UCLA
Working Papers in Phonetics

Title
WPP, No. 41

Permalink
https://escholarship.org/uc/item/5fb7f1kq

Publication Date
1978-04-15

eScholarship.org Powered by the California Digital Library
University of California

UCLA 1978

WPP41

vowels tones lips etc

etc lips tones vowels

vowels tones lips etc

etc lips tones vowels

vowels tones lips etc

WPP41

UCLA

1978

UCLA Working Papers in Phonetics 41

April, 1978

The UCLA Phonetics Laboratory Group

Steve Anderson
Ava Berinstein
Cathe Browman
Ron Carlson
Pat Coady
Vanna Condax
Sandy Ferrari Disner
Vicki Fromkin
Vivian Flores
Linda Galloway
Manuel Godínez
Louis Goldstein
Steve Greenberg
Jean-Marie Hombert
Peter Ladefoged
Wendy Linker
Ian Maddieson
Willie Martin
Breyne Moskowitz
Jonas Narty
George Papçun
Lloyd Rice
Renée Wellin
Anne Wingate
Jim Wright
Eric Zee

As on previous occasions, the material which is presented here is simply a record for our own use, a report as required by the funding agencies, and a preliminary account of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS grant NS 9780
USPHS grant NS 12303
NSF grant SOC75-7158
and the UCLA Department of Linguistics

Correspondence concerning this series should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles, California 90024

*A Comparative study of Some Romance Vowels**

Manuel Godínez Jr.

[Paper presented at the 4th Annual Minnesota Regional Conference
on Language and Linguistics, May 1978]

Studies in Hispanic and Luso-Brazilian linguistics have traditionally described vocalic phenomena in terms of the *tongue arching model of vowel production.*[1] Such an articulatory model conceives of tongue-height as playing a predominantly prominent role in the determination of vowel quality. Following this approach, depiction of vowel qualities in the published literature has generally been made in the form of a triangular arrangment of points. The relative heights of these points are said to be a direct function of the tongue height required for a particular vowel quality. Recent experimental evidence has shown, however, that the tongue arching model is misleading and has served to obscure the actual articulatory parameters involved in vowel production.[2] It is suggested in this paper that vowel sounds be viewed as acoustically specifiable entities resulting from corresponding vocal tract configurations. The representations of vowels in graphic form are thus to be thought of as depicting *relative differences in acoustic vowel quality* and not particular tongue positions.

Since the mid-nineteenth century the tongue arching model has enjoyed widespread use in the description of vowels. Despite its proved unsound physiological basis, as shown notably by the early work of Russell (1928), it still maintains a prominent position in many contemporary phonetic-phonological studies.[3] However, an extensive study of published x-ray tracings of many languages by Wood (1975) has shown that many of the articulatory positions ascribed to particular vowels by the tongue-arching model are clearly in error. An interesting example is the case where the tongue height of the vowel [a] in Spanish is found to lie *above* that of the vowel [o] (See Figure 1.)

Figure 1. A comparison of tongue height of the vowels [a], [ɔ] and [o] in Spanish from published x-ray tracings. (After Wood, 1975).

---

*I would like to express my gratitude to Professor Peter Ladefoged and Sandy F. Disner for their assistance in seeing this project through. I am indebted to Professor Isabel López de Herwig for her kind help in locating the Brazilian subjects for this experiment. I am also grateful to George Papçun for permitting me to use his extensive corpus on Mexican Spanish. Special thanks to Professor Ana M. Borzone de Manrique of the Laboratório de Inventigaciones Científicas (Argentina) for sending me the material on Argentine Spanish.

With the advent of the sound spectrograph it was believed that a one-to-one correspondence could be established between the articulatory maneuvers of vowel production and the acoustic signal.[5] Assuming that the tongue height descriptions of vowels were correct, it was easy to conclude that the locations of formants on spectrograms were directly related to tongue positions. The first formant was thought to be inversely related to vertical height and the second formant was believed to be related to the front-back position of the tongue.

In actuality, however, speakers are striving to create a constant vocal tract configuration for a given vowel. The speaker does this by making the articulatory adjustments in: 1) vocal tract length 2) position of the tongue 3) configuration of the lips.[6]

These parameters are utilized in different ways by different speakers to varying degrees in producing a particular vowel quality.[7] Thus the acoustic signal represented on a spectrogram is the result of the total supralaryngeal vocal tract configuration as produced by the speaker for a given vowel. It would seem to be more accurate, then, to conceive of vowel production in articulatory terms as striving to produce a constant vocal tract configuration rather than the tongue positions predicted by the tongue-arching model.

It is essential to point out also that when an acoustic analysis is being carried out, we are attempting to capture the acoustically significant perceptual contrasts made by speakers of the language under investigation. This point has not usually been clearly stated (e.g. Azevedo Filho, 1968) in the published literature on Spanish and Portuguese phonetics cited above.

We have briefly discussed how the *tongue-arching model* of vowel production has affected the description of vocalic phenomena. It was seen that studies based on this model are inaccurate descriptions of the actual articulatory parameters involved in vowel production. It was also pointed out that constant vocal tract configurations result in particular vowel qualities, and that the acoustic signals generated by these vocal tract shapes result in a particular vowel quality perceived. We are able to measure these acoustic outputs by use, for example, of the sound spectrograph. We will now examine the extent to which Hispanic and Luso-Brazilian linguists have availed themselves of this instrument in their descriptions of vocalic phenomena in the period since spectrographic analysis came into use in the late 1940's.[8]

II. Previous experimental work:

One of the earliest attempts to apply spectrographic analysis to the vowels of Spanish was an unpublished study by Skeleton (1950), *A Spectrographic Analysis of Spanish Vowel Sounds*. His study was based on recordings of 20 male subjects representing the Spanish of various parts of Southwestern United States, Latin America and Spain. The analysis was based on words uttered in isolation with vowels in 1) primary, 2) secondary and 3) unstressed positions. The vowels in the various phonetic contexts were then plotted on formant charts with F1 on the abscissa and F2 on the ordinate.[9]

Another study by Cárdenas (1960) was based on two speakers: the author himself and one female subject. The vowels in this study were also examined on the basis of position of stress. In addition the author included vowel tokens that were diphthongal in nature. The formant frequency values were calculated and plotted on formant charts.

A major pilot study entitled, *Comparing the Phonetic Features of English, French, German and Spanish* was undertaken by Delattre (1964). The section of this study that we are interested in is Part II, p. 45-70, where the vocalic features of the individual languages are characterized and compared with each other. Through this work we are able to see that the labels traditionally used as high vs. low and front vs. back become meaningful statements of vowel quality. Such labels are useful insofar as we refer to the relative positions of the vowels occupying the acoustic space. An acoustic analysis, as was pointed out earlier, is an attempt to relate the objective physical dimensions of the acoustic signal that serve to define the perceptual qualities that serve to distinguish vowels. The vowel points depicted in formant charts correlate pretty well with trained phoneticians' auditory impressions.

There have also been several other recent acoustic studies on the vowels of Spanish. Skelton (1969) presents a summary of his earlier work reformulated in more recent terms. A study on the vowels of Argentine Spanish comparing those with American English was done by Guirao and Borzone de Manrique (1972).[10] A preliminary experimental study by Godínez (1977a) compared the vowel qualities of Mexican and Peninsular Spanish dialects.[11]

Experimental phonetic work on the vowels of Portuguese has, however, not been as active as it has been for that of Spanish. An early auditory based study on European Portuguese vowels was done by Strevens (1954). He compared the Portuguese vowels with those of the cardinal vowel system. Head (1965a) presents a study based on an acoustic distinctive feature analysis of the phonemic systems of Rio de Janeiro and Lisbon speech.[12] A brief acoustic study by Heike (1967) compared the vowel qualities of Porto Alegre and Coimbra speech of two subjects. Godínez (1977a) examined the vowels of Brazilian Portuguese. In this preliminary study the speech of 9 male speakers were sampled representing the areas of São Paulo, Rio de Janeiro, Espírito Santo and Bahia.[13] The recent work of Delgado Martins (1973, 1975a, 1975b, 1977) undoubtedly represents the present stage of experimental phonetic work in Luso-Brazilian linguistics. In Delgado Martins (1973), "Uma análise acústica das vogais orais em português", we are presented with an acoustic study of Lisbon Portuguese vowels. The vowels of 8 male speakers were uttered in selected phonetic frames. The formant frequency values were measured spectrographically and were then submitted to a statistical analysis. Her analysis of the data revealed the formant frequencies for the vowels of Lisbon Portuguese, shown in Table 1.

VOWEL	i	e	ɛ	ɐ	a	ɔ	o	u
F1	294.	403.	501.	511.	626.	531.	426.	315.
F2	2343.	2084.	1893.	1602.	1326.	994.	864.	678.
F3	2985.	2589.	2565.	2559.	2440.	2407.	2414.	1662.

Table 1. Final mean values (Hz) of formant frequencies of 8 speakers representing Lisbon Portuguese. (After Delgado Martins, 1973).

With the exception of Delgado Martins (1973), there are several questions relating to the methodological techniques employed in the studies discussed above and the validity of the general conclusions drawn. Firstly, how valid are the conclusions drawn from an experiment with a relatively small sample population?

Presumably we are dealing with members of speech communities that share common behavioral tendencies. That is, as the result of the interaction between the variables of age, sex, and linguistic experience, speakers will tend to manifest common underlying phonetic habits that identify them as belonging to a particular sector in the speech community. That is what is meant by the use of the term *dialect* - systematic qualitative differences used by a group of speakers not used by another group of speakers, e.g. as in the use of /θ/ for the graphemes "z", or "c" before "i" or "e" in Madrid Spanish as opposed to the use of /s/ in Mexican Spanish in similar circumstances. We are not interested, then, in the behavior of an individual, but rather of how members of a group tend to behave as a whole. The experimenter will have to keep this in mind when selecting a representative sample from the parent population. Another question involves the control of variables such as a) the range in age of the subjects b) sex and c) the linguistic experiences of the subjects used in an experiment. This latter aspect includes, for example, such factors as place of birth, place(s) of residence and socioeconomic status.

Keeping these factors in check, size of sample population and subjects' background, the experimenter will then be able to make valid generalizations from the results of the analyzed data. The validity of the results by Skelton (1950, 1969) Cárdenas (1960), Delattre (1965), for example, were necessarily mitigated due to uncontrolled variables in the experiments.

III. The Experiment:

A report on an ongoing acoustic investigation on the vowels of the Romance languages will now be presented. The purpose of this undertaking is to determine on a scientific basis the underlying phonetic parameters that characterize the vowels of the Romance languages. The experimental techniques utilized in this study and ones tobe used in further analyses, will not only enable us to see the features that identify the vowels of each individual language, but will also permit us to make objective statements concerning the similarities and differences of comparable vowels between each language. The present discussion is a preliminary report of the research in progress.

a) Recordings:

The material presented here represents recorded data on Mexican, Argentine and Peninsular dialects of Spanish and Brazilian Portuguese.

The recorded data on Mexican Spanish consists of 6 speakers representing the speech of Tijuana, Mexico. Subjects of this group were born and have always lived in Tijuana. Four subjects were recorded representing the Argentine Spanish of Buenos Aires.[14] The six subjects of Peninsular Spanish are representative of various geographical locations of Spain. The Brazilian Portuguese data represents the speech of São Paulo (4 , speakers) Rio de Janeiro (2, speakers), Espírito Santo (2, speakers), Bahia (1, speaker) a total of 9 speakers. All subjects recorded were male.

To facilitate reference to speakers,the recorded subjects were divided into groups. Group I represents speakers of the Mexican Spanish dialect, Group II the Argentine Spanish dialect and Group III represents speakers of the Peninsular Spanish dialect. The Brazilian Portuguese subjects are Group IV.

To elicit the vowel phonemes of Spanish, Groups II and III were requested to read the sentence "Digo la palabra____ otra vez" substituting the words *pipa, pepa, papa, popa, pupa* each time the sentence was uttered. Group I was asked to utter the vowel phonemes represented by the graphemes *i e a o u* of the Spanish orthographic system.

The subjects of Group IV representing Brazilian Portuguese speech each repeated the words *se* [i], *sesta* [e] , *sexta* [ɛ] , *sa* [a] , *só* [ɔ], *sou* [o] *and zuzu* [u] in the frame "Digo a palavra ____outra vez."

b) Acoustic Analyses:

The recorded data were then submitted to a spectrographic analysis. Spectrograms were made for each vowel of each speaker, and the formant frequency values F1/F2/F3 were measured. The recorded data were also analysed by means of a computerized LPC (Linear Predictor Coefficient) analysis to determine the formant frequency values.

Both sets of results were checked and corroborated. Tables 2 and 3 list the formant values obtained from both analyses. These values were then plotted on formant charts with F1 at the abscissa and F2-F1 at the ordinate. Ellipses were drawn at two standard deviations from the means around the points representing each vowel. The ellipses of each vowel of each dialect of Spanish were superimposed so as to facilitate comparisons between each dialect.

c) Results:

Figure 2 shows the formant values plotted as ellipses for Mexican Spanish (solid lines), Argentine Spanish (dotted lines), and Peninsular Spanish (broken lines). We can see that the target areas of the vowels [o], [u] and [a] of Peninsular Spanish are positioned more closely than in the other two dialects of Spanish. These latter two dialects seem to exhibit larger acoustic differences in the spacing of these back vowels. There also seems to be some differentiation in the height of the front vowels in Mexican and

		GROUP I			GROUP II			GROUP III		
Subject		F1	F2	F3	F1	F2	F3	F1	F2	F3
1	[i]	299.	2227.	2833.	271.	2092.	2923.	365.	2297.	3124.
2		335.	2222.	2970.	292.	2222.	2864.	305.	2584.	3059.
3		274.	2194.	2921.	279.	2277.	3044.	345.	2194.	2836.
4		314.	2233.	2819.	277.	2339.	3061.	311.	2174.	2606.
5		272.	2606.	3057.				312.	2679.	3341.
6		292.	2437.	3042.				305.	2323.	2839.
1	[e]	458.	1988.	2604.	389.	1969.	2566.	509.	2044.	2535.
2		448.	2204.	2853.	412.	1950.	2479.	387.	1878.	2634.
3		469.	2158.	2628.	360.	2026.	2837.	450.	1944.	2621.
4		460.	1970.	2594.	387.	2071.	2633.	473.	1728.	2342.
5		399.	2340.	2941.				371.	1908.	2512.
6		464.	2315.	2768.				432.	1990.	2888.
1	[a]	692.	1186.	1950.	697.	1246.	2402.	688.	1252.	2499.
2		784.	1357.	2384.	675.	1236.	2206.	634.	1366.	2538.
3		796.	1351.	2650.	824.	1296.	2839.	773.	1106.	2637.
4		721.	1378.	2350.	775.	1196.	2656.	606.	1162.	1649.
5		898.	1358.	2680.				594.	1058.	2412.
6		817.	1502.	2616.				707.	1195.	2468.
1	[o]	474.	903.	2533.	433.	906.	1416.	509.	1027.	2833.
2		563.	845.	1783.	465.	903.	1606.	431.	981.	2613.
3		418.	624.	1567.	458.	951.	2731.	525.	1080.	2779.
4		466.	789.	2334.	456.	772.	1590.	522.	1072.	2581.
5		435.	909.	1981.				430.	1166.	2558.
6		494.	977.	1518.				463.	902.	2333.
1	[u]	334.	872.	1626.	335.	767.	1515.	394.	1164.	2752.
2		339.	868.	2235.	353.	642.	1434.	339.	857.	2510.
3		290.	603.	2036.	342.	956.	1649.	385.	1090.	1662.
4		322.	681.	2255.	305.	606.	2588.	365.	873.	2251.
5		307.	913.	1540.				335.	676.	1295.
6		330.	949.	1552.				333.	897.	3391.

