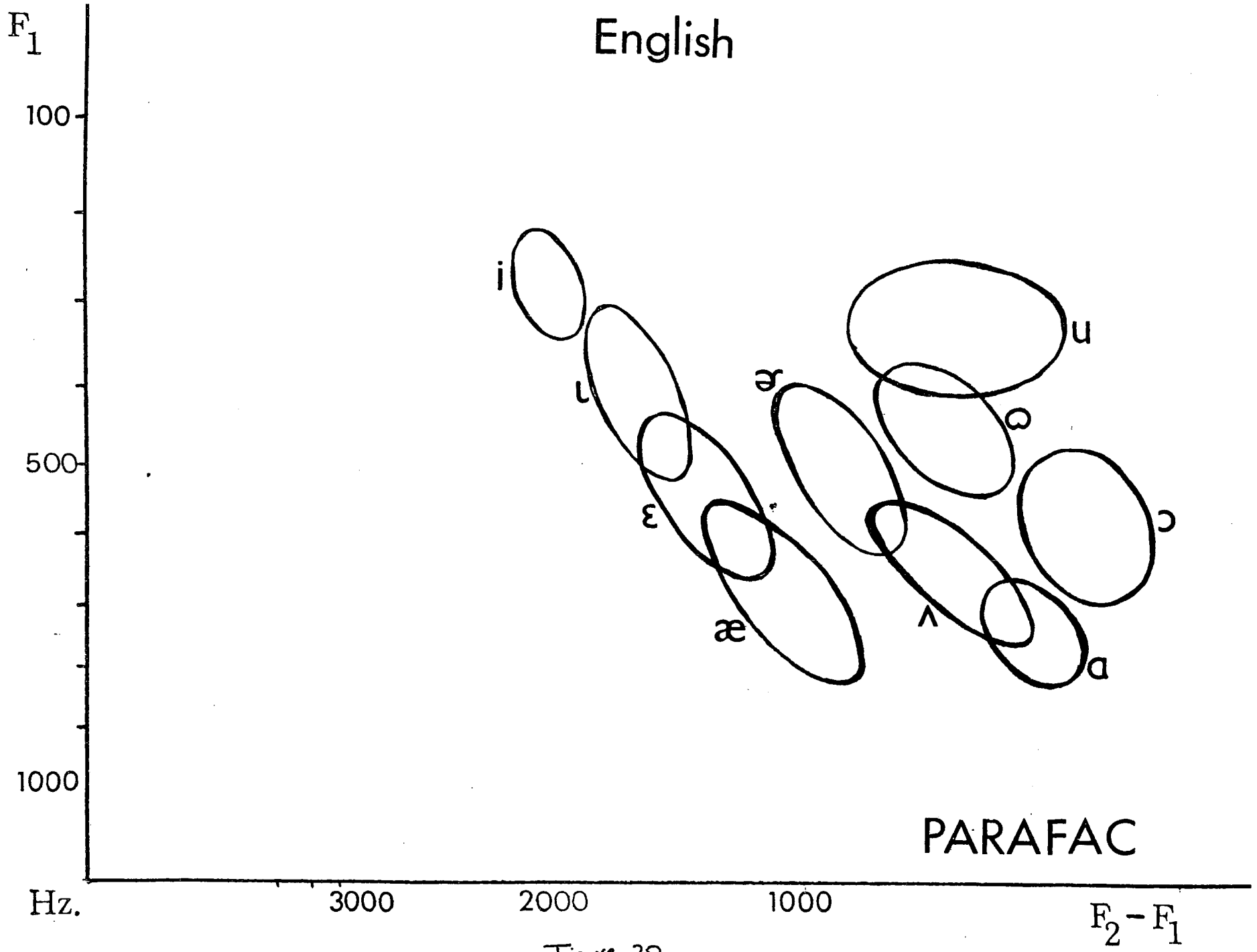

English
F1
100
500
1000
i
ɪ
ɛ
æ
ɚ
ʌ
ʊ
u
ɔ
ɑ
PARAFAC
Hz.
3000
2000
1000
F2 - F1

Figure 20

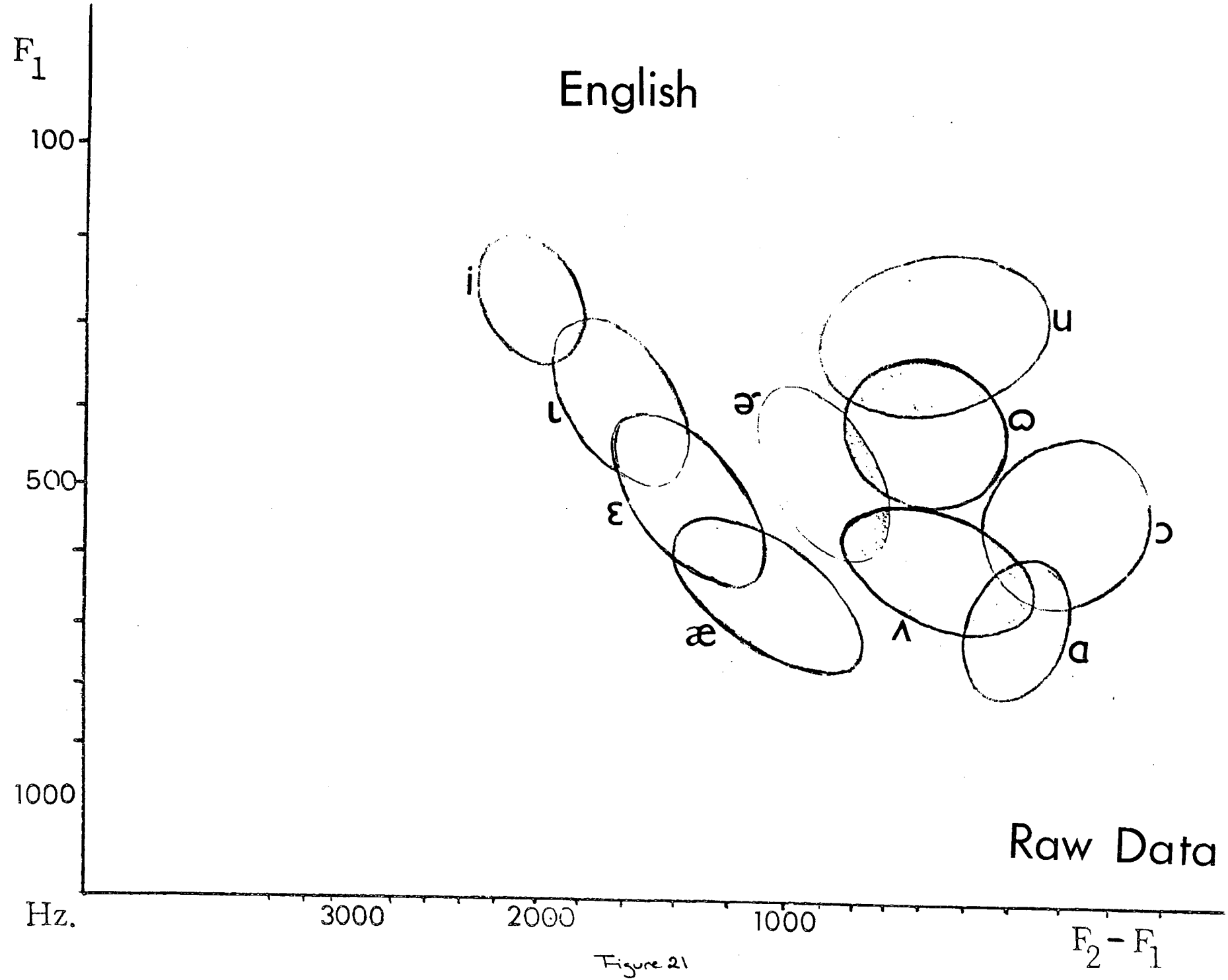

English
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
i
ɪ
ɛ
æ
ɚ
ʌ
u
ʊ
ɔ
ɑ
Raw Data

Figure 21

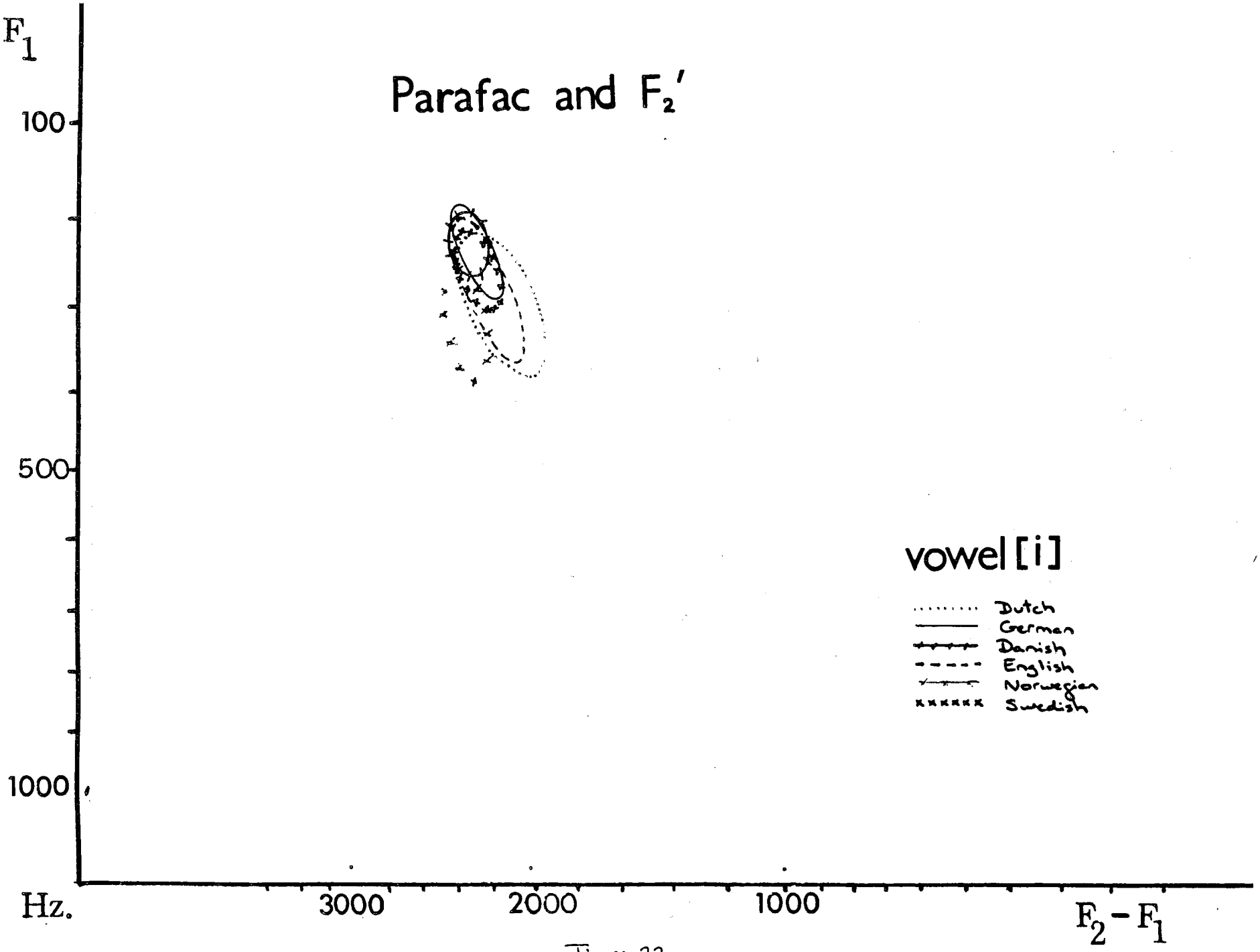

Parafac and F_2'
F_1
100
500
1000
Hz.
3000
2000
1000
$F_2 - F_1$
vowel [i]
Dutch
German
Danish
English
Norwegian
Swedish