Table 2. Formant frequency values representing Mexican (Group I), Argentine (Group II) and Peninsular (Group III) Spanish dialects.

Subject		F1	F2	F3		F1	F2	F3		F1	F2	F3
1	[i]	247.	2021.	2685.	[a]	743.	1380.	2676.	[ɔ]	551.	930.	2751.
2		264.	1995.	2632.		685.	1299.	2087.		485.	865.	2496.
3		296.	2106.	2664.		794.	1346.	2516.		573.	962.	2957.
4		308.	2259.	2923.		742.	1238.	2629.		600.	1011.	2247.
5		324.	1980.	2501.		627.	1273.	2482.		592.	958.	2462.
6		192.	2177.	3127.		693.	1280.	2095.		530.	847.	3147.
7		272.	1929.	2455.		803.	1310.	2119.		524.	797.	2608.
8		207.	2312.	2975.		725.	1394.	2219.		569.	858.	2085.
9		262.	2078.	2802.		792.	1346.	2434.		640.	965.	2374.
1	[e]	387.	2050.	2569.					[o]	377.	755.	2041.
2		387.	1027.	2600.						359.	856.	2518.
3		423.	1872.	2263.						424.	894.	2792.
4		459.	1805.	2425.						504.	1027.	2585.
5		368.	1929.	2467.						389.	764.	2285.
6		317.	1904.	2532.						295.	596.	2213.
7		350.	1858.	2635.						410.	1035.	2076.
8		350.	2108.	2776.						421.	936.	2459.
9		449.	1819.	2282.						429.	926.	2151.
1	[ɛ]	448.	1856.	2461.					[u]	305.	801.	2222.
2		356.	1870.	2490.						282.	719.	2443.
3		587.	1659.	2403.						327.	855.	2356.
4		470.	1708.	2492.						350.	1227.	2424.
5		566.	1715.	2855.						279.	719.	2615.
6		429.	1695.	2206.						283.	967.	1931.
7		362.	1946.	2627.						282.	880.	2797.
8		560.	1774.	2373.						297.	783.	2566.
9		572.	1768.	2805.						353.	1065.	2518.

Table 3. Formant frequency values representing speakers of Brazilian Portuguese (São Paulo, Rio de Janeiro, Espírito Santo and Bahia).

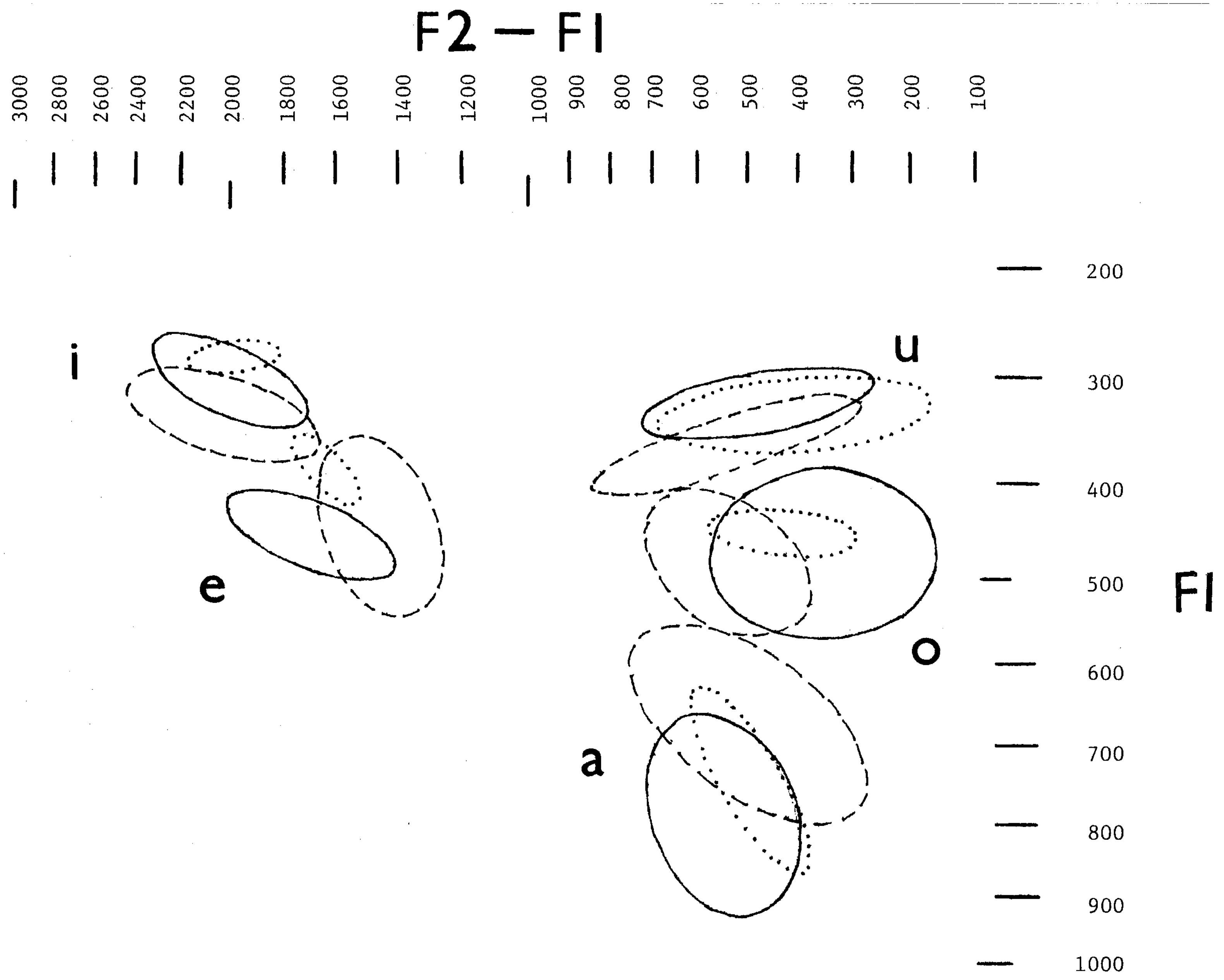


Figure 2. A plot of the formant frequencies from Table 2.

——— Mexican
······· Argentine
- - - - - Peninsular

Argentine Spanish dialects. The sound represented by [e] in Argentine Spanish seems to occupy a higher position in the vowel space than the comparable vowel of the Mexican Spanish dialect. The same type of relationship seems to also hold true for the vowel [i] in Argentine and Peninsular Spanish. It will also be noticed that the vowel traditionally transcribed as [a] in the published literature occupies a back position in the acoustic space rather than a central quality as has been usually been ascribed to his vowel.

As may be seen in Figure 3 the target areas of the front vowels of Brazilian Portuguese show considerable overlap. This is particularly evident in the case of the vowels [e] and [ɛ] relationship where the ellipses depicting [ɛ] intersects with the [e] space. There is also a noticeable degree of variation in the front-back dimension of the vowel [e] and in the high-low dimensions of the vowel [ɛ].

The vowels [e] and [o] in Spanish have also been traditionally depicted in positions equidistant between[i]and[a]and[u]and[a]respectively.[15] The position of the vowel targets for[e]and[o]in this study would tend to disclaim this assertation. The general overall pattern seems additionally not to follow the predictions of the principle of maximal perceptual contrast discussed in Liljencrants and Lindblom (1972). This principle suggests that the vowels of a particular language are maximally dispersed relative to each other in the available acoustic vowel space. The data reported here seems not to tend to support this claim. Further analyses will be necessary, however, before we may be certain concerning this point.

d) Discussion:

The results based on the formant frequency plots would tentatively suggest at this point that differences in the phonetic quality of the vowels in the data on Spanish potentially exist. This would reaffirm on an objective basis the auditory assessments made by Hispanic and Lino-Brazilian linguists in the published literature over the years.[16] However because of the relatively small sample population dealt with in this experiment for Groups I, II, III (a ratio of 6-4-6) conclusions based on the results presented here can only be taken as suggestive. This particularly holds true for the data on Brazilian Portuguese. The particular vowel tokens elicited from the subjects represented as the ellipses in Figure 3, are in fact not representative of these vowel qualities. This is exhibited in the large degree of overlap in the [e] and [ɛ] vowel target areas.

To summarize, we have seen that the tongue arching model has served as an inadequate description of the actual articulatory parameters involved in the production of vowel sounds. It was shown that the relevant physiological features in the determination of vowel quality are 1) the state of the lips 2) the position of the tongue body and 3) the vertical height of the larynx. Tongue height, as was pointed out above, was not an invariant feature in the determination of a particular vowel quality. The labels

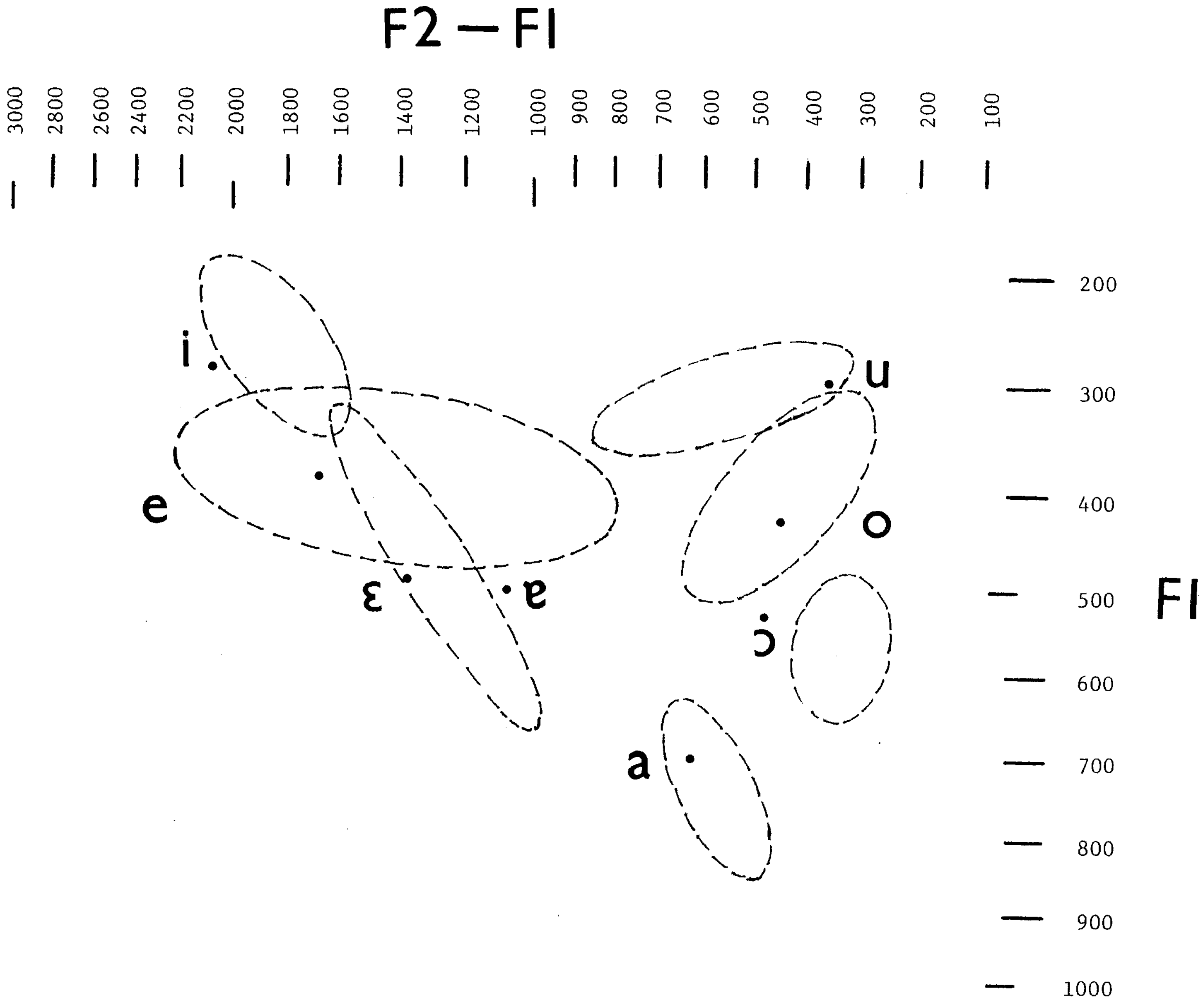


Figure 3. A plot of the formant frequencies from Table 3 representing Brazilian Portuguese. Dots represent values (Table 1) from Delgado Martins, 1973 study on Lisbon Portuguese vowels.

high vs. low and front vs. back that were traditionally purported to describe the highest point of the tongue, are useful insofar as they refer to *the relative positions of vowels in the acoustic space.* The total supralaryngeal vocal tract shape as the result of the three parameters act as constant determiner of vowel quality. The particular vocal tract shape acts as an acoustic filter that permits certain frequencies through called formants. These formant frequencies are the acoustically significant perceptual (i.e. auditory) attributes that identify vowel sounds. By plotting F1 against F2-F1, rather than F1 against F2, we more closely approximate the perceived perceptual distances between vowels in the available acoustic space.[17]

As was pointed out the results presented in this preliminary report are only meant to be suggestive. Before we can draw any general conclusions from our analyses, further analyses will be necessary. Steps in this direction are already underway. At the time the present paper was completed the following research was being undertaken:

1) increase the size of the sample population to a minimum of 12 subjects for each representative dialect of Spanish and of Brazilian Portuguese.

2) the utilization of a set of words that will more readily elicit the vowels [e] and [ɛ] from the Brazilian subjects.

This latter point being important since the great deal of variation noted, for example, in the overlapping of the front vowels of Brazilian Portuguese was due to a large degree to the fact that the words use in this preliminary study were not suitable for eliciting the appropriate vowel phonemes. In addition, the formant frequency values will be submitted to an analysis of variance program so as to determine quantitatively the similarities and differences among comparable vowels in spanish and Portuguese.

Footnotes

1. Chronologically, for example, we have in Spanish: Josselyn (1907); Colton (1909); Navarro Tomás (1927); Entrambasaguas (1952); Flórez (1960); Cárdenas (1960); Alarcos Llorach (1961); Serralta and Tuscón (1967); Navarro Tomás (1968a, 1968b); Dalbor (1969); Quilis and Fernández (1973); Stockwell and Bowen (1973) and in Portuguese we have: Bonaparte (1885); Gonçalves Vianna (1892); Sá Nogueira (1938); Gonçalves Vianna (1940); Lüdtke (1953); Strevens (1954); Nunes (1960); Carvalho (1961, 1962); Azevedo Filho (1971); Redenbarger (1975). For Romance, e.g., Defarrari (1954); Malmberg (1971), Hammartröm (1972).

2. See, for example, discussion by Ladefoged et al. (1972); Wood (1975), (1977) and Lieberman (1976).

3. Wood (1975) and Lieberman (1976).

4. Wood (1975) p. 81.

5. e.g. Delattre (1951)

6. Perkell (1977) p. 62-63.

7. Harshman (1970) p. 48; Riordan (1977).

8. Joos (1948).

9. Skelton (1955) however, does not include the formant frequency values themselves in his dissertation.

10. The same study was also published in Guirao and Borzone de Manrique (1975).

11. I was unable to examine the study by Miyoshi (1969) at the time this paper was being prepared.

12. As with the Skelton (1965) study, Head (1965) does not include the formant frequency values for the vowels examined, despite the author's claim that his work provides an acoustic phonetic analysis of Portuguese (p. 2). Also, a comprehensive survey of the published literature on Spanish and Portuguese phonetics (Godínez, in preparation), has been unsuccessful in locating another acoustic study by the same author (Head, 1964-65).

13. Godínez (1977a, in press).

14. I am grateful to Professor Ana M. Borzone de Manrique of the Laboratório de Investigaciones Sensoriales (Universidad de Buenos Aires) for sending me the recorded material on Argentine Spanish.

15. See footnote 1

16. See footnote 1

17. Ladefoged, in discussion of Lindblom (1972) p. 93.

References

Alarcos Llorach, Emilio (1961) *Fonología española,* 3ª ed. aumentada y revisada. Madrid: Editorial Gredos.

Azevedo Filho, Leodegário de (1968) "Estudo da gramática Portuguesa: Fonologia" *Revista de Portugal* (Lisboa) 33, #267, 318-43.

Azevedo Filho, Leodegário A. de (1971) *Para uma gramática estrutural da lingua portuguesa,* Rio de Janeiro: Edições Gernasa.

Bonaparte, H.I.H. Prince L.L. (1885) "Portuguese Vowels, According to Mr. R.G. Vianna, Mr. H. Sweet, and Myself" *Transactions of the Philological Society* 3-4 pt.III (London) 404-08.

Cárdenas, Daniel N. (1960) "Acoustic Vowel Loops of Two Spanish Idiolects" *Phonetica* 5, 9-34.

Colton, Molton A. (1909) *La Phonétique Castillane: Traité de Phonétique Descriptive et Comparative*, Paris: [Lievens à Saint-Maur].

Dalbor, John B. (1969) *Spanish Pronunciation: Theory and Practice*, New York: Holt, Rinehart and Winston.

Defarrari, Harry A. (1954) *The Phonology of Italian, Spanish and French* Washington, D.C. Roy J. Defarrari

Delattre, Pierre (1951) "The Physiological Interpretation of Sound Spectrograms", *Publications of the Modern Language Association of America* 66, 864-75.

Delattre, Pierre (1964) "Comparing the Vocalic Features of English, German, Spanish and French" *International Review of Applied Linguistics*, 2 #2, 71-97.

Delgado Martins, Maria R. (1964-73) "Análise acústica das vogais orais tónicas em português" *Boletim de Filologia* (Lisboa) #3/4, 303-14.

Delgado Martins, Maria R. (1975a) "Vogais e consoantes do português: estatística de ocorrência, duração e intensidade"*Boletim de Filologia* (Lisboa) 24, 1-11.

Delgado Martins, Maria R. (1975b) "Perception de l'accent portugais" *8th International Congress of Phonetic Sciences* (Leeds)

Delgado Martins, Maria R. (1977) "Aspects de l'accent en portugais: voyelles toniques et atones" Unpublished doctoral dissertation Université de Strasbourg.

Entrambasaguas, Joaquín de (1952) *Síntesis de pronunciación espanõla* Madrid: Consejo Superior de Investigaciones Científicas.

Fant, Gunnar (1960) *Acoustic Theory of Speech Production*. The Hague: Mouton.

Flórez, Luis (1960) *Leccoines de pronunciación*, Bogota: Editorial "Bremen".

Godínez Jr., Manuel (in press) "Preliminaries to a Bibliographic Survey of Spanish and Portuguese Phonetics" *Sumposium on Hispanic and Luso-Brazilian Linguistics* (University of Hawaii, Manoa, 1977)

Godínez Jr. Manuel (1977a) "Preliminaries to an Acoustic Analysis of Brazilian Portuguese Vowels" UCLA Phonetics Lab. Unpublished Manuscript.

Godínez Jr.,Manuel (1977b) "The Specificiation of Vowels in Spanish" UCLA Phonetics Laboratory. Unpublished manuscript.

Godínez Jr. Manuel (in preparation) "A Bibliographic Survey of Spanish and Portuguese Phonetics" UCLA Phonetics Laboratory.

Gonçalves Vianna, Aniceto dos Reis (1940) "Essai de phonétique et de phonologie de la langue portugaise d'apres le dialecte actuel de Lisbonne" *Boletim de Filologia* (Lisboa) 12, 165-243.

Gonçalves Vianna, Aniceto dos Reis (1892) *Exposição da pronúncia normal portuguesa,* Lisboa: Imprensa National

Guirao, Miguelina and Borzone de Manrique, Ana M.B. (1972) "Identification of Argentine Spanish Vowels" *Proc. of the 7th Intern. Congress of Phonetic Sciences* (eds. A. Rigault and R. Charbonaeau) The Hague: Mouton. Also in *Journ. of Psycholinguistic Res.* 4,#1 (1975) 17-25.

Harshman, Richard A. (1970) "Foundations of the PARAFAC Procedure: Models and Conditions for an 'Explanatory' Multi-Modal Factor Analysis" *UCLA Working Papers in Phonetics* 16, 1-84.

Hammarström, Göran (1972) "The Romance Languages" in (eds. James M. Anderson and Jo Ann Creore) *Readings in Romance Linguistics.* The Hague: Mouton.

Head, Brian F. (1964-65) "Esboço de uma classificação acústica dos fonemas do português" Paper presented at the *XI Congresso Internacional de Linguística e Filologia Românicas*, Madrid, 1965. Cited in *Revista Portuguesea de Filologia* (Coimbra) 13 #1/2 687.

Head, Brian F. (1965) *A Comparison of the Segmental Phonology of Lisbon and Rio de Janeiro.* Unpublished dissertation, U. of Texas at Austin.

Heike, Georg (1967) "Zur Methodik phonetisch-phonologischer Vergleichung: Ein Beitrag zur luso-brasilianischen Aussprache" *Phonetica* 17, 231-40.

Carvalho, José G. Herculano de (1961) "Le vocalisme atone des parlers créoles du Cap Vert" *Boletim de Filologia* (Lisboa) 20 #1/2, 3-12.

Carvalho, José G. Herculano de (1962) "Sincronia e diacronia nos sistemas vocálicos do crioulo caboverdiano" in *Miscelânea Homenaje a André Martinet,* Vol. I (ed. Diego Catalán) Canarias: Universidad de La Laguna.

Joos, Martin (1948) *Acoustic Phonetics, Language* Monographs #23 Suppl. to *Language* 24 #2.

Josselyn, Freeman M. (1907) *Études de phonétique espagnole,* Paris: H. Welter.

Ladefoged, Peter (1975) *A Course in Phonetics*, New York: Harcourt Brace Jovanovich, Inc.

Ladefoged, Peter and DeClerk, J. and Lindau, M. and Papçun, G. (1972) "An Auditory-Motor Theory of Speech Production" *UCLA Working Papers in Phonetics* 22.

Lieberman, Philip (1976) "Phonetic features and physiology: a Reappraisal" *Journal of Phonetics 4*, 91-112.

Liljencrants, Johan and Lindblom, Björn (1972) "Numerical simulation of vowel quality systems: The role of perceptual contrast" *Language* 48, 4, 839.

Lindblom, Björn E.F. (1972) "Phonetics and the description of language" in *Proc. of the 7th Intern. Congress of Phonetic Sciences*, (A. Rigault and R. Charbonneau, eds.) The Hague: Mouton.

Lüdtke, H. (1953) "Fonemática portuguesa II: Vocálismo" *Boletim de Filologia* (Lisboa) 14 #1/2, 197-217.

Malmberg, Bertil (1971) "La structure phonétique de quelques langues Romane" in *Phonétique générale et romane: Etudes en Allemand, Anglais, Espagnol et Francais*. The Hague: Mouton.

Nunes, José Joaquim (1960) *Comprêndio de gramática histórica, portuguesea: Fonetica e morfologia*, Lisboa: Livraria Clássica Editora.

Miyoshi, J. (1969) "Reflexión acústico-fonética de las vocales espanõlas: (II) *Hispanica* 14, 1-19.

Navarro Tomás, Tomás (1927) *Comprendio de orthología española*, Madrid: Librería y Casa Editorial Hernando (S.A.).

Navarro Tomás, Tomás (1968) *Manual de pronunciación española*, Madrid: Consejo Superior de Investigacioes Científicas. Publicaciones de la *Revista de Filología Española, III*.

Navarro Tomás, Tomás (1968) *Studies in Spanish Phonology*. (Trans. by Richard D. Abraham). Florida: University of Miami Press.

Perkell, Joseph S. (1969) *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*, Cambridge: The MIT Press.

Quilis, Antonio and Fernández, Joseph A. (1973) *Curso de fonética y fonología españolas*, 7ª edicíon revisada y aumentada. Madrid: Consejo Superior de Investigaciones Científicas.

Redenbarger, Wayne J. (1975) "Lusitanian Portuguese [ɐ]is [+ATR] and [+CP]" in *Studies in Romance Linguistics: Proceedings of the Fifth Linguistic Symposium on Romance Languages* (ed. Michio P. Hagiwara). Rowley, Mass.: Newbury House Publications.

Riordan, Carol J. (1977) "Control of Vocal-Tract Length in Speech" *Journ. of the Acoust. Soc. of Amer.* 62.4, 988-1002.

Russell, George O. (1928) *The Vowel*, Columbus: Ohio State Univ. Press.

Sá Nogueira, Rodrigo de (1938) *Elementos para um tratado de fonética portuguesa* Lisboa: Imprensa Nacional de Lisboa.

Serralta, F. and Tuscón, V. (1967) *Phonétique espagnole pratique,* Paris : Université de Toulouse.

Skelton, Robert B. (1950) "A Spectrographic Analysis of Spanish Vowel Sounds" Unpublished doctoral dissertation, U. of Michigan.

Skelton, Robert B. (1969) "The Pattern of Spanish Vowel Sounds" *International Review of Applied Linguistics 7*, #3, 231-37.

Stockwell, Robert P. and J. Donald Bowen (1973) *The Sounds of English and Spanish,* Chicago: U. of Chicago Press.

Strevens, P.D. (1954) "Some observations on the phonetics and pronunciation of modern Portuguese" *Revista do Laboratório de Fonética Experimental* (Coimbra) 2, 5-29.

Wood, Sidney (1975) "The weakness of the tongue-arching model of vowel articulation" *Working Papers,* Phonetics Laboratory, Lund University, 11, 55-108.

Lip Positions and Forment Frequency in American English Vowels[1,2]

[Paper delivered at the 94th meeting of the Acoustical Society of America]

Wendy Linker

In a previous paper delivered to the Acoustical Society of America, Ladefoged and Rice showed that the distance between the upper and lower lips could be predicted from a particular combination of formant frequencies, and that the correlation between the observed lip distance and predicted lip distance was 0.78.[3] Their data consisted of x-rays taken of five subjects producing 10 American English vowels. The aim of the present investigation is to examine in more detail the extent to which lip position is predictable from acoustic data, and to determine the lip measure or measures that are most highly correlated with formant frequency.

Simultaneous frontal and lateral photographs were taken of eight male subjects pronouncing nine words: heed, hid, hayed, head, had, hod, hoed, hood, who'd. A device, which had two cameras mounted on it, was used to keep the subject's head steady during the production of these words. The subject was instructed to sustain the vowel of each word until after the simultaneous photographs were taken. The entire session, conducted in a sound attenuated booth, was recorded. In the recording, an audible click from the camera shutters indicates the exact time at which the photographs were taken.

In order to measure lip protrusion, it was necessary to decide on a fixed reference point to measure from. The upper front teeth were selected, since this structure moves much less than any other structure in the front of the mouth. A narrow paper strip marked in 5 millimeter sections was attached to a very thin metal wire, and the wire was fastened in between the subject's upper front teeth. From this, relevant distances to

---

[1] This research was supported by a grant from NIH to the UCLA Phonetics laboratory.

[2] I would like to express my gratitude to Peter Ladefoged, Ian Maddieson, Willie Martin, George Papcun, Lloyd Rice and particularly to Louis Goldstein for their much needed assistance in the preparation of this paper.

[3] A description of this study may be found in Ladefoged, Harshman and Goldstein, "Vowel Articulations and Formant Frequencies" , UCLA Working Papers in Phonetics 38.

the upper teeth could be calculated. Care was taken to ensure that the stick did not touch either the upper or lower lip of the subject. In addition, a millimeter scale was placed in the frontal view.

To obtain measures of lip position, negatives of the front and side views were enlarged,and tracings were made. Exact distances were calculated from the scales using a computer program written by Lloyd Rice. With this program it was possible to use a Grafpen connected to the PDP-12 computer to retrace the tracings of the negatives. The program used the points marked by the Grafpen as input to computer 24 lip measurements. The measurements on the lateral view were taken with reference to the five points shown in Figure 1: (1) Point B, a fixed point on the upper teeth, (2) Point C, the corner of the mouth, (3) Point D, the most forward point of contact between the lips, (4) Point E, the outer point on the lower surface, and (5) Point F, the outer point on the upper lip surface. Every possible direct distance from each of these five points to one another was found. In addition, selected perpendicular distances were taken from points B, C, and D to: (a) the tangent line through points E and F and (b) perpendicular lines drawn through points B, D, and F. Finally, the area enclosed by the lips was calculated. Only 12 of these side view variables are considered in the analysis.

Figure 2 shows the front view measurements that were calculated. Measures 20, 21, and 22 represent the vertical distance between the lips taken at equal intervals, with 21 being at the center of the of the lips. Measure 23 is the distance between the two corners of the mouth, and measure 24 is the front view area. Thus, 12 side view variables and 5 front view variables entered into the analysis, giving a total of 17 lip variables.

Wide band spectrograms were made of each utterance photographed and the formant frequencies F1, F2 and F3 were measured at a point immediately prior to the click of the camera shutters. (This click was easily visible on the spectrogram). In some cases, where interpretation was difficult, an LPC program was used to supplement the spectrograms. From these three primary acoustic measurements, 22 additional acoustic varaibles were obtained by calculating products and ratios of combinations of formant frequencies.