Figure 22

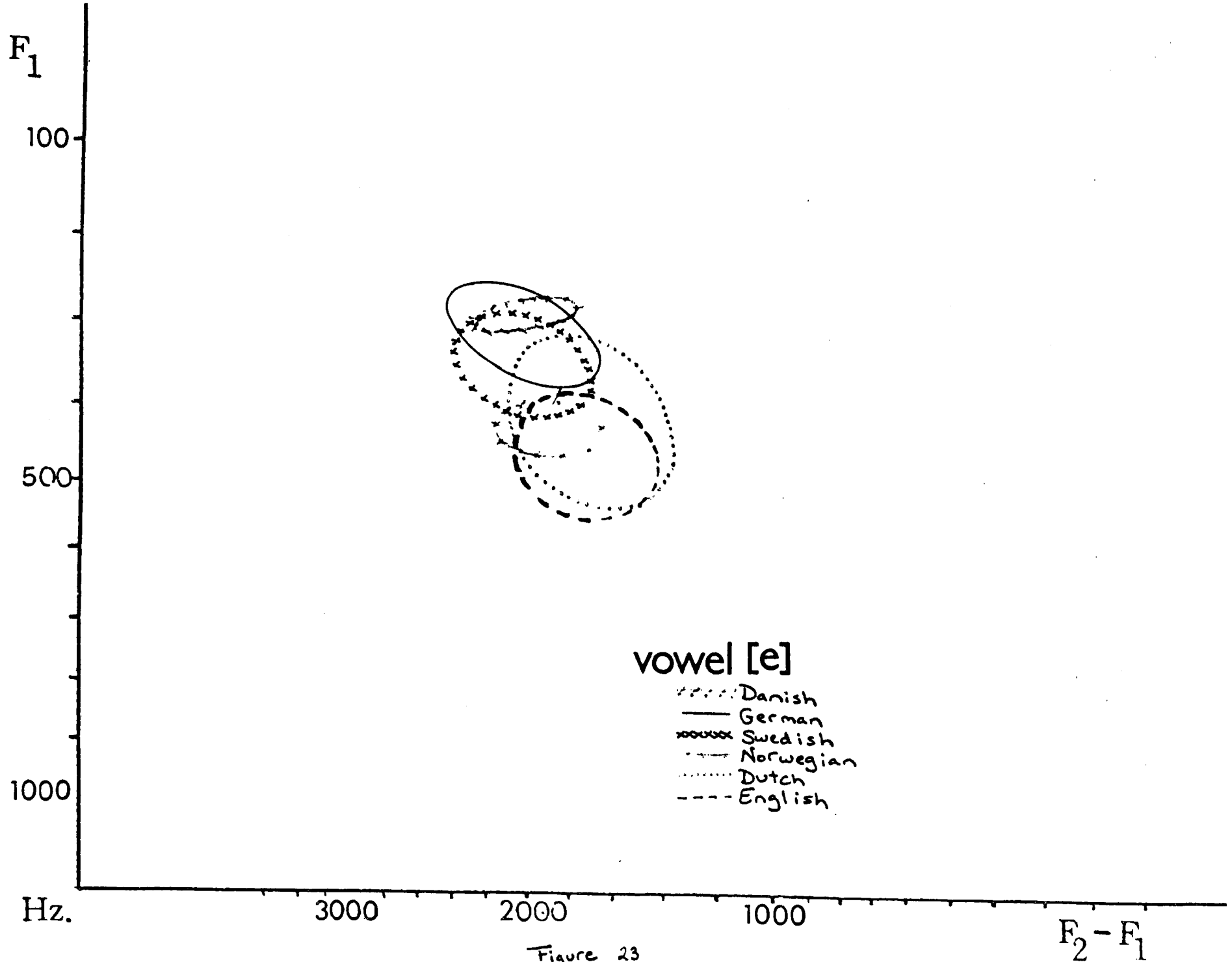

F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
vowel [e]
Danish
German
Swedish
Norwegian
Dutch
English

Figure 23

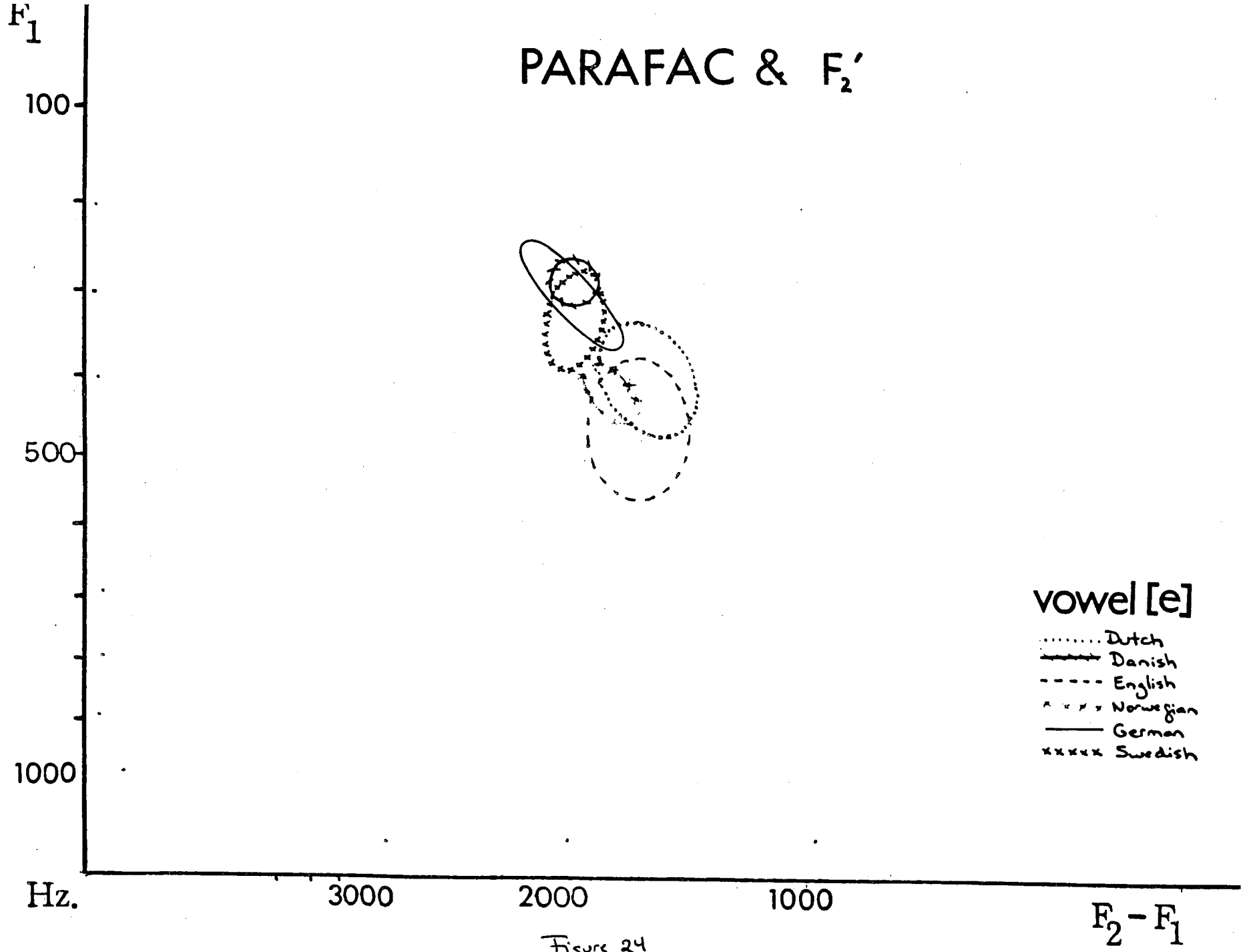

PARAFAC & F_2'
F_1
100
500
1000
Hz.
3000
2000
1000
$F_2 - F_1$
vowel [e]
Dutch
Danish
English
Norwegian
German
Swedish

Figure 24

There are further grounds on which to question the SDNorm results, and these are fundamentally articulatory in nature. Fant (1965) has observed that "the Russian and Scandinavian [i] vowels...are prepalatal, whereas the [i] is articulated more toward the midpalatal region in English," giving the former "a sharper quality."[17] Fant attributes this to the respective cavity structures of the vowels. A true palatal [i] (as in English) is characterized by a wider oral cavity and a narrower pharyngeal cavity than a corresponding prepalatal [i]. Both of these cavity configurations are associated with lowered values of F_3.

It is generally agreed that articulatory backness is associated with the frequency of the second formant. A relatively low F_2 value (and, to some extent, also F_3) will tend to correspond to relatively advanced place of articulation, all other things being equal. This has important implications for our evaluation. We have seen that the various SDNorm procedures reduce the vowel ellipses in a way which results in English [i] having higher F_2 than Swedish [i] (Figures 12, 14, 19); this fact suggests that the former is very likely prepalatal and the latter, midpalatal. Yet Fant's data shows that the very opposite holds true.

An examination of the PARAFAC results for the vowel [i] shows English to have higher F_1 and lower F_2 than Danish, and somewhat higher F_1 and only very slightly lower F_2 than Swedish. These trends are all in the predicted direction. However, as the PARAFAC procedure is unable to remove much of the scatter in the data, the results remain somewhat inconclusive. The former case may well be a reliable difference; the latter is probably not. Statistical tests, as will be described in the following section, are needed to determine the significance of the differences with greater precision.

As a final note, it should be emphasized that the validity of the PARAFAC procedure is not necessarily diminished by its failure to show <u>significant</u> differences in some of the expected directions. There may in fact be a good deal of variation in dialect and in the overall sampling between the phoneticians' data at hand. Indeed, a different data set of English (not reported in this study), drawn from five speakers of Midwestern American English, shows quite marked differences from the other Germanic languages with respect to [i]. The differences in the

---

17 Fant points out that the cases of prepalatal [i] which he has observed have mostly come from languages which contrast [i] with [y] or [ɨ] qualities. As the latter qualities have characteristically low values of F_3, "the contrasts are enhanced by having maximally high F_3 in [i]." This may thus be viewed as an argument for a principle of perceptual separation.

raw data are in the direction indicated by the auditory descriptions and by Fant's articulatory observations. The most we can legitimately ask of a normalization is that it reduce the scatter in a data set; it cannot be expected to introduce trends which are not present in the raw data, and it should in fact be less highly valued if it does so. Auditory and articulatory information of the type described is useful in making a choice between two competing normalization procedures; in fact, this step in the evaluation procedure must necessarily take precedence over considerations of scatter reduction. Yet it should be kept in mind that this information is only suggestive, and does not constitute an end in and of itself.

Analysis of Variance

In order to ascertain what differences actually obtain between any two languages, it is necessary to compensate for systematic speaker-particular variance,much as the phonetician does in assigning phonetic symbols. As we have seen, several different normalization procedures may be used, each of which has its particular advantages and drawbacks.

It is also possible to achieve this goal, while bypassing many of the inherent problems of normalization, by performing an analysis of variance on the data. This makes it possible to compare directly the amount of difference between the languages and the amount of speaker-dependent difference within the languages. The analysis of variance procedure takes as input the raw formant data in mels and calculates the overall mean of each language as well as the mean of each set of comparable vowels across languages. (The mean of each speaker's formant frequencies is also calculated, but as it plays no role on the systematic phonetic level, this information is not utilized in the present study.) Overall differences and pattern differences are determined from the distribution of the means, and their significance is ascertained by utilizing the results of the analysis of variance.