A factor analysis was carried out together with Louis Goldstein using Harshman's PARAFAC procedure, in order to see how the lip measures covaried with one another for different vowels and different speakers. It was found that only one factor was needed to account for 92% of the variance. Perhaps here it is worth noting briefly how PARAFAC works. As is shown under (1) in Table I, this kind of factor analysis gives as a result three sets of constants: the first set of constants represents the weighting of each of the 17 lip measurements on the one lip factor that was extracted; the second set of constants represents the weightings of the 9 individual vowels on the lip factor, and the third set represents the weightings of the 8 individual speakers with respect to the underlying lip factor. Under (2) in Table I, the values for each of the 9 vowel loadings are given. It can be observed that the values for the loadings tend to be higher for spread vowels (numbers 1 through 6) and somewhat lower for rounded vowels (numbers

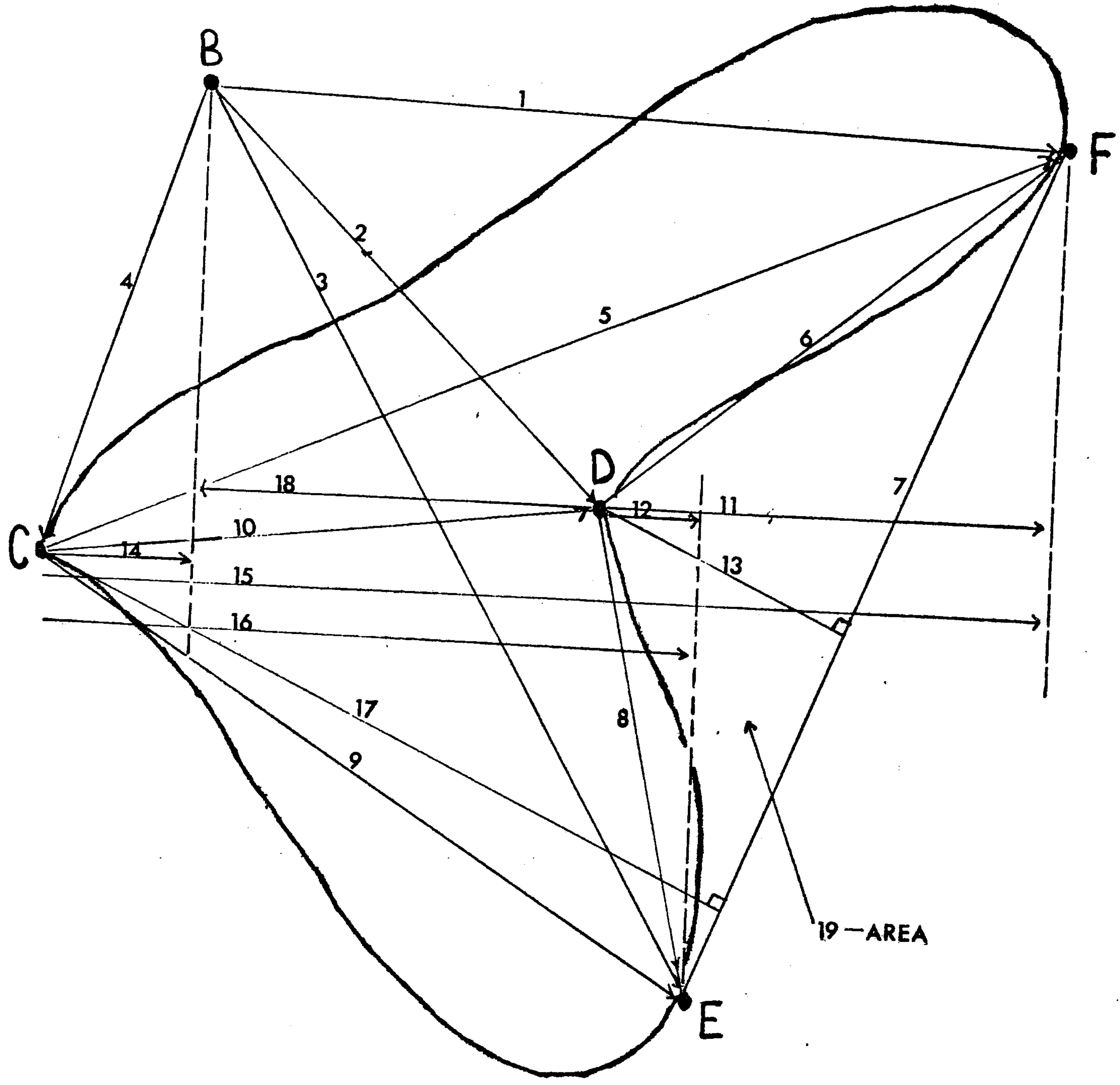


Figure 1. A diagram showing the measurements taken from the lateral view of the lips.

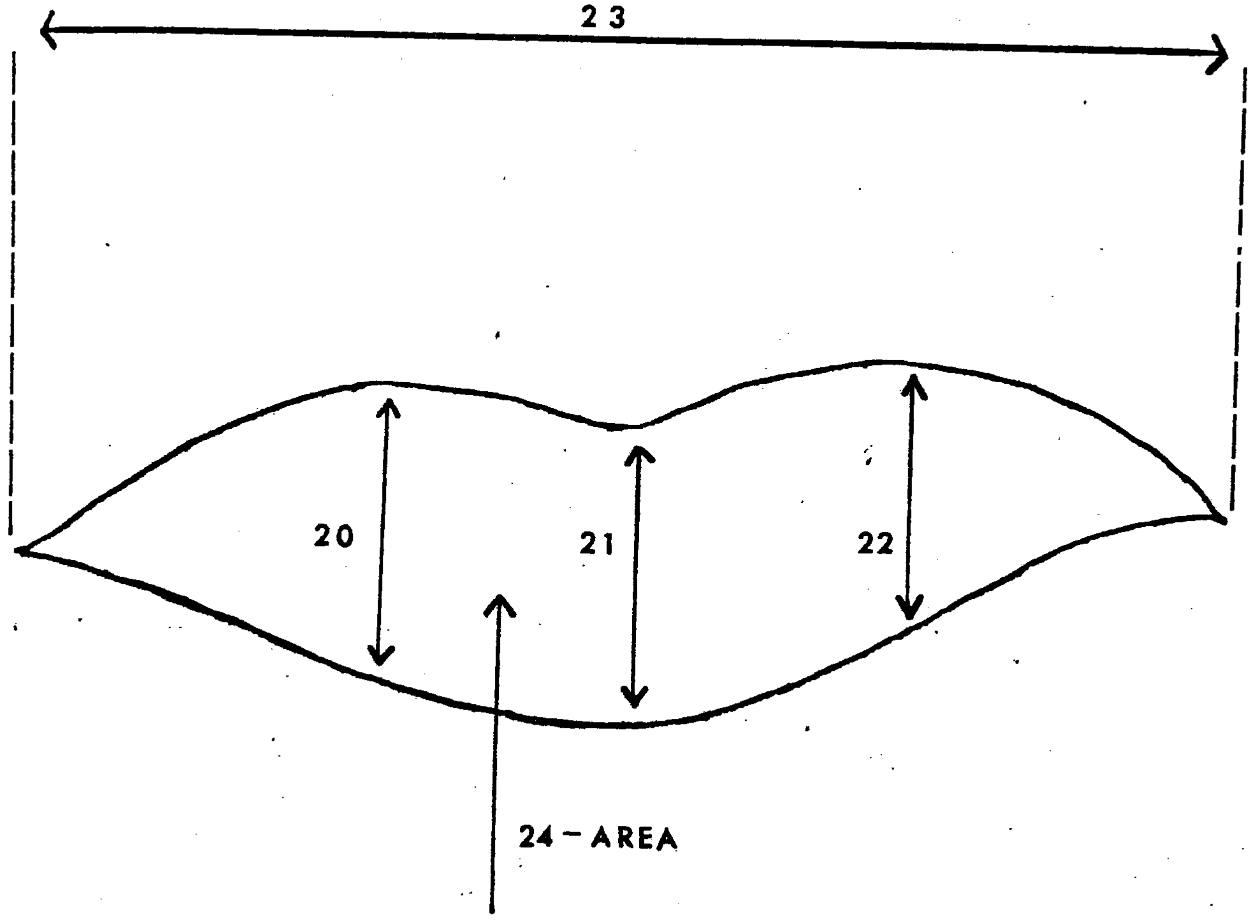


Figure 2. A diagram showing the measurements taken from the frontal view of the lips.

7 through 9). Thus, this lip factor is closely related to a traditional description of lip position in vowels.

Given that the 17 lip variables are so highly correlated with one another with the extraction of only one factor, and given that this factor lends itself to a reasonable interpretation, the factor may be taken to represent the underlying determiner of lip position for these data. Therefore, we would like to predict this one lip factor from formant frequencies, in order to arrive at a measure of the correlation between the acoustic data and the lip variables taken overall. It would be possible to predict each of the 17 lip measures from the formant frequencies, using 17 different equations. But these predictions would not use the fact that all 17 measures are highly correlated, and should not be allowed to vary independently. It is therefore appropriate to predict them from the same equation.

Table I.

(1) Predicted distance $_{ijk}$ = measure$_k$ x vowel $_j$ x speaker$_k$ (r=.961)

(2) Vowel loadings

1.	heed	1.038
2.	hid	1.033
3.	hayed	1.133
4.	head	1.131
5.	had	1.195
6.	hod	1.111
7.	hoed	0.762
8.	hood	0.808
9.	who'd	0.631

(3) Regression equation to predict underlying lip parameter L
72 tokens (8 speakers, 9 vowels) r=.831

$$L = 20.44 \frac{F2}{F3} + 0.005 \frac{F1.F3}{F2} - 1.610$$

To determine how to predict this underlying lip parameter from the formant frequencies, a stepwise multiple regression program was run,using the individual products of the 8 speaker loadings and 9 vowel loadings given by the PARAFAC analysis. As can be seen under (3) in Table I, a fairly high degree of correlation -0.831- was achieved with only two terms. Having predicted the values of the factor loadings on vowel tokens, it was now possible to take the actual loadings of lip measurements given by PARAFAC, together with the vowel token loadings predicted by the regression, to reconstruct a predicted set of lip measures, that depend on formant frequency. The correlation obtaining between these predicted lip measurements and the actual lip data is 0.945.

We can now look at the difference between predicted and observed values for each individual lip measure across all vowels and subjects, to see which measures were best recovered. Here it is important to remember that only one regression equation is being used, namely the one that predicts the underlying lip parameter, and it is from this parameter that the other measures are predicted.

The correlation coefficients for the 17 lip variables are listed under (2) in Table II. It can be seen that the variable that is recovered best is number 6, which is a measure of upper lip protrusion from the most forward point of lip contact. The measure that is predicted the most poorly is number 1, the protrusion from the tooth to the upper lip. Although it might be thought that both these measures are good descriptions of the amount of lip protrusion, apparently only the distance from the most forward point of lip contact can be predicted accurately. This may indicate that measure number 6 is the best possible measure of the outer limits of the vocal tract.

Table II.

(1) Correlation between predicted and observed 17 measures, 9 vowels, 8 speakers r = .945

(2) Individual correlations for the 17 lip measures (n = 72):

<u>Side view</u>

	1.	Distance from tooth (B) to upper lip point (F)	0.191
	2.	Distance from B to most forward point of lip contact (D)	0.683
	3.	Distance from B to lower lip point (E)	0.332
	4.	Distance from B to corner of mouth (C)	0.670
	5.	Distance from C to F	0.641
*	6.	Distance from D to F	0.828
	7.	Distance from E to F	0.683
	8.	Distance from D to E	0.744
	11.	Distance from D to perpendicular line through F	0.810
	12.	Distance from D to perpendicular line through E	0.690
	13.	Distance from D perpendicularly to line connecting E and F	0.738

<u>Front view</u>

20.	Distance between lips, 1	0.409
21.	Distance between lips, 2	0.426
22.	Distance between lips, 3	0.425
23.	Distance between corners of mouth	0.610
24.	Area	0.514

On the identifiability of isolated vowels

Daniel Kahn[1]

It is well established that the most important parameters in the identification of vowels are the frequency values of the first two amplitude peaks, or formants, in the speech spectrum. Yet the actual frequencies of the formants corresponding to a given perceived vowel quality vary a great deal from speaker to speaker. For example, in their well-known experiment, Peterson and Barney (1952) observed F1/F2 values ranging from 600/1600 Hz to 1100/2600 Hz for the vowel [æ] of had; furthermore, the region in F1/F2 space around 600/1600 Hz, for example, was populated by tokens of the vowels [æ, ɛ, ʌ, ω, ɝ].

In spite of this bewildering lack of one-to-one correspondence between vowel quality and small regions of F1/F2 space, normal communication does not suffer from vowel mis-identification, even in a language like English, with many phonemically distinct vowel qualities. Two types of explanation suggest themselves.

I will refer to the first as the internal-parameter hypothesis. According to this hypothesis, all information necessary for the identification of a vowel is contained in the acoustic signal. While F1 and F2 are the most important cues for vowel identification, listeners make important use of other cues, such as the frequencies of higher formants, vowel-duration, fundamental frequency contour, degree of diphthongization, etc.

The second possible explanation is what I will call the normalization hypothesis, according to which speaker-normalization accounts for the disambiguation of F1/F2 values which could correspond to more than one vowel. Under this hypothesis, formant transitions from consonant to vowel and vowel to consonant as well as previous exposure to certain known or inferrable vowels of a given speaker would provide sufficient information about the parameters of that speaker's vocal tract that a given point in F1/F2 space could correspond to only one vowel.[2]

---

[1] Daniel Kahn is currently at Bell Telephone Laboratories and the University of Massachusetts. Most of the work described in this paper was performed while he was at the UCLA Phonetics Lab. An earlier version of this paper was presented at the 94th meeting of the Acoustical Society of America, Miami Beach, Florida, 16 December 1977.

[2] The experimental work described in Ladefoged (1967) demonstrates that this sort of normalization affects listeners' judgement about vowel quality, at least in the case of synthetic vowels. The "normalization"

While it is certainly possible, and I believe correct, that the factors mentioned in both explanations contribute to our highly reliable vowel identification ability in ordinary situations, the results of certain experiments suggest that normalization and consonantal context are much more important than the secondary internal parameters - higher formants, duration, etc.

Consider an experiment in which several speakers produce isolated vowels which are then randomized by speaker and vowel quality and presented to listeners for identification. The normalization hypothesis predicts very poor identifiability, since the absence of consonants in the stimuli prevents the use of consonantal transitions for speaker normalization, while the speaker randomization eliminates the possibility of normalization by continued exposure to the same speaker's vowels.

On the other hand, all the vowel identification cues of the internal-parameter hypothesis are retained in such an experiment (provided that the recording and playback are of high quality), so that if this latter hypothesis is correct, one would expect very high identification scores. In fact, however, the scores are typically very poor.

In a series of experiments performed by Strange, Verbrugge, Shankweiler, and Edman (1976), an error rate of 43% was observed under the speaker-randomized isolated vowel condition: that is, nearly *half* of the vowels presented to listeners were mis-identified. As outlined above, this sort of result can be viewed as evidence for the normalization hypothesis and against the internal-parameter hypothesis.

I undertook the experiment to be described in this paper because I was concerned that the methods employed in previous experiments might be such as to unfairly downplay the contribution made by purely internal parameters to vowel identification. Furthermore, it seemed to me that on the basis of the very high error rates observed in speaker-randomized isolated vowel experiments, the view was becoming prevalent that sufficient information for accurate identification of a single isolated vowel is simply not contained in the acoustic signal; it is important to know whether this view is correct.

---

hypothesis should perhaps be broadened. In discussing why vowels in CVC syllables are easier to identify than isolated vowels, Strange, Verbrugge, Shankweiler, and Edman (1976:214) first point out that consonantal transitions may supply information necessary for speaker normalization, then say, "On the other hand isolated vowels may be difficult to perceive for a more fundamental reason. It is possible that listeners ordinarily rely upon information distributed throughout the whole syllable for identification of the vowel...If it is the case that...transitions specify the vowel as well as the consonants, we could assert that the vowel is inseparable from the syllable, that it is not specified by formant frequencies at any particular cross section in time, but rather is carried in the dynamic configuration of the whole syllable."

There are certain factors which immediately come to mind as possible sources of theory-irrelevant errors in an experiment in which speaker-randomized isolated vowels are to be identified:

(1) speaker and listener may vary in dialect

(2) speakers may make errors in production. In addition to gross errors such as production of a vowel other than the one intended, a speaker might, because of momentary confusion or inattention, produce a vowel which is slightly "off" in quality. Also, a certain percentage of utterances made by untrained speakers will be muffled or unclear because of poor articulation, extraneous mouth noises, etc.