Let us examine how the analysis of variance (ANOVA) procedure may be used to test the significance of trends which appear in the raw data. Fig. 25 shows the raw values of the four comparable front vowels (i e ɛ æ) of Danish and California English, plotted on the same graph. The ANOVA results reveal that in the F_1 domain, there is a highly significant language effect ($p < .0001$) -- that is, that the overall mean of the Danish vowels is higher than the overall mean of the English vowels. There is also a very significant ($p < .001$) "pattern effect" (technically, the language by vowel interaction effect), which shows the sets not only to be centered at different locations in the vowel space, (schematically: ), but

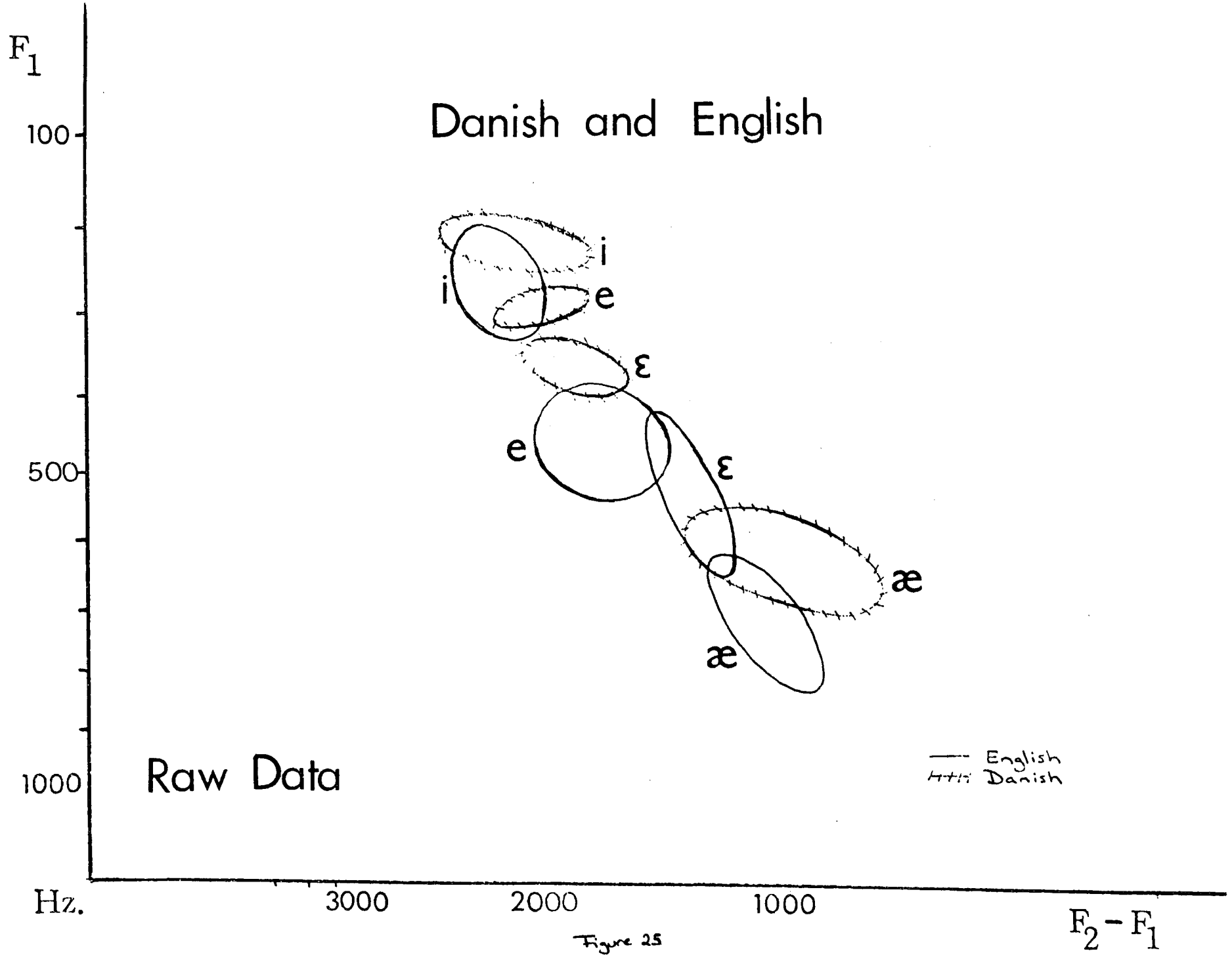

Danish and English
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
i
i
e
ε
e
ε
æ
æ
Raw Data
English
Danish

Figure 25

also to be patterning differently around these means (schematically: ▽ ◁). That is to say, while all the vowels of Danish are higher, on average, than the vowels of English, not all the vowels participate equally in determining this difference. It can be seen from the graph that Danish [e] and [ɛ] are more different from their English counterparts than are Danish [i] and [æ], at least with respect to their F_1 values.

The F_2 values, however, reveal a rather different set of facts. There is still a very significant pattern effect ($p < .001$) in the front-back dimension, but the language effect has largely disappeared ($p < .15$)[18] In other words although there is no significant difference in backness between English and Danish as a whole, nevertheless there are differences between English and Danish in the degree of backness of individual vowels.

The non-significance of the overall mean differences in F_2 is suggestive in its own right. It indicates that the significant differences in F_1 between Danish and English are real phonetic differences, rather than differences due to physiological factors separating Danes from Americans as a whole. If, for example, Danes were on the average significantly larger than Americans one would expect both F_1 and F_2 to display a language effect. This follows from a simple tube model which claims that both formants are affected by tube length in analogous fashion. According to such a model, all of the formants will covary in response to a change in tube size. The lack of such covariance in the Danish/English data, then, suggests that the differences are probably not physiological, but rather, phonetic.

Having ascertained that significant differences exist between the overall means of the languages and between their respective patterns, we may go on to determine how the individual vowels are distributed around these means. It is this step in the ANOVA procedure which, like the various normalization procedures, seeks to determine what significant differences hold between individual vowels in different languages which are transcribed with the same phonetic symbol. Once the significance of an overall difference betwen the F_1 or F_2 values of different languages has been calculated by the ANOVA procedure, it is possible to determine where in the pattern the significant differences reside.

---

18 See discussion of this point, p 58 .

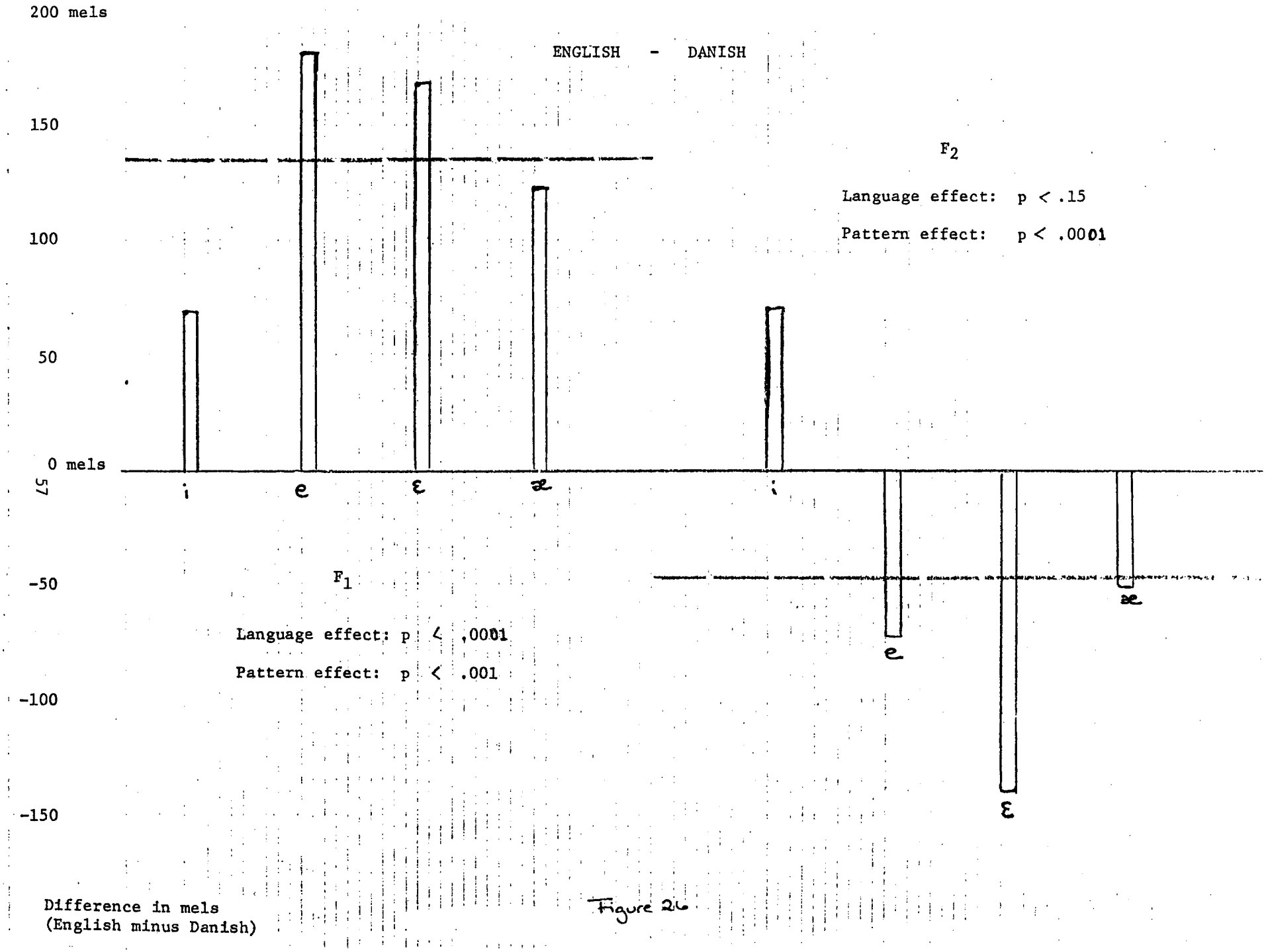

200 mels
150
100
50
0 mels
-50
-100
-150
ENGLISH - DANISH
F_2
Language effect: p < .15
Pattern effect: p < .0001
F_1
Language effect: p < .0001
Pattern effect: p < .001
i
e
ɛ
æ
i
e
ɛ
æ
Difference in mels
(English minus Danish)

Figure 26

Results: English-Danish.

Fig. 26 illustrates this point for the Danish/English data which we have been examining. The mean difference in F_1 values between the English data (μ=577 mels) and the Danish data (μ=442 mels) is 135 mels; this value is indicated by the horizontal dashed line. Each of the bars shows the difference, in mels, between the F_1 values of a particular vowel in Danish and in English. The actual results of the analysis of variance are listed along with the bar graphs.

It is clear from this graph, as it was from the plotted formant frequencies (Fig. 25), that the vowels [e] and [ɛ] pattern together in the F_1 dimension, displaying greater than average differences between the languages.

One should not always accept a significant language difference at face value. If a significant pattern difference co-occurs, the mean may be affected disproportionately by one of the vowels. Thus, there may be said to be a significant difference between the means of two languages only if the vowels of one language are consistently higher (or consistently lower) than the vowels of the other. The bar graphs show, for example, that the F_1 language effect is consistent throughout all the Danish and English vowels, and thus its significance is confirmed. In contrast, the language effect in the Dutch and English data (see Fig. 29, below) which is nearly significant at just above the $p < .05$ level, is not consistent across all vowels. The apparent significance of this language effect thus reflects the distribution of a subset of the vowels rather than the language as a whole.

Dutch-German, Dutch-English and the base of articulation.

Many of the language comparisons reveal signficant pattern differences of the type described above, with or without concomitant differences in the means. In some, the individual vowels which pattern together form a natural class; this can be seen in the bar graph of Dutch and German, Fig. 27, where the F_2 values pattern together in accordance with the feature Round ([-round] vowels [i e ɛ a] on one side of the mean, [+round] vowels [y ø u o] on the other).

In other cases, however, a disparate group of vowels will pattern together, such as [ɪ a ɔ] in the Dutch-English comparison Fig. 29.

The latter case is, on the whole, a more interesting situation for the linguist. Pattern effects that are not associated with a single phonetic feature -- particularly if they are not associated with differences between languages as a whole --

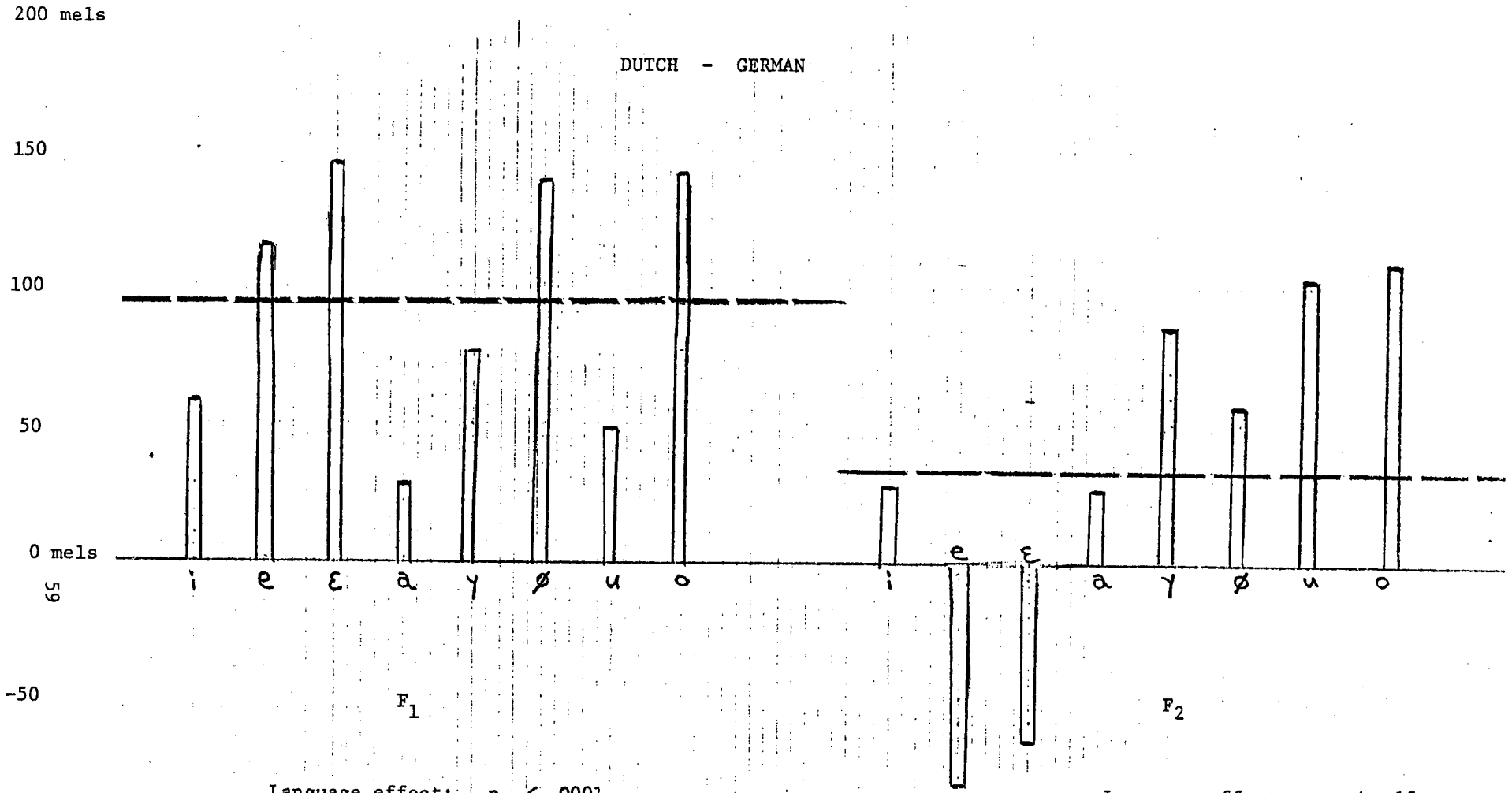

DUTCH - GERMAN
200 mels
150
100
50
0 mels
-50
-100
-150
i e ɛ a y ø u o
i e ɛ a y ø u o
F1
F2
Language effect: p < .0001
Pattern effect: p < .0001
Language effect: p < .15
Pattern effect: p < .005
Difference in mels
(Dutch minus German)

Figure 27

clearly indicate differences that cannot be ascribed solely to differences in the base of articulation, if base of articulation is to be defined as a language-specific neutral position of the tongue.[19]

Further applications of analysis of variance.

The ANOVA procedure does not transform the data as normalization procedures do; rather, it allows us to test specific hypotheses. Let us therefore begin by hypothesizing that the quality of a given vowel is predictable on the basis of the overall shape of the phonological system to which it belongs. We may then wish to test whether, for example, two languages with similar numbers of vowels will necessarily have vowels of similar quality, or whether languages with different vowel systems (both in the number and the distribution of their vowels) will have corresponding vowels of predictably different quality.

English-Dutch.

Let us first consider the example of English and Dutch. Despite certain differences in phonological shape--Dutch has a set of front rounded vowels; English has phonemic central vowels [ʌə]--the languages have similar numbers of vowels overall. Dutch has 12 contrasting vowels; English has 11. The analysis of variance indicates that differences in the mean formant frequencies between the languages are so small as to be insignificant compared to the large speaker differences within each language. This finding holds for F_1 as well as for F_2. The raw values of the six vowels common to both languages are plotted in Fig. 28, and the results of the ANOVA procedure in Fig. 29.

It is evident from Fig. 29, as we have already mentioned, that there are no differences in quality consistent across all vowels of Dutch and English. What seems to be almost a significant difference in the F_1 domain may be attributed in fact only to the vowels i,u, and ɛ. What differences do hold between the languages seem to center on the vowel [ɛ]. Interest-

---

19 The notion of base of articulation has been invoked by linguists for at least three centuries, in forms ranging from quite concrete ("the real-time physiological nature of the Basis" cited by Drachman (1975)) to abstract, and from language-particular to universal. Drachman discusses this matter at length, in particular the notion, described above, that language-specific differences in the neutral position of the tongue may account for the observed phonetic differences. It is this particular interpretation of the base of articulation which is inadequate for the data at hand. However, it remains to be determined through future investigation whether other interpretations might account for the differences more satisfactorily.

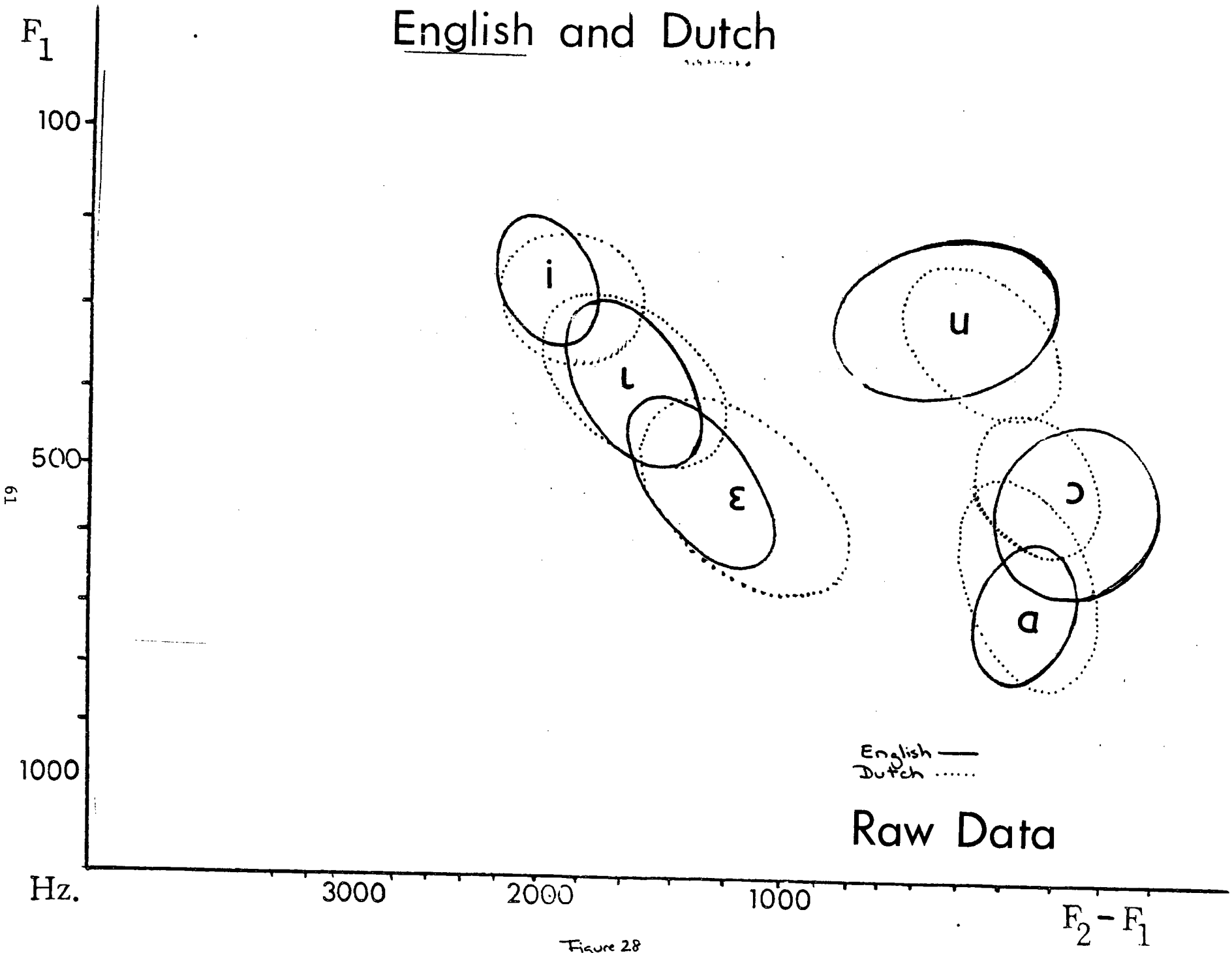
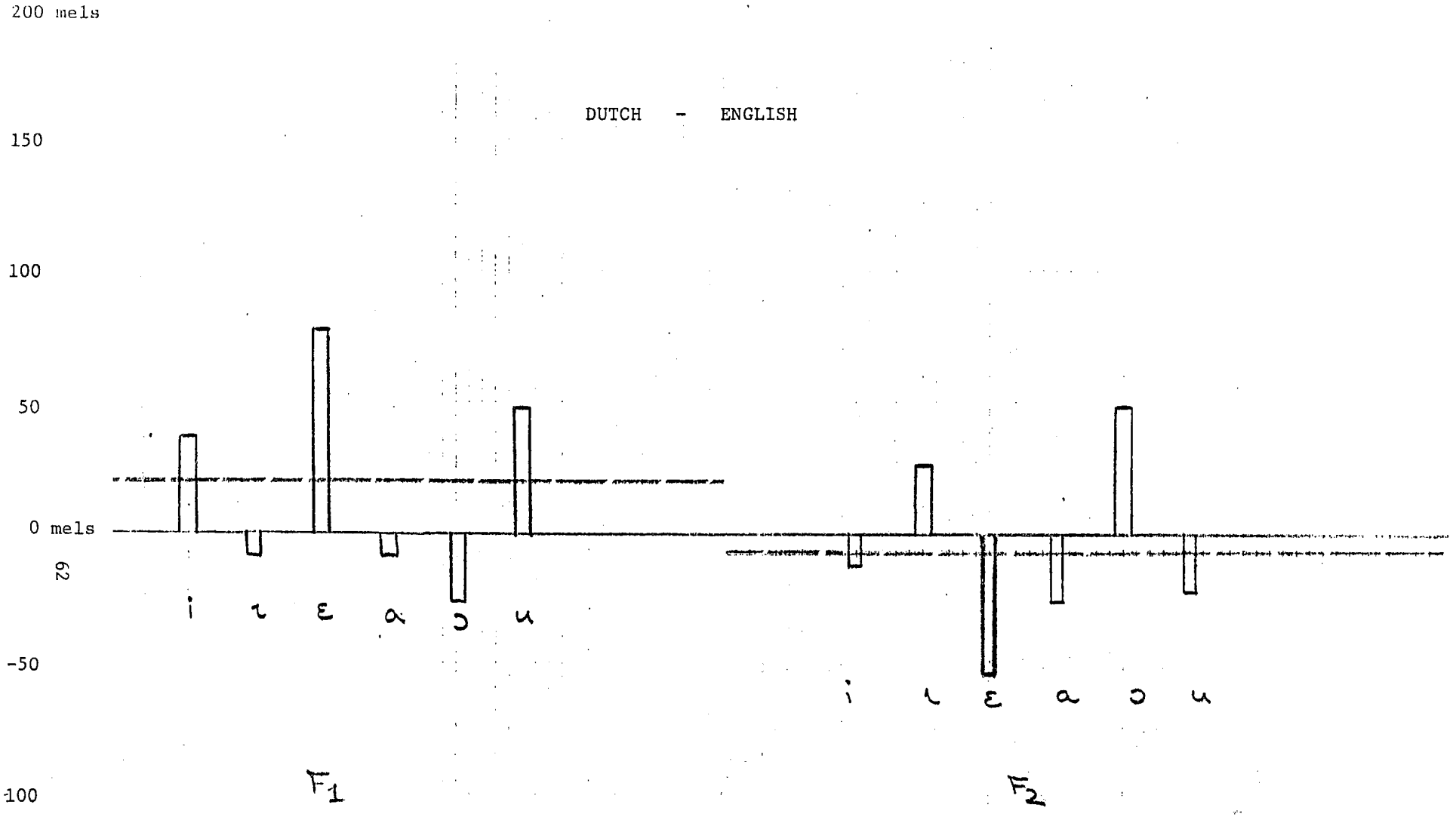