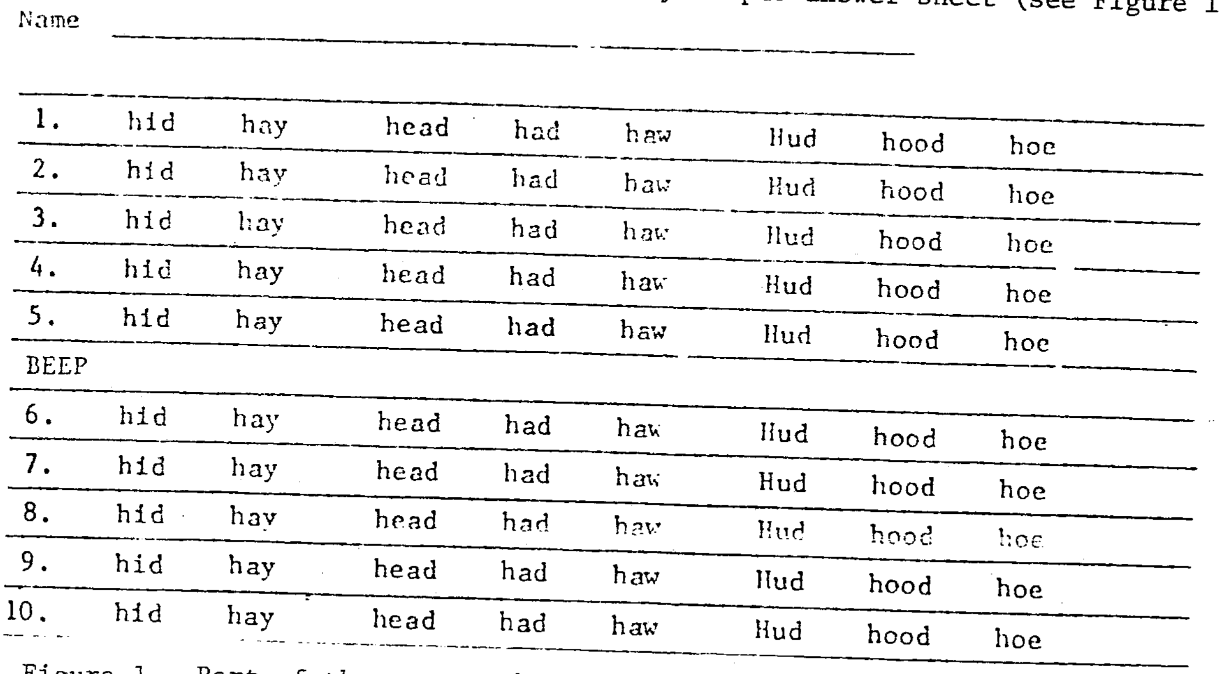
(3) listeners may make simple errors in transcription, correctly perceiving [ʌ], for example, but circling [ʊ] on the answer sheet.

In the design of the present experiment, I attempted to circumvent these problems in the following ways:

First of all, it was my intention to use speakers and listeners from a well-defined dialectal area. Practical considerations prevented me from accomplishing this goal, though the results are perhaps even more interesting because of the dialectal variation tolerated. Most of the twenty speakers were natives of the Los Angeles area, though none of the listeners were.

Secondly, the production of stimuli by speakers was carefully monitored, and any productions which appeared to be deviant or unclear were discarded.

In addition, to minimize extraneous errors on the part of the listeners, pains were taken to produce an optimally simple answer sheet (see Figure 1),



Name

1.	hid	hay	head	had	haw	Hud	hood	hoe
2.	hid	hay	head	had	haw	Hud	hood	hoe
3.	hid	hay	head	had	haw	Hud	hood	hoe
4.	hid	hay	head	had	haw	Hud	hood	hoe
5.	hid	hay	head	had	haw	Hud	hood	hoe
BEEP								
6.	hid	hay	head	had	haw	Hud	hood	hoe
7.	hid	hay	head	had	haw	Hud	hood	hoe
8.	hid	hay	head	had	haw	Hud	hood	hoe
9.	hid	hay	head	had	haw	Hud	hood	hoe
10.	hid	hay	head	had	haw	Hud	hood	hoe

Figure 1. Part of the answer sheet used for recording listener's responses.

and subjects were allowed to listen again to a stimulus in cases of uncertainty. (In fact, most listeners took little or no advantage of this option). In addition, the subjects tested so far have all had phonetic training and are familiar with the type of task they were asked to perform.

In the design of this experiment, the following problem was encountered. Naive subjects, and to a certain degree even trained phoneticians, find it more difficult to pair isolated sounds with phonetic transcriptions than actual words with standard spellings, and are thus more likely to make extraneous errors of production and transcription if use is not made of actual words of English. On the other hand, most isolated vowel sounds do not correspond to English words. I attempted to overcome this problem by choosing [h] + Vowel as the basic token shape. Where an English word of this form exists, as in the case of *hay*, the speaker was instructed to simply say the word, while listeners found the printed word *hay* on the answer sheet. Where the vowel was such that [h] + Vowel was not an English word, as in the case of the vowel [ι], the speaker was instructed to say the word *h*-vowel-*d*, leaving off the *d*. Most speakers were able to do this in a very natural way, and none of the stimuli had any trace of consonant. For such vowels, the word h-vowel-d appeared on the answer sheet, and the listeners understood that they would hear only the [h] and the vowel.

This procedure may seem to represent a compromise. One is able to work entirely in terms of English words, but on the other hand one is dealing with [h] + Vowel syllables, rather than with isolated vowels. However, it is known that the syllable-initial [h] of English is merely a voiceless vowel of the same quality as the following (voiced) vowel. For this reason, [h] should not be expected to introduce consonantal transition information of the type hypothesized as aiding in vocal tract normalization.

Eight vowels were used in this experiment. Past tests have shown [iy], [uw], and [er] to be extremely easy to identify,[3] so these vowels were omitted in the interest of maximizing the number of presentations of "difficult" vowels in a given amount of time. Furthermore, in current Los Angeles speech the usual distinction between the low back vowels [ɑ] and [ɔ] has been lost, reducing the number of available vowels by one. Thus the vowels used were [ι, e, ɛ, æ, ɑ, ʌ, ɷ, o].

Each of the twenty speakers recorded eight stimuli, each stimulus consisting of two repetitions of [h] followed by one of the vowels; thus, [hι hι] was one of the stimuli, [he he] another, and so on. Of these 160 stimuli, only 80 were used, 10 tokens of each of the eight vowels. Thus on the average each of the 20 speakers was responsible for only four of the 80 vowels heard by the listeners.[4] These 80 stimuli were randomized and presented to listeners for identification.

---

[3] For example, in the Peterson and Barney (1952) experiment, [i] had a rate of correct identification of 99.9%, [u] 99.2%, and [er] 99.6% (in [h --- d] context).

[4] Consider an experiment in which a small number of speakers record a large number of vowels. Even if these stimuli are speaker-randomized, the possibility exists that listeners will become familiar with vocal characteristics

To date, ten listeners have been tested, and the identification scores have been very high. Two of the listeners identified all 80 vowels without error and another four made only one error. Of the four out of ten who made more than one error, the lowest score was 74 correct out of 80, or 92.5%. The overall rate of correct identification was better than 97%. Table 1 shows the confusions between individual vowels.

Table 1. Results of the listening test.

Intended vowel	[ι]	[e]	[ε]	[æ]	[a]	[ʌ]	[ɷ]	[o]
[ι]	100	--	--	--	--	--	--	--
[e]	--	100	--	--	--	--	--	--
[ε]	--	--	88	12	--	--	--	--
[æ]	--	--	1	99	--	--	--	--
[a]	--	--	--	--	98	2	--	--
[ʌ]	--	--	--	--	6	93	1	--
[ɷ]	--	--	--	--	--	1	99	--
[o]	--	--	--	--	--	--	--	100

High as these scores were, there is reason to believe they would have been even higher if dialect uniformity had been insisted upon. No fewer than 52% of the errors that were made involved identifying intended [ε] of *head* as [æ] of *had*. The significance of this fact is that most of the subjects responsible for the bulk of the errors were natives of the Midwestern dialect area, a dialectal characteristic of which is the raising of [æ] toward the mid tongue-height position of [ε]. Thus *had*, pronounced [h æ d] by the Los Angeles speakers, is commonly pronounced [h εˇə͓d] in the Midwestern region. The latter pronunciation is close enough to the Los Angeles pronunciation of *head* to offer an obvious explanation of why these Midwestern subjects tended to hear Los Angeles *head* as *had*. In contrast, only one error involved reporting *head* for intended *had*, supporting the view that dialectal variation, rather than lack of speaker normalization, is behind the majority of the observed errors. In any case, however, it should be kept in mind that the overall error rate was less than 3%.

---

of particular speakers, allowing a certain amount of normalization to take place in spite of the randomization. I believe I have avoided this problem by keeping the ratio of the number of speakers to number of stimuli quite high.

In spite of the large difference in observed error rate in the present speaker-randomized isolated-vowel experiment and that of Strange, Verbrugge, Shankweiler, and Edman, for example, the results of the present experiment do not challenge the validity of the findings of these researchers. Obviously the task presented to their listeners was a much more difficult one, resulting in much higher error rates, in turn allowing further experimentation to determine what additional factors would aid identification. (Strange's group found consonantal context, but not previous experience with a speaker's vowels, to significantly improve vowel identifiability). For its part, the present experiment seems to indicate that all necessary information for the unambiguous identification of an isolated vowel is contained in the signal and that while exposure to consonantal transitions and to several of a speaker's vowels makes the identification process easier by increasing redundancy, these latter factors are not essential ones.

References

Ladefoged, P. (1967). *Three Areas of Experimental Phonetics*, Oxford University Press.

Peterson, G. and Barney, H. (1952) 'Control methods used in a study of the vowels.' *J. Acoust. Soc. Amer.* 24:175-184.

Strange, W., Verbrugge, R., Shankweiler, D. and Edman, T. (1976) 'Consonant environment specifies vowel identity', *J. Acoust. Soc. Amer.* 60: 213-221.

Phonetic differences within and between languages

Peter Ladefoged

[Paper given at the 52nd annual meeting of the Linguistic Society of America]

A good description of a language describes not only the phonological patterns within the language but also how the sounds may be characterised in terms of some absolute phonetic standards. This is a necessary procedure if we are going to say how the sounds of one language differ from the sounds of other languages. I want to consider the problems that doing this poses for linguistic theory. I will use data from six languages that have contrasts in phonation type such as murmur or breathy voice versus regular voice. I want to point out that that this additional contrasting phonation type is not phonetically the same in the different languages.

Contrasts involving murmur in the Owerri dialect of Igbo, a Kwa language spoken in Nigeria, are illustrated in (1). Like Hindi, in addition to contrasting murmured and regularly voiced stops, Owerri Igbo also contrasts aspirated and unaspirated voiceless stops; but these are irrelevant here. We will be concerned only with the difference between murmured and regularly voiced sounds.

(1)	ίba	'to get rich'
	ίb̤ʱa	'to peel'
	ίda	'to cut'
	ίd̤ʱà	'to peel'

Figure 1 shows spectrograms of murmured velar stops in Owerri Igbo, as pronounced by a female and a male speaker. The murmur after the release of the stops can be seen at the time indicated by the arrow. Notice that it is not very long -- about 30 msec -- and that the vertical striations corresponding to the vibrations of the vocal cords are clearly visible throughout the release in the lower part of the spectrogram. Hindi also has a contrast between voiced and murmured aspirated stops as shown in (2). But it seems to me that the murmured sounds in Hindi are not like this. Figure 2 shows a murmured alveolar stop in Hindi, as pronounced by a female speaker. Notice that at the time of the murmur -- shown by the

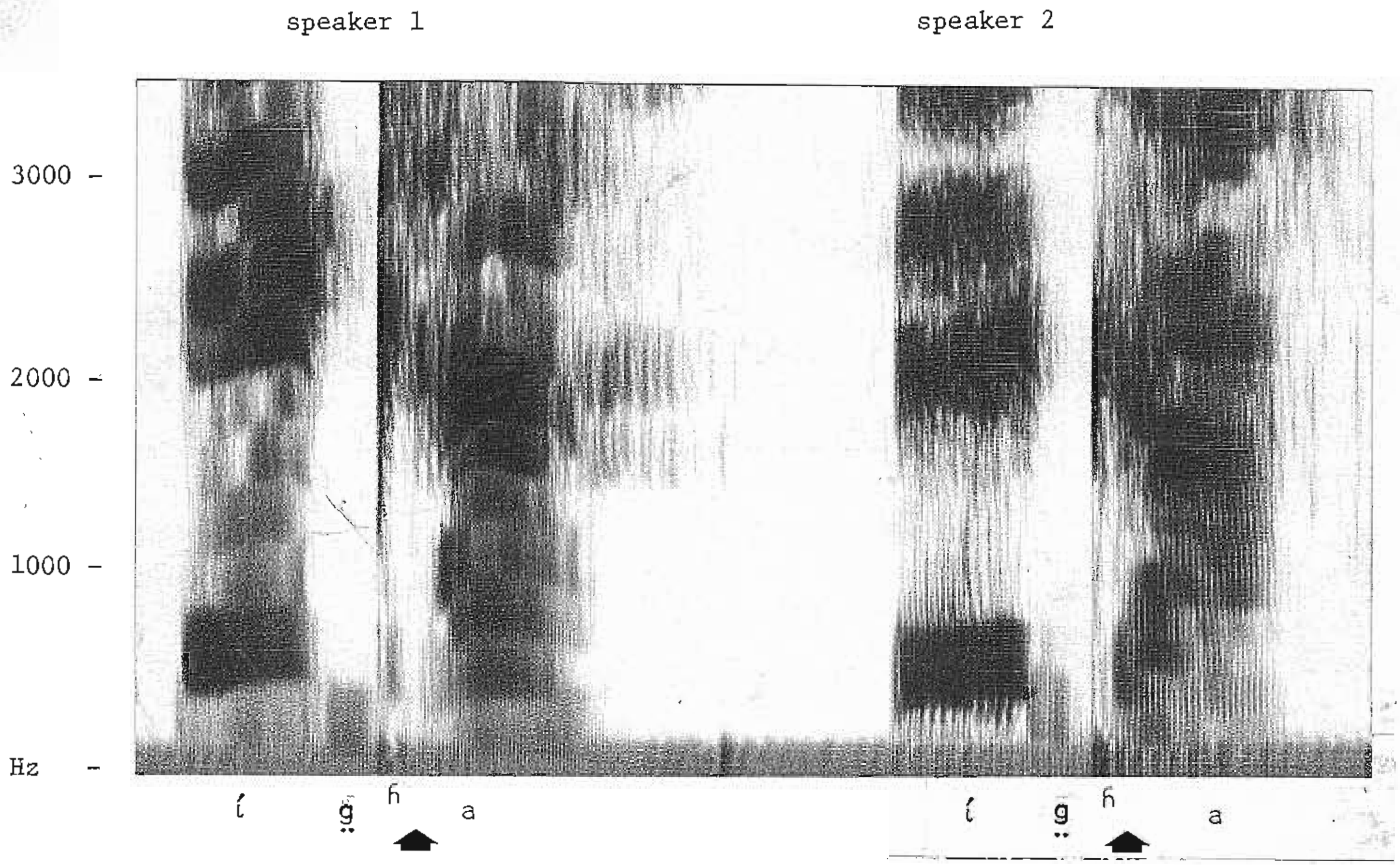


Figure 1. Spectrograms illustrating murmured aspirated plosives in Owerri Igbo. (After Ladefoged et al 1976).