English and Dutch
F1
100
500
1000
Hz.
3000
2000
1000
F2 - F1
i
ɪ
ɛ
u
ɔ
ɑ
English
Dutch
Raw Data

Figure 28



Language effect: $p < .06$

Pattern effect: $p < .0001$

Language effect: $p < .60$

Pattern effect $p < .005$

Figure 29

Difference in mels
(Dutch minus English)

ingly enough, of all the vowels in this study [i ɪ ɛ ɑ ɔ u], [ɛ] is the only one noted by Koolhoven as having different phonetic quality in Dutch and English. (Dutch [ɛ] is "more open than the English vowel in bed.") It may be possible to find independent phonological or diachronic linguistic explanations for this difference.

German-Swedish.

The vowel inventories of German and Swedish are also quite similar to one another. German has eight vowels; Swedish, nine. Each has a set of front rounded vowels, and each also has a set of short allophonic variants with slightly different phonetic quality. (The short vowels are not taken into consideration in the present study.) Fig. 30 shows a plot of the raw values of these long vowels; Fig. 31 shows the ANOVA results. The language effect is small but consistent across all vowels in the F_1 domain ($p < .10$), while there is no consistent language effect in the F_2 domain. We may note, however, that the front rounded vowels pattern together quite unmistakeably, and that only in these vowels ([y ø]) are the Swedish F_2 values higher than the German ones. Lindau (1977) notes this difference, and attributes it to articulatory factors: "The Swedish [y] is usually made with a larger lip opening than the German [y]." In fact, it is because of this significant pattern effect that there is no language effect. Further study, then, should investigate whether any significant language or pattern differences hold for the remaining vowels. For example, the formants of Swedish, as compared to German, appear to be comparatively lower in the back vowels than in the front.

One additional detail that bears mention is the nature of German [ɛ]. This vowel is a comparatively recent innovation, introduced to preserve the distinctions that were obliterated by the umlaut process. It is used by educated speakers only, in some regions more than in others, and as it is usually learned in school, it is subject to phonetic inconsistency. This is reflected in the considerable amount of acoustic variability that characterizes this vowel (see Fig. 31). The ANOVA results and bar graph indicate that German [ɛ] and Swedish [ɛ] have different phonetic quality, particularly in the F_1 dimension. Given the special status of German [ɛ], this fact should not be viewed in the same light as the other differences between the two languages. Further investigations of the problem may in fact have to disregard this vowel.[20]

---

20 The evolution of the verb nehmen ("to take") serves to illustrate this point.

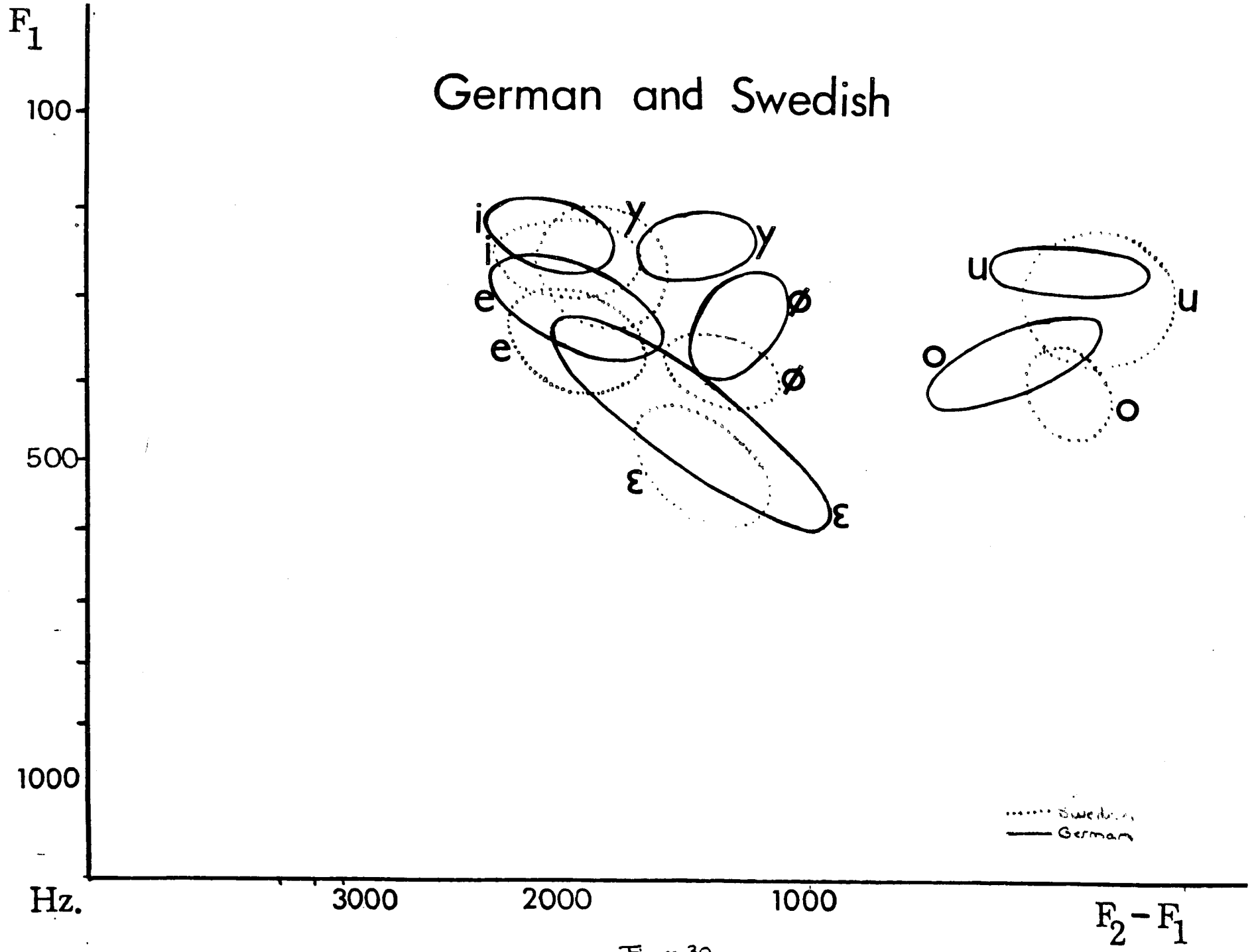
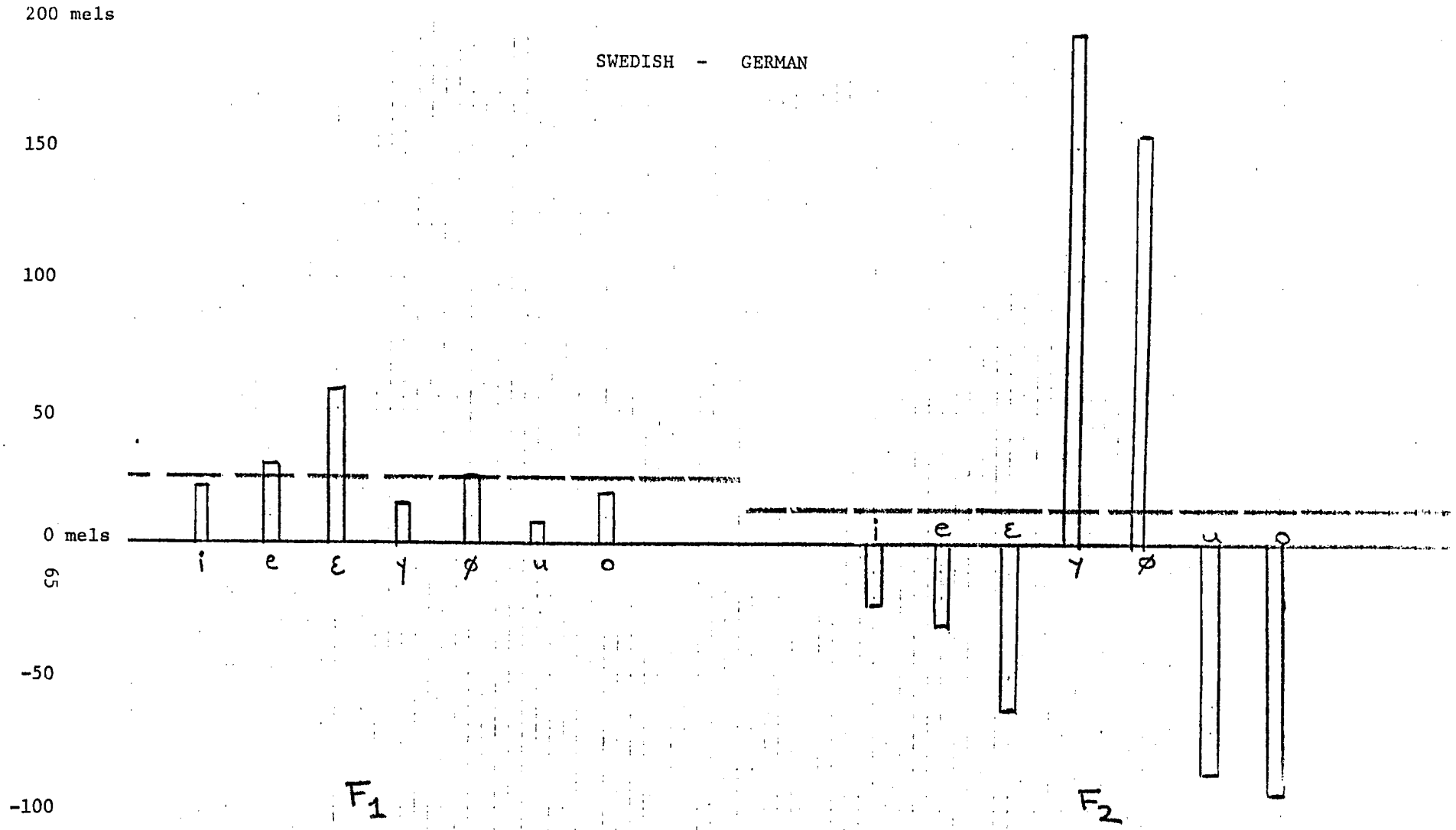

German and Swedish
F1
100
500
1000
Hz.
3000
2000
1000
F2 – F1
i
i
y
y
e
e
ø
ø
ɛ
ɛ
u
u
o
o
Swedish
German

Figure 30



Language effect: $p < 0.10$
Pattern effect: $p < 0.35$

Language effect: $p < 0.60$
Pattern effect: $p < 0.0001$

Figure 31

Difference in mels
(Swedish minus German)

Utilization of the Vowel Space

It will be recalled that Danish and English do show significant differences (cf. Figs. 25 and 26). The differences in the mean F_1 values of English and Danish may be attributed, at least in part, to the number of vowels in the system. English, in fact, has a fifth front vowel [ɪ], which falls within the gap between [i] and [e], with some overlap. According to the principle of maximal contrast (Jakobson 1941, Martinet 1955, Liljencrants and Lindblom 1972), the vowels [ɪ e ɛ] will tend to arrange themselves at more or less equal intervals between the endpoints of the system, namely [i] and [æ], while Danish [e ɛ] will also tend to arrange themselves at equal intervals between the endpoints [i] and [æ]. Similarly, Norwegian, with only three front vowels, will tend to have its [e] close to midway between [i] and [æ]. Moreover, the principle of maximal contrast (as set forth by Liljencrants and Lindblom) can be taken to imply that the endpoints of all three systems should have identical quality--specifically, that [i] should always be as high as possible, and [æ] should be as low as possible, subject only to pressures resulting from its having to be kept distinct from other vowels such as [ɑ]. However, as indicated by the raw data (Fig. 25) and confirmed by the ANOVA results (Fig. 26), the vowels transcribed as [i] and [æ] can vary considerably in quality from language to language.