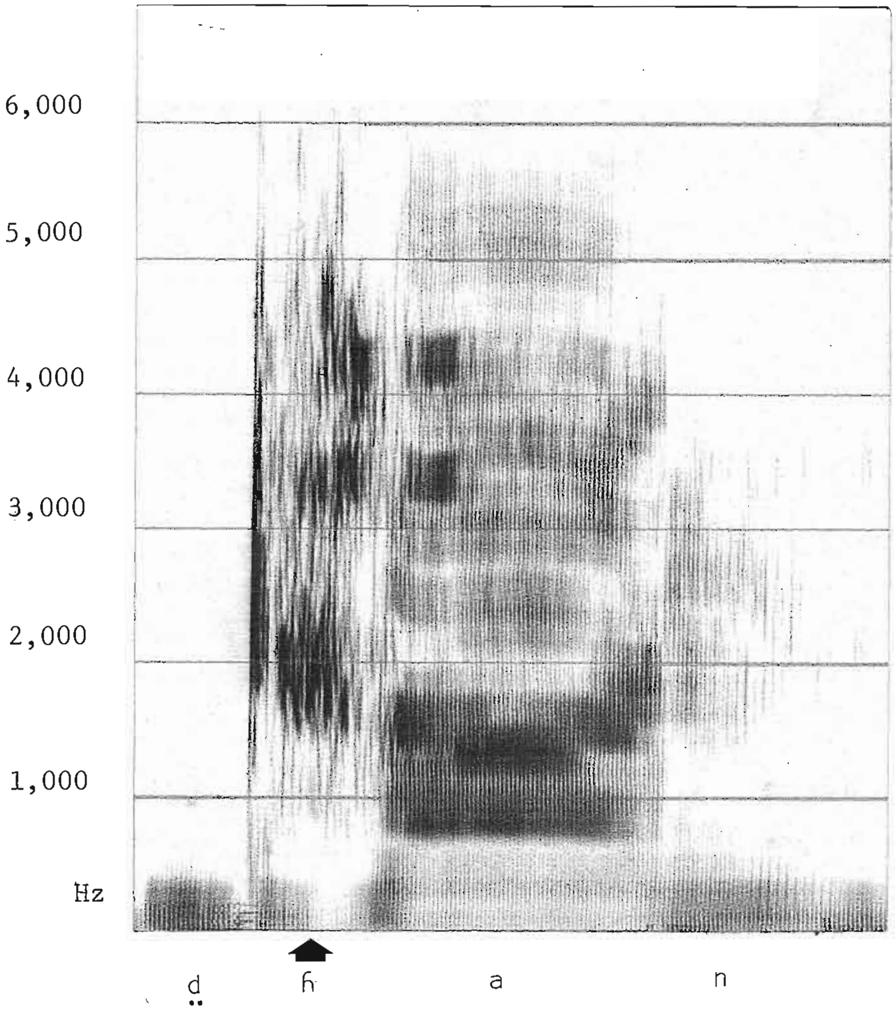


Figure 2. Spectrogram illustrating a murmured aspirated plosive in Hindi.

(2)	bal	'hair'
	b̤ʱal	'forehead'
	d̪an	'charity'
	d̪̤ʱan	'paddy'

arrow -- the voicing component is less visible than it is in Owerri Igbo. In addition, there is much more random fricative energy in the upper part of the spectrogram, and the murmured release lasts for considerably longer -- about 60 msec. On the basis of this limited evidence it seems as if the breathy voiced stops in Hindi are more breathy and less voiced than those in Igbo. This is also the impression one gets from listening to these languages.

Next I want to make the point that what is called murmur in Marathi may involve a different voice quality. Marathi has contrasting oral stops which are perhaps more like those of Igbo than of Hindi. In addition, as is shown in (3) Marathi also contrasts murmured and voiced nasals. Figure 3

(3)	mɑ:r	'caste'
	m̤ɑ:r	'beat'

shows spectrograms of these words. The vibrations of the vocal cords are clearly visible during the murmured nasal. (Here, as throughout this paper, I am using two dots below a symbol to indicate that it is in some sense murmured). Furthermore, although I cannot prove it, because I have not investigated enough speakers, I get the impression that in the oral stops Marathi has heavier voicing than Hindi.

Contrasts involving voiced and murmured vowels in Gujarati are given in (4). Some of these contrasts are illustrated in Figure 4. Again you can

(4)	bɑr	'twelve'
	bɑ̤r	'outside'
	pɔr	'last year'
	pɔ̤r	'dawn'

see that in this case the murmured vowel in the right hand spectrogram contains very substantial vibrations of the vocal cords. It differs from the regularly voiced vowel on the left in that the murmured vibrations of the vocal cords do not produce so much energy at the higher formant frequencies.

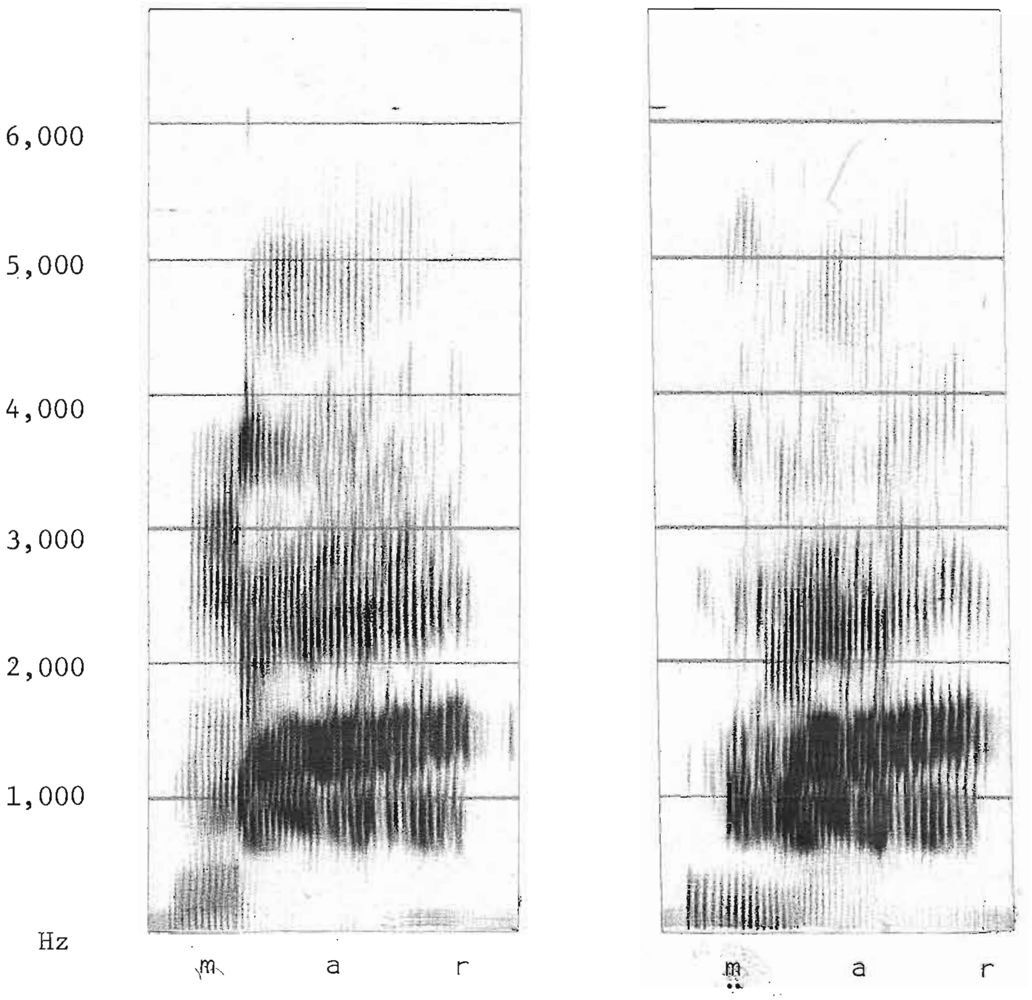


Figure 3. Spectrograms illustrating the contrast between voiced and murmured nasals in Marathi.

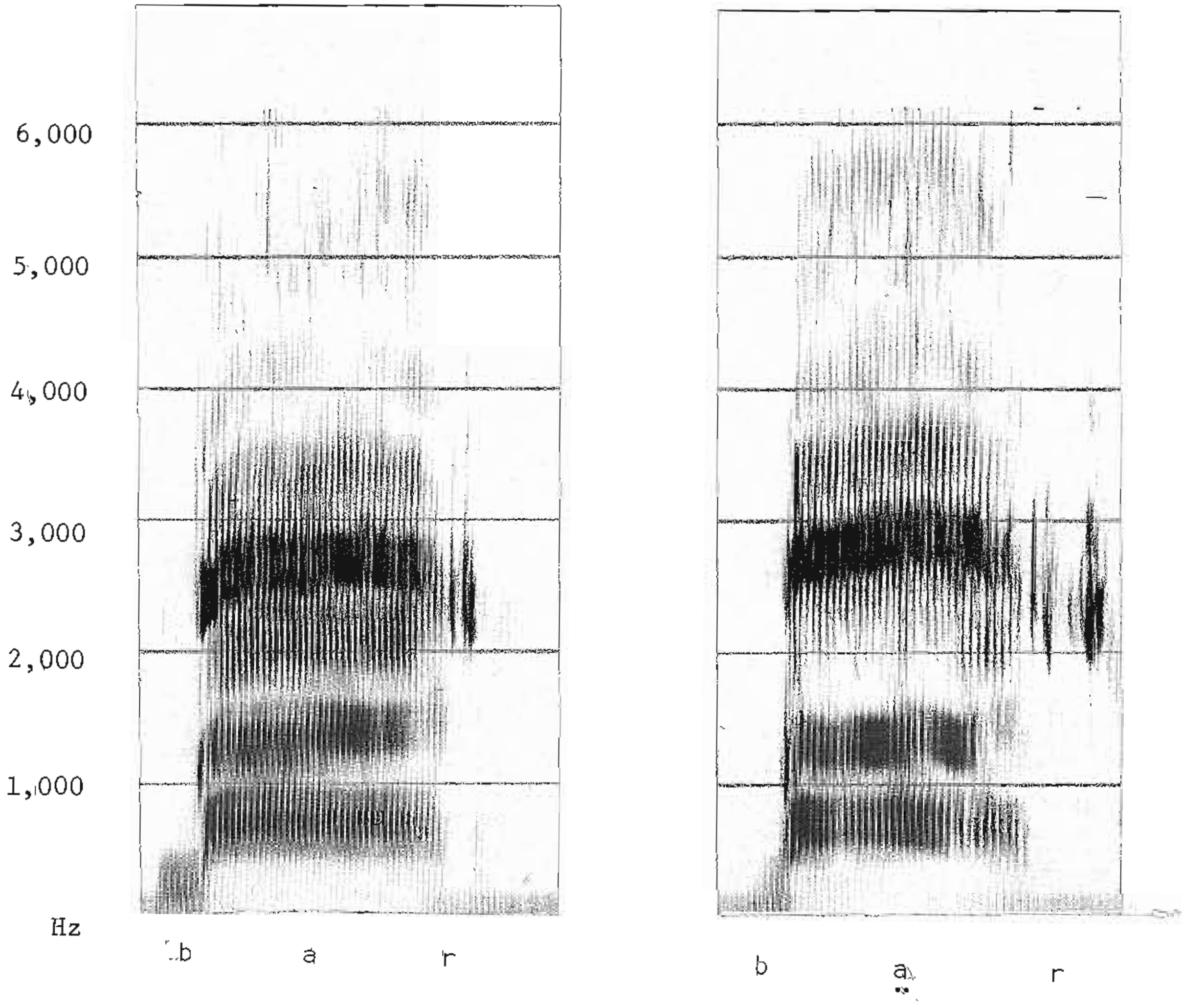


Figure 4. Spectrograms illustrating the contrast between voiced and murmured vowels in Gujerati.

Some Javanese contrasts are shown in (5). Phonologists usually trans-

(5)		
	pipi	'cheek'
	bibi	'aunt'
	kalī	'river'
	galī	'to dig'
	papa	'blunt'
	baba	'scene from a play'

cribe Javanese as if there were a series of voiced and a series of voiceless stop consonants. But it would be equally appropriate to say that the contrast is between a murmured stop, symbolized by [b], and a regularly voiced stop, confusingly symbolised by a [p]. Both stops have vibrating vocal cords; and at least part of the contrast is in the quality of the adjacent vowel, rather than in the stop. This can be seen to some extent in Figure 5. The first formant frequencies of the vowels in the first word are much lower than those in the second word. You can also see that both stops have voicing vibrations during the first part of the closure; and that both stops have about the same duration. As far as I can tell, the main physiological difference is that the so-called murmured or voiced stop is pronounced with a lowered larynx and somewhat lax vocal cords.

Now let me try to sum up the implications of all this for linguistic theory. We have seen that many languages use an additional glottal gesture, such as murmur. As far as I know, no language uses more than one such glottal gesture, in addition to the voiced and voiceless glottal states. But it is also apparent that what is called murmur in one language is not the same as murmur in another language.

Chomsky and Halle (1968) proposed that each phonological feature can be defined in terms of a single physical scale. In their view, both contrasts within a language and differences between languages should be expressed in terms of the values along some set of scales. Ladefoged (1971 and 1975) has a similar erroneous approach. It seems to me now that the relation between phonological features and phonetic scales is not so straightforward. If we ask what languages are made of the answer is in part, phonological oppositions that can be specified in terms of values of features, like voiceless, voiced, murmured, laryngealized. But the labels specifying phonological oppositions cannot be interpreted in terms of a single physical scale for each feature. If we ask what speech sounds are made of the answer has been provided for us by acoustic phonetics. It is things like formant frequencies, fricative amplitudes and source characteristics. Or, if we want to use physiological parameters, then it is vocal tract shapes, glottal strictures, and such things. From an acoustic or a physiological point of

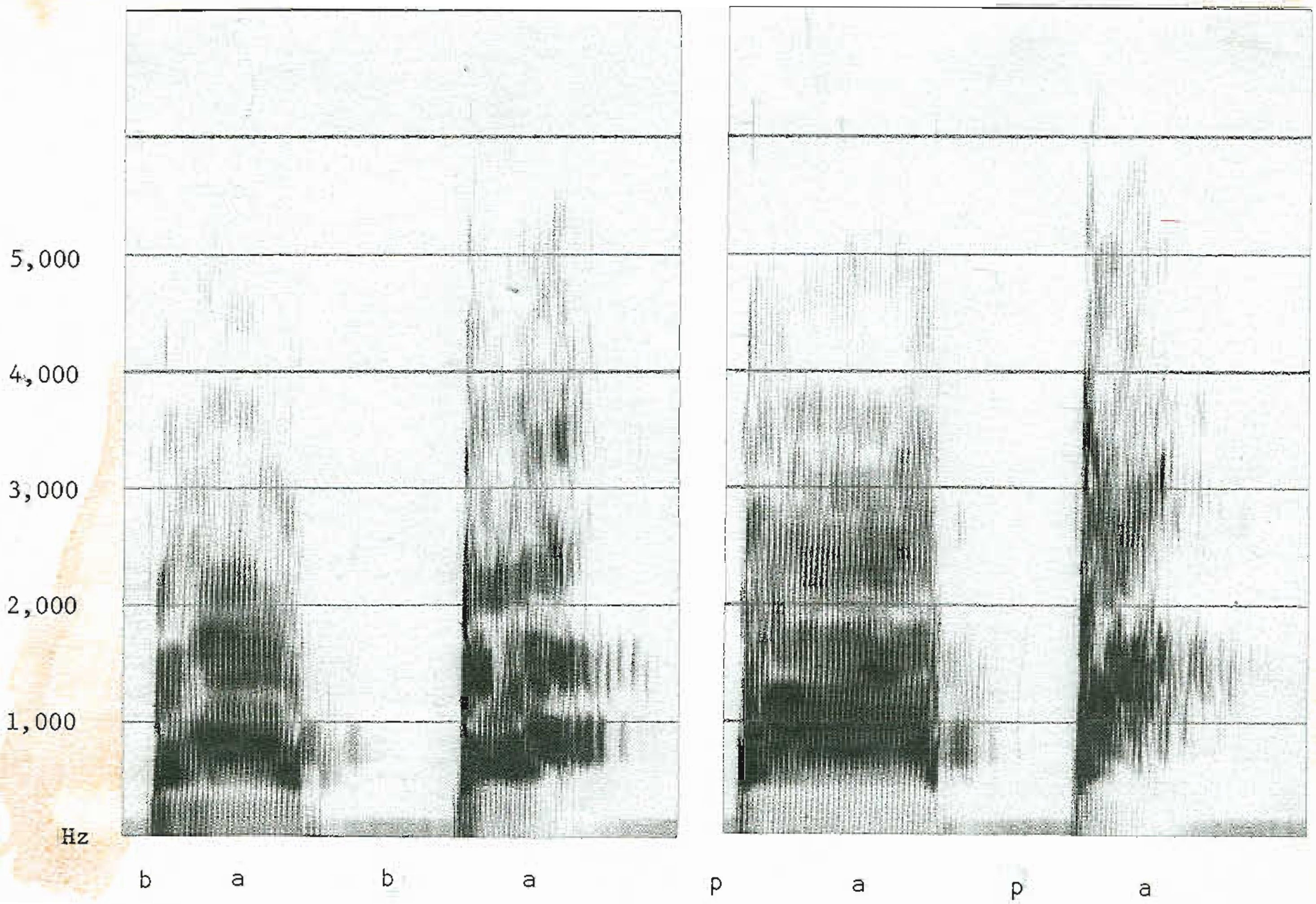


Figure 5. Spectrograms illustrating contrasting bilabial stops in Javanese.

view most phonological features are cover features, definable only in terms of complexes of phonetic parameters.

When we are just talking about the phonological oppositions within a single language we can get away with fairly casual descriptions of the phonetic facts. But when we are trying to compare languages, or to give accurate phonetic descriptions of a single language, then we must learn to reinterpret phonological oppositions in terms of complexes of the real phonetic elements of languages.

References

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English.* Harper and Row.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics.* Univ. of Chicago Press.

Ladefoged, P. (1975). *A Course in Phonetics.* Harcourt Brace and Jovanovich.

Ladefoged, P., Williamson, K., Elugbe, B. and Uwalaka, Sr. A.A. (1978). 'The stops of Owerri Igbo' *Studies in African Linguistics,* Supp. 6, 147-163.

Expectation affects identification by listening

Peter Ladefoged

[Paper presented at the 94th meeting of the Acoustical Society of America]

In many criminal cases suspects are identified simply by listening to tape recordings. The identifications are sometimes made by close acquaintances of the suspect, but they are also made by law enforcement officers who do not know the suspect very well. I have heard police officers claim to identify a suspect as a result of having heard his voice on one day six years previously "on another matter" (a legal euphemism for the occasion of his previous trial for the same offence).

It is a matter of common experience that our expectations affect the way we identify voices. When you are not expecting a telephone call from a particular person it is often hard to identify them. But if you are forewarned, or know that that person is likely to call, it is relatively easy. Similarly, if you suspect that a certain person might belong to a particular group, you are more likely to claim that you recognize his voice in a recording of that group --- even if in fact his voice has not been recorded. The frequency of such false positive identifications has now been demonstrated in an experiment.

At the time of the experiment there were 25 people actively associated with the UCLA Phonetics Laboratory, most of whom saw one another almost daily, conversed together frequently, and had known one another well, often for a period of many years. Of the 25 people, 10 were women and 15 were men. There were only two American Blacks in the group, both of them men. Recordings were made of 11 of the lab group, four women and seven men. Neither of the two Blacks was recorded. Instead a recording was made of another Black male, who was relatively unfamiliar to most members of the group.

All the recordings were made over the lab telephone, using a direct tap on the line. Speakers were called at other telephones, and were asked to identify themselves by a number that was assigned to them. They then talked for a period of 30 seconds to a minute on some neutral topic that could not be used to identify them, such as the weather or a program they had seen on T.V. It has been shown in previous experiments (Hecker 1971) that, on a good quality recording, there is no improvement in the ability to identify through listening to longer passages. Speakers either can or cannot be identified after listening for a short time. The comparatively high quality of the recordings was attested by spectrographic investigations.

The background noise was at a much lower level than in any of the recordings used in any legal case that I have ever been associated with. The frequency range was also far superior, extending (with some slight losses at both ends) from 100 to 4,000 Hz (as compared with 100 to 2,500 Hz in most of the recordings presented in evidence).

The recordings were edited, arranged in a random order, and played to 10 members of the Phonetics Lab group. The listeners were simply asked to try to identify the speakers. They were not told whether the speakers were, or were not,members of the Phonetics Lab. But since it was common knowledge that during the preceding week recordings had been made from the tapped phone in the Phonetics Lab, it is reasonable to assume that the listeners expected to hear the voices of the members of the group.

All the listeners except one correctly identified all the 11 members of the Phonetics Lab. Nearly all the identifications were made in less than 5 seconds for each speaker. The one listener who made an error in this part of the task confused two speakers who were not well known to him. Three of the ten listeners knew the Black speaker, who was not a member of the Lab group, and correctly identified him. Of the remaining seven listeners, only two said that they could not recognize this speaker. Five listeners wrongly identified this speaker as one of the two Blacks in the Phonetics Lab.

It is worth noting that four of the five listeners who made the wrong identification were highly skilled, experienced phoneticians. Two of them hold doctoral degrees in the field, and the other two were post M.A. students working on their Ph.D. theses. The fifth listener who made this error was an undergraduate computer programmer who had had some training in phonetics. All of them realized that the speaker was Black, and simply wrongly identified him as one of the Blacks in the Lab, because of their expectations.

References

Hecker, M. L. (1971). *Speaker recognition: An interpretive survey of the literature*. ASHA Monograph 16.

To appear in *Berkeley Linguistics Society* *IV. (1978)*

The Frequency of Tones.

Ian Maddieson
University of California, Los Angeles.

Introduction

In order to establish reliable estimates of the frequency of occurrence of tones of certain types, a sample of over 200 tone languages has been studied.

The language sample

There are a number of well-known difficulties in the way of constructing high-quality language samples, even if only currently-spoken languages are considered. A proper random sample cannot be taken since the number of languages spoken is not known - both because some languages have escaped attention altogether and because there are no agreed criteria as to what count as separate languages as opposed to dialects of the same language. Furthermore, the quality and quantity of information available on known languages varies enormously, as does the accessibility of the data. These last factors can lead to a language sample which is biassed through inclusion only of the most familiar languages, or those whose descriptions are most accessible.

The language sample in this study was selected according to a principle designed to avoid as many of these problems as possible. Each small group of closely-related languages, according to the best available genetic classifications, was considered. One tone language was selected to represent each group if one or more of the languages was tonal. Factors such as the clarity and accessibility of data only played a role in deciding which language to include from within a group. Languages with no (known) close relatives were treated as the equivalent of a group and were included. This approach ensures a language sample which includes the maximum of genetic diversity while representing language families in a manner which is roughly proportional to the number of tone languages they include. The problem of language/dialect discrimination is avoided, since cases which raise this problem obviously concern members of the came language group and only one representative can be included from each group. And the principle directs an outward search for the data to satisfy the requirements of the quota structure, thus reducing the influence of 'bibliographical convenience' (Bell 1978) on the resulting sample. Naturally, many problems concerning inadequacy and insufficiency of data still cannot be overcome, and time has placed limits on the extent to which the ideal sample has yet been approached. If all gaps could be filled, over 400 languages would have been included in the sample.

Interpretation and standardization of data

To make comparisons between languages meaningful, and derive valid generalizations, the languages must be described in a similar fashion. Otherwise, differences in descriptive paradigm are confounded with differences between languages at some common descriptive level. In this study the objective is to standardize the descriptive level at some central point along what is roughly an 'abstract-concrete' parameter ranking phonological descriptions according to the depth of the analysis in a given language and the degree of abstraction permitted by the theoretical model used. The reasons for this choice are partly pragmatic: many available descriptions do not supply enough information about higher-level processes to enable a more highly 'abstract' account to be developed; likewise, many descriptions give inadequate detail on phonetic realizations, etc, for a more concrete account to be developed. However, most of the descriptions which would fall toward either extreme of an 'abstract-concrete' parameter do provide the information required to move them toward the center.

The procedures involved in standardizing the data can be thought of as involving essentially three aspects:-

i) determining the number of basic contrastive tone units,
ii) determining the shapes of the tones,
iii) determining the number of contrastive pitch heights used in the tones and deriving a transcription.

The steps taken in determining the number of tones include attempting to correct over- or under-differentiation. Over-differentiation, resulting in a claim that a language has more tones than it is assigned here, can occur when predictable variant shapes of the same tone are considered distinct; when nontonal features such as contrasting syllabic structures or phonation types are held to distinguish 'tones' when the pitch patterns are in themselves similar; or when tone patterns are considered to be unitary despite evidence that a sequence of more basic tonal units is concerned.

Under-differentiation, resulting in a claim that a language has fewer tones, can occur when tones are considered the same because of a complementarity of distribution despite a lack of similarity in the tones themselves; when abstract or diacritic features are used to predict tonal differences; or when more complex tone shapes are unjustifiably regarded as composed of a sequence of simpler tones simply on reductionist principles rather than because evidence exists to show their sequential nature.

Each language was examined to determine how many tones it would have in the light of such considerations as the above. An example may help to illustrate how this was done. Standard Thai has a type of closed syllable with a short vowel followed by a stop consonant. These syllables can be either high-toned or low-toned. However, as the open or sonorant-final syllables may also have high or low tones at similar pitch levels, the case for recognizing additional distinct 'short tones' is not accepted, as this would be to over-differentiate. On the other hand, Gandour

(1975) has proposed that the (long) rising and falling tones should be regarded as sequences of low and high, and high and low tones respectively. This argument is based only on distributional criteria, rather than on showing that there are some ways that falling and rising tones behave like other sequences of high and low tones in Thai, or result from a coalescence of such units. Consequently, the elimination of rising and falling tones from the inventory is not accepted, since this would result in under-differentiation. The result is that Standard Thai is considered here as a language with 5 tones (it also has a mid tone) rather than a language with either 7 or 3 tones.

Beyond the kind of considerations outlined above, tones which are considered 'marginal' to a system have been eliminated from the basic inventory. These cases include tones which are of very rare occurrence in a language. Such marginal items are eliminated for two principal reasons. First, many available descriptions of languages reflect a level of familiarity too imperfect to be sure that such marginal items would have been reported. Hence comparability of descriptions is best served by matching such de facto elimination of marginal tones by eliminating them from the more detailed descriptions. Secondly, there is a sense in which the structure of an inventory of tones is best represented by an account which reflects the relative frequency of occurrence of tones within the language. The elimination of marginal tones represents a first crude step towards such an account.

A basic shape was determined for each tone in a system in terms of the relative pitch height of each tone and the direction and extent of any movement. The major effort was directed to correctly identifying basically level tones, requiring only one 'pitch target' for their characterization, and essentially contour tones, requiring two (or more) pitch targets for their characterization. In speech, F_0 is rarely held in a steady state. Pitch movement occurs because of co-articulation between tones in sequence, because of differences of phonation type and other segmental influences, because of stress placement and the superimposition of intonational contours, and because of mechanical factors inherent in the production of F_0. Whatever pitch movement can be reasonabl attributed to such factors is discounted in determining basic tone shapes. Thus many phonetically-occurring contours are regarded as resulting from the effects of such natural phonetic factors on tones which can be described with a single pitch target, i.e. they are level tones.

Likewise, factors such as co-articulation of tones and imposition of intonational patterns can result in an enlargement or a reduction or levelling of the pitch movement associated with an underlying contour tone. A basic shape for each contour tone when it is unaffected by the influence of these factors has been sought.

Some of the influences on tone shapes which are discounted can be summarized by such 'natural phonetic rules' as the following:-
a) unmarked intonational effects tend to cause level tones to fall, especially finally (this effect seems most apparent with the lowest

level tone), and tend to compress the range of rising tones while extending the range of falling tones;
b) tones tend to begin at a pitch level which is more central in the pitch range than their (first) target level.
The way in which such rules are used to interpret tone shapes may be illustrated with reference to Standard Thai again. One set of shapes for the 5 Standard Thai tones is given by Abramson (1962). These, shown below on Chart 1, represent 'citation shapes' of the tones and show quite marked pitch movement for all five tones. They may be compared with the shapes for tones in connected speech found by Gandour and Maddieson (1976), shown on Chart 2.

Chart 1

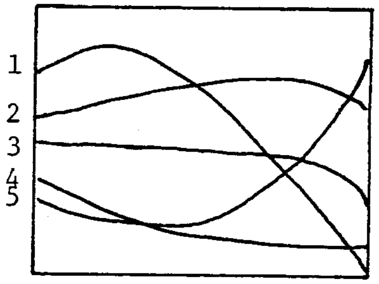
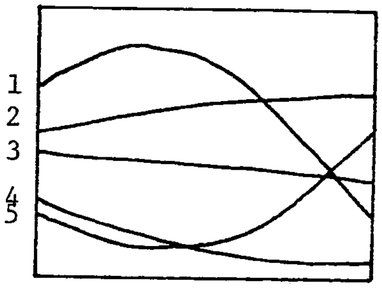


Chart 2



An effect of the citation context can be seen in that the final falling sections of Abramson's tones 2 and 3 can be seen to be absent in the corresponding tones on Chart 2. The overall downward speech melody still makes tones 3 and 4 descend somewhat on Chart 2, but in both charts tones 2, 3 and 4 occupy separate portions of the pitch range. Their shapes can be accounted for if only one pitch target is assumed for each of them, plus the effects of rules (a) and (b) above. They are considered to be high, mid and low level tones. Note that the high and low tones both begin at points closer to the mid tone than their end-points are in Chart 2. A centralized beginning to the tones can be seen also in the two remaining tones, 1 and 5. These two tones cross the pitch ranges of the level tones and thus must be characterized by more thzn one target. Because the centralized beginning is general it is not assumed to be part of the basic tone shape and hence these tones can be described with only two targets each. Tone 1 is regarded as a falling tone (not a rising-falling) and tone 5 as a rising tone (not a falling-rising). The greater extent of the pitch movement for the falling tone relative to the rising tone, which is more noticeable on Chart 2, can also be understood as an adaptation to an unmarked downward intonation contour. Imposition of falling intonation is facilitated by extending the downward pitch movement of a tone and truncating the upward pitch movement of a tone. Because of these considerations, the underlying pitch range covered by these two contour tones is assumed to be essentially the same, both traversing a distance equivalent to that between the high and low level tones.

Following determination of the tone shapes, tones were coded by means of a standard transcription employing a 5-point pitch scale. Each tone inventory was examined to determine how many distinct beginning- or end-points ('pitch heights') were required to characterize the tones in the inventory, and each pitch height was assigned to a number on the 5-point scale. Thus, for Standard Thai, 3 pitch heights were recognized, with a level tone at each height, a tone falling from the highest to the lowest of the three, and a tone rising from the lowest to the highest of the three. This transcription enables comparisons to be made concerning the proportion of the pitch-range covered by contour tones in those languages which use 3 or more pitch heights. In these cases the contour tones can be classed as using the whole of the pitch-range used for tones in the language, or as using only a part of it. It is not possible to make direct cross-language comparisons of the pitch heights themselves, as the transcription is derived independently for each language. This is necessary because the data to establish cross-language comparability of pitch phenomena is generally lacking, except in situations such as bilingualism or when borrowings occur between tone languages, see Maddieson 1977.

Results

At the present time, data from 207 languages has been examined. Some of the results are given below.

i) Number of tones in the inventory. The totals have been computed for the number of languages having phonological inventories containing 2 tones, 3 tones, 4 tones, etc. The totals are given in Table I.