It may be possible to account for the relative qualities of vowels in these languages with a modification of Liljencrants and Lindblom's general hypothesis. Maddieson (1977), for example, suggests that a principle of adequate, rather than maximal, separation governs a number of tonal phenomena in a wide range of languages. (cf. Terbeek 1977, and Lindblom 1975) In contrast to the theory of maximal separation, which claims that vowels will be maximally dispersed in phonetic space, this theory of adequate separation would claim that vowels need only be

---

	Old German		Middle High German		
indicative	neme	"I take"	ne:me		
subjunctive	nami		na:mi	→	næme
			umlaut		

The product of the umlaut process was [næme]. Some dialects set up a new phoneme [ɛ:] in order to disambiguate the two forms, while others simply collapsed [æ] and [e] as [e].

separated by a sufficient interval. Thus the endpoints of a set of five front vowels, as in English, would tend to be more widely separated from one another than would the endpoints of a set of three or four front vowels, as in Norwegian or Danish. Both theories, however, make similar claims about the relative placement of the intermediate vowels in a given system: specifically, that they will fall at roughly equal intervals between the endpoints. According to the theory of adequate separation, then, the vowel systems of Danish and English should assume relative configurations as shown in Figure 32.

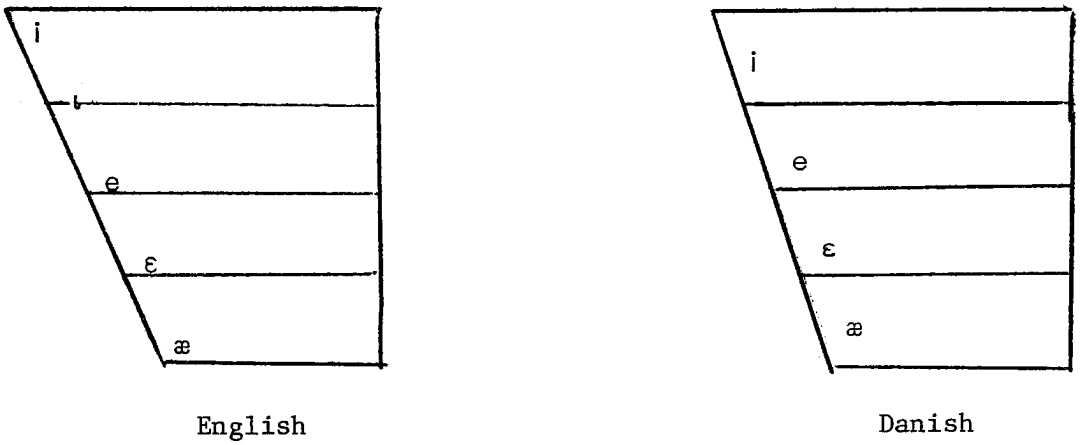


Figure 32.

If we compare these predictions to the actual data for English and Danish, we find that the theory of adequate separation accurately predicts the relative placement of [e ɛ æ], which are lower (=higher F_1) in the former language than in the latter, but it makes the wrong prediction in the case of [i], which should be higher.

By the same token, the theory of adequate separation predicts that a language such as Norwegian, which contrasts fewer vowel heights than either English or Danish, would have vowels in the relative configuration shown in Figure 33.

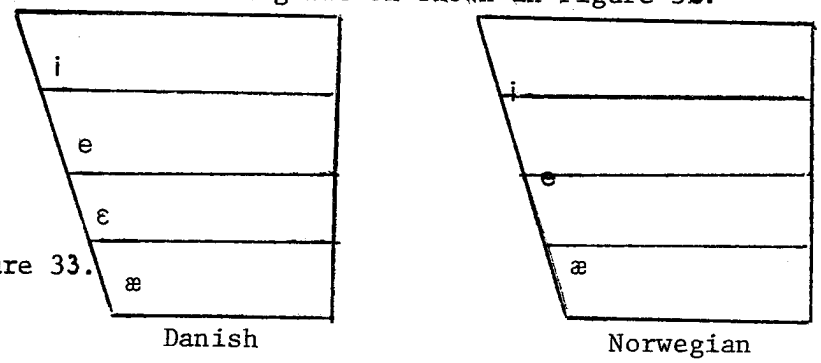


Figure 33.

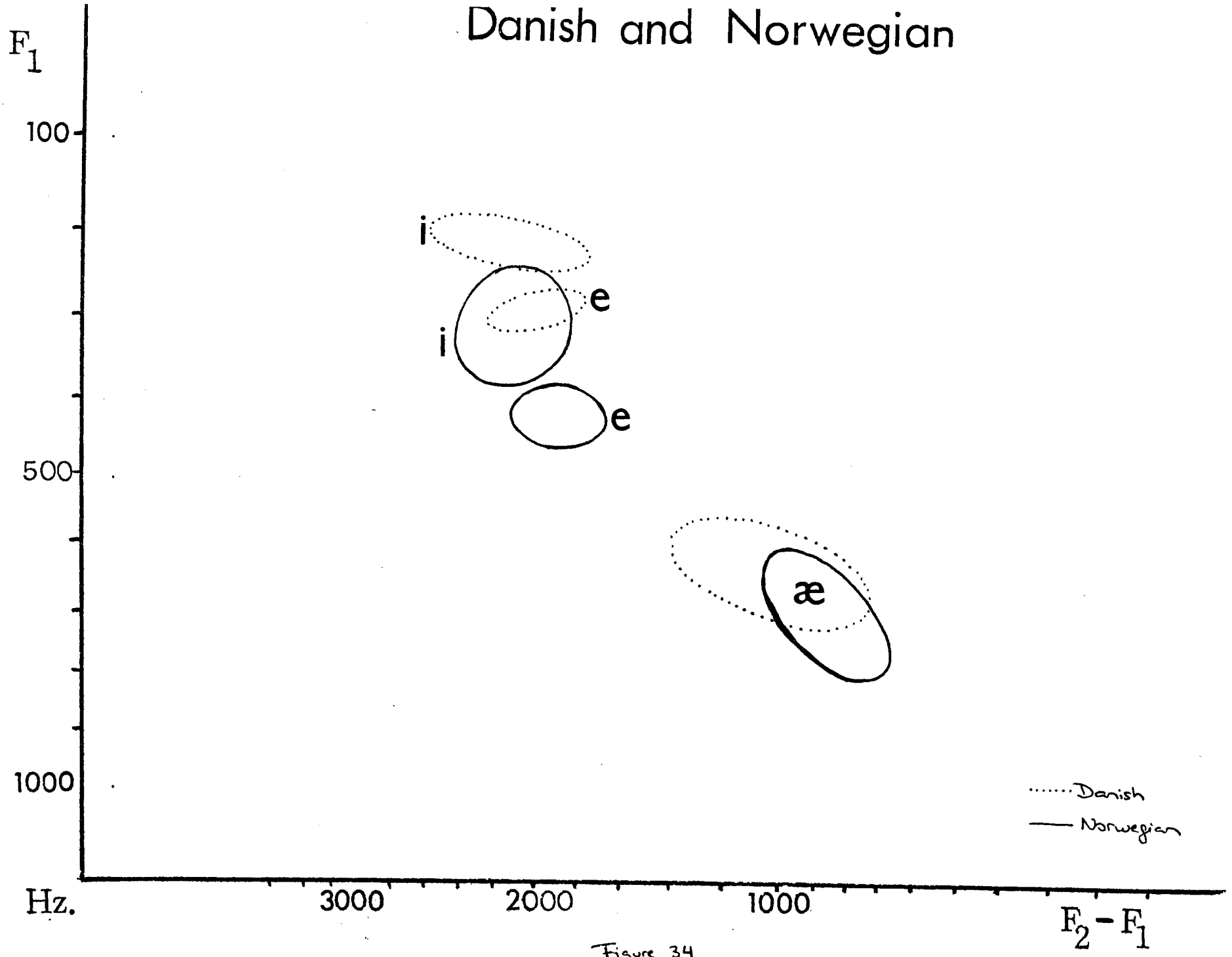

Danish and Norwegian
F1
100
500
1000
i
e
i
e
æ
Danish
Norwegian
Hz.
3000
2000
1000
F2 - F1

Figure 34

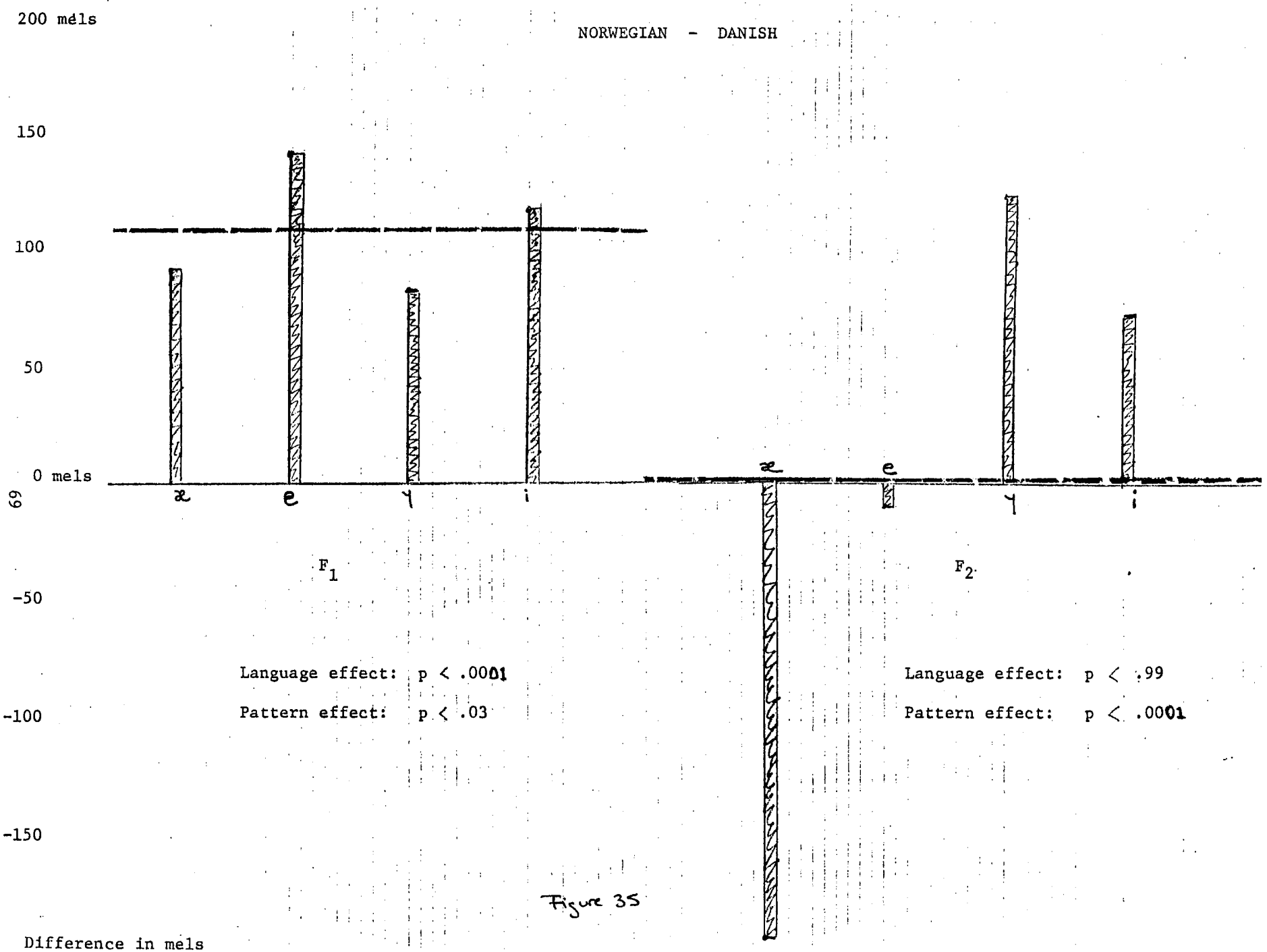

NORWEGIAN - DANISH
200 mels
150
100
50
0 mels
-50
-100
-150
æ
e
ɣ
i
æ
e
ɣ
i
F1
F2
Language effect: p < .0001
Pattern effect: p < .03
Language effect: p < .99
Pattern effect: p < .0001
Difference in mels
(Norwegian minus Danish)

Figure 35

We may note from the bar graph of Norwegian and Danish (Fig. 35) that the largest share of the significant difference between the means of Norwegian and Danish is indeed to be found in the vowel [e]. This is in accordance with both of the "separation" theories, in light of the fact that Danish has [ɛ] in its vowel inventory, whereas Norwegian does not. The theory of adequate separation correctly predicts that Norwegian [i] will be lower than Danish [i], but wrongly predicts that Norwegian [æ] will be higher.

The relationships between comparable vowels in the three languages, as predicted by the theory of adequate separation, are indicated schematically by the dotted arrows in Figure 36. The theory as presently formulated claims that all intervals between the vowels of a given language should be equal.

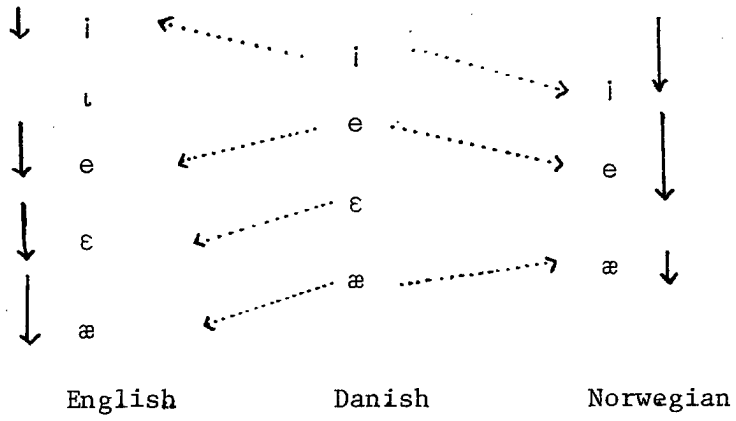


Figure 36. Relative heights of selected vowels in three languages.

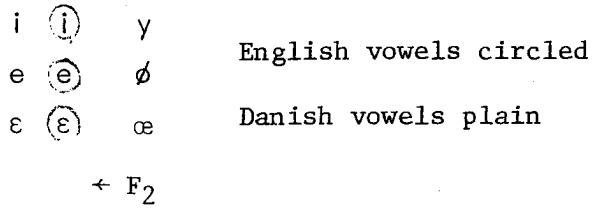
The solid arrows in the figure represent the observed relationships, and their relative length represents the relative magnitude of the difference.

The theory of adequate separation predicts that in a language with five front vowels [i ɪ e ɛ æ], as in English, the [i] should be higher and the [æ] lower than the corresponding vowels in a language with only four, [i e ɛ æ], as in Danish. As we have seen, only the latter holds true. Conversely, in a system with three front vowels, [i e æ], as in Norwegian, the [i] should be lower and the [æ] higher than the corresponding Danish vowels. Here, only the former holds true. What the two wrong predictions have in common, however, is the fact that they seem to be minimized in the actual data. English and Danish [i] are less different than are English and Danish [æ] (the latter correctly predicted by the theory of adequate separation); Norwegian and Danish [æ] are less different than are Norwegian and Danish [i] (again, the latter correctly predicted).

It should be noted that the theory of maximal separation, as formulated by Liljencrants and Lindblom, makes the wrong prediction in all four cases. Moreover, the differences in phonetic quality across languages also seems to constitute an argument against the quantal theory proposed by Stevens (1972).

An alternative explanation may, of course, be that the Liljencrants and Lindblom model correctly predicts the relationships between most vowel systems, but that the Danish system constitutes an exceptional case. As will be recalled, the vowels of Danish are not normally distributed around the mean, and the mean itself is significantly higher than the means of other Germanic languages. In checking the results of their model, Liljencrants and Lindblom make use of sources which are "based on ... phonemic analyses, and often fail to comment on fine phonetic detail." (Liljencrants and Lindblom, p. 845) It is therefore possible that the phonetic transcriptions utilized by Liljencrants and Lindblom conceal other exceptional cases of the type we have found in Danish.

The F_2 results are somewhat different than expected on phonological grounds, in the case of Danish and English. The principle of maximal contrast would predict that the front unrounded vowels of a language which also has a set of front rounded vowels (e.g. Danish) would be more peripheral than their counterparts in a language without front rounded vowels (e.g. English).



While [e ɛ æ] follow this principle, [i] does not; English [i] actually has a slightly higher F_2 value than Danish [i]. (See Fig. 25) This anomaly is most probably due to the omission, in the ANOVA study and in the plots of the raw data, of the third formant. F_3 is "a decisive factor in the auditory impression of [i], ... especially so in languages which, like Danish and Swedish [and unlike English], have a very close [i] with a relatively low and weak F_2 and a high and strong F_3." (Fischer-Jorgensen, 1972) We may recapture a good deal of this salient information by applying F_2' to the raw data; these new results are shown in Fig. 37. Furthermore there may be more appropriate methods of integrating the values of F_2 and F_3 than Fant's F_2', and these may be better able to separate the high vowels of English and Danish in the predicted direction. Presumably, the analysis of variance would reveal a significant language effect if an effective second formant such as F_2' rather than F_2 alone, were chosen as a dependent variable.

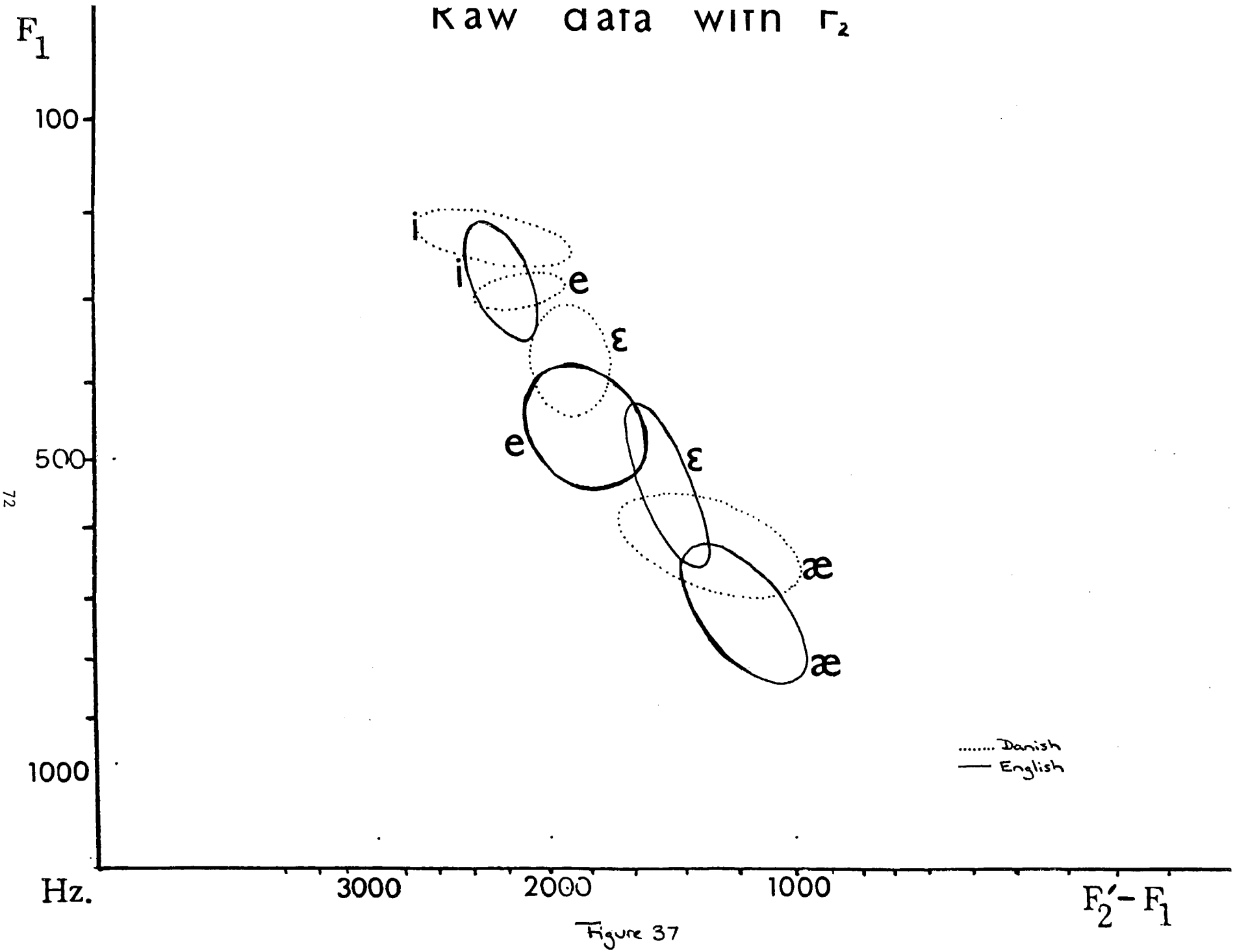

Raw data with F2
F1
100
500
1000
i
i
e
ɛ
e
ɛ
æ
æ
..... Danish
—— English
Hz.
3000
2000
1000
F2′ – F1

Figure 37

The pattern differences between the F_2 values in Danish and those in English may be attributed, at least in part, to the fact that these Danish vowels are all phonologically [+long], whereas the corresponding English vowels are further distinguished by the(phonologically redundant) feature [± long]. In fact, in the case of the Danish front vowels, only very minor qualitative differences accompany differences in vowel length. Fig. 25 shows that the Danish vowels, separated from each other by their quality alone, quite noticeably avoid overlap, while the English vowels, which can be disambiguated on the basis of their length, if need be, overlap rather freely.

Multivariate Analysis

One additional point to be kept in mind is that the significance of vowel quality differences cannot be determined validly on the basis of univariate analysis alone. Non-significance in the case of either or both formants is not necessarily indicative of non-significance for the bidimensional formant space. Accordingly, multivariate analyses were performed on the data, with F_1 and F_2 serving as simultaneous dependent variables.

In the multivariate analysis of the Danish-English data, highly significant differences ($p < .0001$) were found to hold between the means of the languages, as well as between the patterns. The same held true for the Danish-Norwegian ($p < .0001$) data. In the Dutch-English ($p < .0642$) and the German-Swedish ($p < .2864$) cases, however, the means were not significantly different. All these results are comparable to those predicted independently by F_1 and F_2.

In summary, the data which we have examined shows that there are differences between languages which cannot be predicted on the basis of their phonological systems alone. Knowledge of the languages themselves, rather than their systems, is needed to account for the differences between Dutch and English [ɛ], for example, or between German and Swedish [y]; these differences are not associated with any overall differences between the languages. Such evidence points to the inadequacy of any theory which seeks to attribute all such phonetic differences to differences in the base of articulation. It is also direct counter-evidence to a quantal theory of vowel quality, and it offers problems for a theory predicting tendencies towards maximal perceptual separation. It is most in accord with a theory which makes only the weak prediction that the vowels of a language will be adequately separated.

LANGUAGE		i	e	ɛ	a	æ	ɑ	ɣ	ø	œ	ʉ	ɞ	ɐ		TOTALS
DUTCH	R	10690	14859	21815	22462		15940	12864	11196	15751	12302	10053	9952	R	157884
	S	3034	6427	7728	14168		13458	7879	4926	8576	7693	8595	8545	S	91029
	F	3377	5843	6773	10386		9236	"	"	"	4838	6515	6126	F	74475
	P	5453	6324	7835	16832		15852	7577	5375	9273	11259	9431	8686	P	103897
GERMAN	R	5183	8482	16116	9365			4423	5843		4382	7389		R	61183
	S	432	1394	5717	3367			1482	1244		1017	2239		S	16892
	F	414	1130	4008	2243			"	"		508	1784		F	12813
	P	1046	3110	11083	4467			3179	4146		5504	8884		P	41419
SWEDISH	R	6675	7539	8293			4825	6192	10373		8906	4008		R	56811
	S	1558	2686	2111			6503	2071	4860		6248	6946		S	32983
	F	1548	2010	1932			3223	"	"		3798	3826		F	16337
	P	3581	3188	5284			4580	3223	3864		6430	5551		P	35701
NORWEGIAN	R	8476	3760			8708	4209	6635	5733					R	37521
	P	2714	2205			5529	4618	1976	3496					P	28674
DANISH	R	4043	2143	5881		10568		2591	1102	7351	2576	73	1592	R	37920
	S	955	1244	753		2874		565	691	3562	2814	1891	980	S	16329
	F	961	967	911		2613		"	"	"	1583	1093	794	F	13740
	P	1709	1407	1771		6446		1658	791	7037	3455	1551	2827	P	28652
ENGLISH (Peterson/ Barney)	R	7596		12833		14090	8444				20988		17241	R	81192
	S	3015		5648		5956	9226				9057		14174	S	47076
	F	"		"		"	"				"		"	F	"
	P	4429		9880		11655	5654				17140		12063	P	60821
CALIFORNIA ENGLISH (Kahn)	R	6047	10178	5576		7791	14027				12902	5428		R	61949
	S	3769	6031	5827		6575	18346				4052	3166		S	47766
	F	"	"	"		"	"				"	"		F	"
	P	7049	8959	6330		5130	14451				12666	5277		P	59862

Table I. Areas of the Ellipses Produced by: Raw Data only (R)
SD Normalization (S)
SD Norm' (without Front Rounded Vowels) (F)
PARAFAC Normalization (P)

REFERENCES

Anderson, S.R. 1974. The Organization of Phonology. New York: Academic Press.

Bernstein, J. 1977. Vocoid psychoacoustics, articulation, and vowel phonology. Phonetics Laboratory, The University of Michigan.

Bredsdorff, E. 1958. Danish, An Elementary Grammar and Reader. Cambridge: Cambridge University Press.

Broad, D. and Wakita, H. 1977. Piecewise-planar representation of vowel formant frequencies. J. acoust. Soc. Am. 62: 1467-1473.

Chen, M. 1973. On the formal expression of natural rules in phonology. Journal of Linguistics 9:223-249.

Chiba, T., and Kajiyama, M. 1941. The Vowel: Its Nature and Structure. Tokyo: Tokyo-Kaiseikan Publishing Co., Ltd.

Chomsky, N. 1964. Current issues in linguistics. In J.A. Fodor and J.J. Katz, The Structure of Language: Readings in the Philosophy of Language. Englewood Cliffs, N.J.: Prentice-Hall.

Chomsky, N., and Halle, M. 1968. The Sound Pattern of English. New York: Harper and Row.

Cleveland, T.F. 1976. The Acoustic properties of voice timbre types and the importance of these properties in the determination of voice classification in male singers. Speech Transmission Laboratory Quarterly Progress and Status Report 1/1976:17-29. Royal Institute of Technology, Stockholm.

Davis, S.B. 1976. Computer evaluation of laryngeal pathology based on inverse filtering of speech. SCRL Monograph No. 13.

Drachman, G. 1973. Phonology and the base of articulation. Die Sprache 19:1-19.

Fant, G. 1959. Acoustic description and classification of phonetic units. Ericsson Technics 15:3-108.

Fant, G. 1965. Formants and cavities. Proc. Fifth Int. Cong. of Phonetic Sci., pp. 120-141. Basle: Karger.

Fant, G. 1966. A Note on vocal tract size factors and non-uniform F-pattern scalings. Speech Transmission Laboratory Quarterly Progress and Status Report 4/1966 Royal Institute of Technology, Stockholm.

Fant, G., Henningsson, G., and Stålhammar, U. 1969. Formant frequencies of Swedish vowels. Speech Transmission Laboratory Quarterly Progress and Status Report 4/1969. 26-31. Stockholm: Royal Institute of Technology.

Fischer-Jørgensen, E. 1972. Formant frequencies of long and short Danish vowels. Studies for Einar Haugen, ed. by E.S. Firchow et al. The Hague: Mouton.

Gamnes, H. 1965. En Akustisk undersökelse og maling av formantfrekvenser i öst-norske vokaler. Unpublished thesis, University of Oslo.

Gerstman, L.H. 1968. Classification of self-normalized vowels. IEEE Trans. Audio Electroacoust. AU-16, pp.78-80.

Halle, M. 1959. The Sound Pattern of Russian. The Hague: Mouton.

Harshman, R. 1970. PARAFAC: Models and conditions for an "explanatory" multi-modal factor analysis. Working Papers in Phonetics No. 16, Phonetics Lab, UCLA.

Harshman, R., Ladefoged, P., and Goldstein, L. 1977. Factor analysis of tongue shapes. J. acoust. Soc. Am. 62:693-707.

Haugen, Einar 1935. Beginning Norwegian. New York: Appleton Century Crofts.

Haugen, Einar 1976. The Scandinavian Languages. Cambridge, Mass.: Harvard University Press.

Heike, G. 1964. Zur Phonologie der Stadtkolner Mundart. Marburg: N.G. Elwert.

Hyman. L.M. 1975. Phonology: Theory and Analysis. New York: Holt, Rinehart, and Winston.

Jacobson, L.C. 1977. Dho Luo vowel harmony. Unpublished manuscript, U.C.L.A.

Jakobson, R., Fant, G., and Halle, M. 1951. Preliminaries to Speech Analysis. Cambridge, Mass.: M.I.T. Press.

Jakobson, R. 1941. Kindersprache, Aphasie, und allgemeine Lautgesetze. Reprinted in Selected Writings I, 328-401. The Hague: Mouton 1962.

Jones, D. 1917. An English Pronouncing Dictionary. London: J.M. Dent and Sons.

Jones, D. 1956. Outline of English Phonetics. Eighth edition. Cambridge; W. Heffer and Sons.

Joos, M. 1948. Acoustic Phonetics. Supplement to **Language** 24.

Jørgensen, H.P. 1969. Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen. **Phonetica** 19:217-245.

Kahn, D. 1977. Near-perfect identification of speaker-randomized vowels without consonantal transitions (Abstract). J. acoust. Soc. Am. 62; S-101.

Koolhoven, H. 1968. Dutch. London: The English Universities Press.

Ladefoged, P. 1964. A Phonetic Study of West African Languages. Cambridge: Cambridge University Press.

Ladefoged, P. 1967. Three Areas of Experimental Phonetics. London: Oxford University Press.

Ladefoged, P. 1971. Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press.

Ladefoged, P. 1975. A Course in Phonetics. New York: Harcourt Brace Jovanovich.

Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. 1972. An auditory-motor theory of speech production. UCLA Phonetics Lab, Working Papers in Phonetics 22:48-75.

Lieberman, P. 1977. Speech Physiology and Acoustic Phonetics: An Introduction. New York: Macmillan.

Liljencrants, J., and Lindblom, B. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. Language 48:839-862.

Lindau, M. 1975. [Features] for vowels. UCLA Phonetics Lab, Working Papers in Phonetics No. 30.

Lindau, M. 1977. Vowel features. UCLA Phonetics Lab, Working Papers in Phonetics 38: 49-81.

Lindau, M., and Wood, P. 1977. Acoustic vowel spaces. UCLA Phonetics Lab, Working Papers in Phonetics 38: 41-48.

Lindblom, B. 1975. Experiments in sound structure. Paper Paper read at the Eighth International Congress of Phonetic Sciences. Leeds, August 1975.

Lindblom, B., Lubker, J., and Gay, T. 1977. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation (Abstract). J. acoust. Soc. Am. 62:S15.

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. J. acoust. Soc. Am. 49:606-608.

Maddieson, I. 1976. The intrinsic pitch of vowels and tones in Foochow. UCLA Phonetics Lab, Working Papers in Phonetics 33:191-202.

Maddieson, I. 1977. Tone loans: a question concerning tone spacing and a method of answering it. UCLA Phonetics Lab, Working Papers in Phonetics 36:49-83.

Martinet, A. 1955. Économie des changements phonétiques. Berne: Francke.

McClean, R.J. 1969. Swedish, a Grammar of the Modern Language. London: The English Universities Press.

Moulton, Wm. G. 1962. The Sounds of English and German. Chicago: University of Chicago Press.

Nearey, T. 1977. Phonetic feature systems for vowels. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Nielsen, N. and Hjorth, P.L. 1971. Dansk sproglaere for svenskere. Lund: Gleerups Förlag.

Nordström, P.-E., and Lindblom, B. 1975. A normalization procedure for vowel formant data. Eighth International Congress of Phonetic Sciences, Leeds.

Ohala, J. 1971. The role of physiological and acoustic models in explaining the direction of sound change. Univ. of Calif., Berkeley, Project on Linguistic Analysis 15:25-40.

Öhman, S. 1966. Coarticulation in VCV utterances: spectrographic measurements. J. acoust. Soc. Am. 39:151-168.

Öhman, S. 1967. Numerical model of coarticulation. J. acoust. Soc. Am. 41:310-320.

Peterson, G., and Barney, H.L. 1952. Control methods used in a study of vowels. J. acoust. Soc. Am. 20:528-535.

Pétursson, M. 1976. Quantität und vokalische Klangfarbe im modernen Isländischen. Zeitschrift für Sprachwissenschaft und Kommunikationsforschung 29: 49-63.

Pols, L., Tromp, H., and Plomp, R. 1973. Frequency analysis of Dutch vowels from 50 male speakers. J. acoust. Soc. Am. 53:1093-1101.

Rice, L. 1971. A new line analog speech synthesizer for the PDP-12. UCLA Working Papers in Phonetics 12:58-75.

Spore, P. 1965. La Langue danoise. Copenhagen: Akademisk Forlag.

Stevens, K. 1972. The quantal nature of speech: evidence from articulatory-acoustic data. In P.B. Denes and E.E. David Human Communication: A Unified View.

Stevens, K., and Blumstein, S. 1975. Quantal aspects of consonant production and perception: a study of retroflex stop consonants. Journal of Phonetics 3:215-234.

Terbeek, D. 1977. Some constraints on the principle of maximum perceptual contrast between vowels. CLS 13.

Uldall, H.J. 1933. A Danish Phonetic Reader. London: University of London Press.

Wakita, H. 1977. Normalization of vowels by vocal tract length and its application to vowel identification. IEEE Trans. Acoust. Spch. Signal Process. ASSP-25: 183-192.

Walshe, M. O'C. 1965. Introduction to the Scandinavian Languages. London: André Deutsch.

Wang, Wm., and Fillmore, C. 1961. Intrinsic cues and consonant perception. Journal of Spch. and Hearing Research 4:130-136.

Williamson, K. 1977. Multivalued features for consonants. Language 53:288-296.