Table I

no of tones :	2	3	4	5	6	7	8(+)
languages :	106	62	14	12	11	2	0

A striking result is that 2-tone systems are overwhelmingly the most frequent. Of the 207 languages, roughly 51% have 2-tone systems. This very striking predominance of 2-tone systems does not seem to have been noted before in the literature on general tonology. Systems with 3 tones are also common, being found in about 30% of the languages sampled. There are only small numbers of languages found with more than 3 tones, with systems of 4, 5, and 6 tones being about equally common (or perhaps more accurately, equally scarce). There are only about 6 or 7% of the languages surveyed that fall in each of these categories. It will be shown later that the step from systems of 3 tones to systems of 4 or more tones also corresponds with the point at which contour tones begin to play a significant part in the make-up of the tone inventories. A further sharp reduction occurs in the frequency of systems with more than 6 tones. Only two 7-tone systems are included in the sample, and there are none with 8 or more tones. Of course, this does not mean that such systems do not occur. The purpose of this study is not to compile a list of all the possible tone systems that do occur, but rather to derive estimates of how frequently systems with given characteristics occur. We may conclude only that probably less than 1 in 200 tone languages contrasts 8 or more tones.

ii) Frequency of tone types. The total numbers of tones of various types contained in the inventories of the languages sampled were also computed. The frequency of tone types is given in Table II.

Table II

level tones	:	465			84%
contour tones	:	87	falling	42	8%
			rising	38	8%
			fall-rise	4	.01%
			rise-fall	3	
total	:	552			

The most striking fact is that, overall, 84% of the tones in the languages sampled are level tones: the number of level tones vastly outweighs the number of contour tones. It will be shown below that this is largely a result of the structure of the common 2- and 3-tone systems.

Among the contours, falling and rising tones are effectively equally frequent in occurrence. The complex bidirectional tones occur very rarely, but both types (falling-rising and rising-falling) occur about as often as each other. These results conflict with earlier claims that falling tones outnumber rising tones in the languages of the world (cf Cheng 1973, Ohala 1973, Hombert 1978). The liklihood is that this claim reflects patterns observed in surface pitch contours, rather than reflecting inherent shapes of tones. It will be seen later that an implicational universal stating that an inventory must have a falling tone before it can have a rising tone would also be false. It seems that any productive or perceptual mechanism that may favor falling pitch in speech does not lead to more falling than rising underlying shapes for tones.

iii) Frequency of tone systems. The frequency of tone systems of various structures was also examined. A summary of the distribution by languages of contour tone types is given in Table III.

Table III

languages with only level tones	:	163	(79%)
languages with level(s) & simple contour(s)	:	37	(18%)
languages including complex contours	:	7	(3%)

Languages whose tone systems contain only level tones comprise the vast majority of tone languages. There are only 37 languages in the sample with systems consisting of a level tone or tones and a simple contour tone or tones, and only 7 including complex contours as well.

The makeup of each individual system in terms of level (L), falling (F), rising (R), falling-rising (FR) and rising-falling

(RF) tones was also analyzed and the results are given in Table IV. This shows the number of languages in the sample with systems containing each different combination of tones found. Systems with 2 tones, 3 tones, 4 tones, etc are listed separately. So, for example, under 3-tone systems an entry reading 2L + F : 7 means that there were 7 languages whose systems consisted of 2 level tones and one falling tone.

Table IV

2-tone systems 2L : 106

3-tone systems 3L : 49 2L + F : 7 2L + R : 3
L + F + R : 2

4-tone systems 4L : 7 3L + F : 2 3L + R : 2
2L + F + R : 1 L + F + R + FR : 2

5-tone systems 5L : 1 3L + 2F : 1 3L + 2R : 1
3L + F + R : 7 2L + F + R + FR : 1
L + F + 2R + FR : 1

6-tone systems 4L + 2R : 1 3L + F + 2R : 3 4L + F + R : 1
3L + 2F + R : 1 2L + 2F + 2R : 1 2L + F + 3R : 1
2L + 2F + R + RF : 1 2L + F + 2R + RF : 1

7-tone systems 3L + 2F + 2R : 1 3L + F + 2R + RF : 1

In all of the languages with inventories containing only 2 tones, both of the tones are level. Although other types of 2-tone systems may occur, they must be regarded as decidedly rare - probably occurring in less than 1 in 200 tone languages. 3-tone systems are also predominantly found to contain only level tones. This is the case for 80% of the 3-tone languages, the remaining 20% having systems including 1 falling or rising tone, or, in a few cases, both. Although the most frequent single type of 4-tone system has 4 level tones, there are as many 4-tone systems which include a contour or contours as those which consist of only level tones. A language with 5 tones which are all level is a rarity. There is only one in the sample. The most common type of 5-tone system is one containing 3 levels, one fall and one rise. This is represented by 7 of the 12 languages concerned, i.e. by about 58% of them. Clearly, when the system contains as many as 5 tones contours are normally found to be included. All the 6- and 7-tone systems contain contours; no systems with 6 or 7 level tones are included in the sample and, in fact, no such systems have been reported in the literature.

Thus, although Wang's (1967) claim that systems with 4 tones must include at least one contour tone is clearly false, contour tones are certainly not unexpected in systems with 4 tones, they are highly probable in systems with 5 tones, and are required in systems with 6 or more tones. As noted earlier, contour tones are rare in 3-tone systems, but become quite common in 4-tone systems. This dividing line which falls between 3 and 4 tone systems is also marked by the restriction of complex contours to those languages which have 4 or more tones.

An odd result is that falling-rising tones only occur in 4- and 5-tone systems, and rising-falling tones only occur with 6- and 7-tone systems in the languages sampled. In view of the small numbers involved, this is likely to be only a sampling error with no significance.

iv) The extent of contours. When 3 or more pitch heights are contrasted in a language, the portion of the pitch range which contour tones cover can be compared. For example, in a system which includes 3 level tones and 1 falling tone, the fall may extend over the whole pitch range from high to low, or may extend only over a part of the range, i.e. from high to mid or from mid to low. The data in Table V show how many of the simple contours in languages with 3 or more contrasting pitch heights cover the whole range or a part of the pitch range. Data are given separately for languages whose inventories include only 1 falling tone (F) or 2 falling tones (2F), and only 1 rising tone (R) or 2 rising tones (2R), as well as totals for all contour tones, all falling tones and all rising tones.

Table V

	falling		rising		totals		
	F	2F	R	2R	falls	rises	all
whole pitch range	10	5	4	5	15	9	24
partial pitch range	3	7	5	11	10	16	26

An interesting pattern emerges in Table V. The extent of contours is dependent on the direction of movement of the contour and the number of tones in the system that move in a given direction. Two assymetries stand out in the table (the crucial numbers are underlined). Although full-range and partial-range contours are about equally frequent overall, when there is only one falling tone in the system, it most often extends over the whole tone range. On the other hand, when there are 2 rising tones in the system, the most commonly found type covers only a part of the tone range. It is clearly not the case that simple contours usually cover only a part of the tone range, as suggested by Wang (1967).

The observed assymetries require explanation. As a first hypothesis, it may be suggested that there is systemic pressure to distinguish 2 tones which move in the same direction by separating them into 2 distinct registers. Both 2F and 2R systems contain more partial-range contours. But there may also be phonetic factors, such as the tendency of falling tones to be extended and rising tones to be truncated, which result in a greater effect of this pressure on the underlying shapes of falling than rising tones.

Where distinguishing between 2 tones which move in the same direction is not a factor, a falling tone is generally found to cover the whole pitch range.

Tone universals

This study also provides an opportunity to test the reliability of generalizations concerning tones, including proposed 'tone universals'. In this connection, it has already been noted that falling and rising tones are equally common and that systems with 4 level tones and no contours occur. Wang's claim that no 4-tone systems occur without at least 1 contour could have been tested by a different strategy involving a direct search for counterexamples; but generalizations such as the claim that falling tones outnumber rising tones can only be tested by constructing a language sample in which the frequencies of tone types are represented in their true ratios. In research on language universals, the direct construction of hypotheses and search for counterexamples is complemented by studies such as the present one using sampling techniques.

For example, an implicational heirarchy has been proposed concerning tone types (Maddieson 1978). This is summarized by the two statements below:-

a) If a language has contour tones it also has level tones.
b) If a language has complex contours it also has simple contours.

These say that no language would have a contour tone without at least 1 level tone in its inventory, and no language would have a complex tone without at least 1 simple contour in its inventory. None of the languages in the sample violates this heirarchy. Because of the way the sample is constructed it can be said that at the very least the normal pattern of distribution of tone types is captured by this heirarchy.

It is important to note that this heirarchy imposes more highly-structured constraints on the distribution of simple and complex contours in tone inventories than are imposed by the simple fact of their rarity. For example, if contour tones were randomly distribin languages in the same overall ratio to level tones represented in the sample, several 2 or 3-tone systems containing only contours would have been expected to occur.

In fact, the distributional limits on tone types seem even more highly structured than this heirarchy indicates. This can be seen from the data in Table VI, which compares the overall ratio of 84% level tones to 16% contours in the whole sample with the separate percentages of level and contour tones in systems containing different numbers of tones.

Table VI

	number of tones in the system						
	2	3	4	5	6	7	overall
% level	100	93	78	58	47	43	84
% contour	0	7	22	42	53	57	16

The overall ratio is most closely matched by the ratio in 4-tone systems. However, this overall ratio is not at all typical of the systems with other than 4 tones. Instead the level to contour ratio is highly dependent on the number of tones in the system; the percentage of levels decreases and the percentage of contours increases as the number of tones increases. In systems with 6 or more tones a majority of the tones involved are contours.

Concluding thoughts

This paper has reported on the frequency of tones and tone systems. It has served to highlight such facts as that the most 'normal' tone system is one with only two level tones. It has provided new, reliable data concerning the relative frequency of level tones and contours of different types. It has provided an opportunity to test certain ideas concerning the constraints on the composition of tone inventories, and revealed that there is more to be done before these are understood better. It provides a source of data from which to derive expectations of the likely historical changes as tone systems evolve.

Note: This work was supported by a grant from the NSF.

References

Abramson, A. 1962. "The vowels and tones of Standard Thai: acoustical measurements and experiments" International Journal of American Linguistics, Publication 20.

Bell, A. 1978. "Language samples" in Universals of Human Language, edd. J.H. Greenberg et al., Stanford University Press.

Cheng, C-C. 1973. "A quantitive study of Chinese tones" Journal of Chinese Linguistics 1: 93-110.

Gandour, J. & I. Maddieson. 1976. "Measuring larynx movement in Standard Thai using the cricothyrometer" Phonetica 33: 241-267.

Hombert, J-M. 1978. "A model of tone systems" in Tone and Accent, ed. D.J. Napoli, Georgetown University Press, Washington D.C.

Maddieson, I. 1977. "Tone loans: a question concerning tone spacing and a method of answering it" Studies on Tone (UCLA Working Papers on Phonetics 36): 49-83.
Maddieson, I. 1978. "Universals of tone" in Universals of Human Language, edd. J.H. Greenberg et al., Stanford University Press.

Ohala, J.J. 1973. "The physiology of tone" in Consonant Types and Tone, ed. L.M. Hyman, University of Southern California.

Wang, W. S-Y. 1967. "The phonological features of tone" International Journal of American Linguistics 33: 93-105.

The interaction of tone and vowel quality

Eric Zee

[Revised version of a paper presented at the 94th meeting of
the Acoustical Society of America]

1. INTRODUCTION

It has been widely attested that the production of tones often involves the movement of the larynx, and also that a higher pitch is correlated with a higher larynx position. Parmenter et. al. (1933) conducted an X-ray investigation of the changes which take place in the articulation of a vowel when it is pronounced first on a low pitch and then on a high pitch. Three speakers, speaking English, French and Spanish respectively, were used. They concluded that important modifications occur in the articulation of a vowel when its pitch is raised an octave. These modifications include (1) the raised larynx, (2) changes in the supraglottal cavities which are probably due to the raised larynx, and (3) changes in these cavities which are probably not due to the raised larynx. The correlation of higher pitch with higher larynx height has been further investigated recently by phoneticians, using different techniques, such as Vanderslice's cricothyrometer (1967) and Ewan and Krones' thyroumbrometer (1972, 1974). Results similar to that of Parmenter's were reported. It is quite obvious that a raised or lowered larynx will alter the length of the pharynx and thus the length of the entire vocal tract, leading to a change in formant frequencies. Based on the assumption that a raised or a lowered larynx will shorten or lengthen the pharynx, Sundberg and Nordström (1976) investigated the effect of raised larynx on vowel formant frequencies. Two subjects, a phonetician and a singer, were used in their experiment. Both were able to control fairly well the positioning of their larynges. During phonation, the subjects were instructed to keep a finger on the thyroid cartilage to check the larynx position. Their findings in regard to the effects on the formant frequencies of a raised larynx were: (1) a substantial rise in the second formant frequency in high front vowels, (2) a rise in both the first and the second formant frequencies in open vowels, and (3) a rise in several vowels in both the third and the fourth formant frequencies.

The present study investigates the effect of tone on vowel quality by using a tone language. It differs from Sundberg and Nordström's in that it does not attempt to isolate pharyngeal change as a single factor that affects formant frequencies. Thus, speakers were not instructed to consciously lower or raise their larynges during phonation or to consciously sustain a vowel, nor were they instructed to keep a finger on the thyroid cartilage. The study explores how formant frequencies change when a vowel is pronounced with different tones in natural speech produced by speakers of a tone language.

2. PROCEDURE

Five Taiwanese Chinese vowles [i], [e], [a], [ɔ] and [u] were used for investigation. These vowels were uttered with two contrasting tones, which are the high tone and the low tone. These vowels, associated with either the high or the low tone, are meaningful monosyllabic morphemes in this dialect, as shown below:

	High tone	Low tone
[i]	'he'	'meaning'
[e]	'to nudge'	'to strangle'
[a]	'Asia'	'secondary'
[ɔ]	'dark'	'bay'
[u]	'mud'	'dirty'

A word list containing these vowels, 10 repetitions for each tone class, was prepared. These test words were arranged in a random order and each vowel was placed in the carrier frame as shown below:

[guá	tçÍ-má	t^hàʔ	___	tài-kēi	t^hĨ́a]
I	now	read	___	everyone	listen
'I	now	read	___	everyone	listens'

Three male native Taiwanese speakers, all in their late twenties and all from southern Taiwan were instructed to produce these tokens at a normal rate of speech. The recording was done in a single session for each speaker in a sound treated room. The data was analyzed by using a PDP-12 computer at the UCLA Phonetics Lab. Formant frequency values were obtained from LPC spectra and fundamental frequency from the corresponding Cepstrum. For each vowel, values were obtained for five tokens with a high tone, and five tokens with a low tone. In some cases, additional tokens were analyzed when the LPC did not give complete formant frequency values for a vowel. The LPC has a window size of 25.6 msec and formant frequency values were obtained every 10 msec. Each vowel associated with either the high or the low tone was analyzed entirely, that is, from the beginning to the end of its waveform . The formant frequency values for each 10 msec interval were averaged for each token. The averaged value which was taken to be the formant frequency (F1, F2, or F3) value for a vowel is shown in Tables, I, II and II for Speakers 1, 2 and 3. The reason for taking the average formant frequencies for each token is because there is an interdependency of formant frequencies and harmonics (Fant, 1973). The dominating harmonics have an influence on the formant frequencies, as indicated in Figure 1 which shows the F_0 and F1 contours as well as the 5th harmonics for vowel [e] produced with a low tone by Speaker 1.Notice how F1 is closer to the fifth harmonic at the start,and then becomes dominated by the fifth harmonic.Thus the value at any one time point for the apparent F1 does not necessarily represent the true value of the formant frequency. Fant (1973) suggests a method of visual interpolation, which is plausible in many circumstances. However, in a long vowel in which F_0 is changing,as it is in the low tone, it is more accurate to average the values across all the time points. This will of course average out any real variations in formant frequency that are due to changes in the vocal tract configuration during the production of the vowel. But, in the case of the present study, this will merely support the null hypothesis (that are no variations in formant frequency associated with variations in tone) and thus will not lead us to any false conclusions.

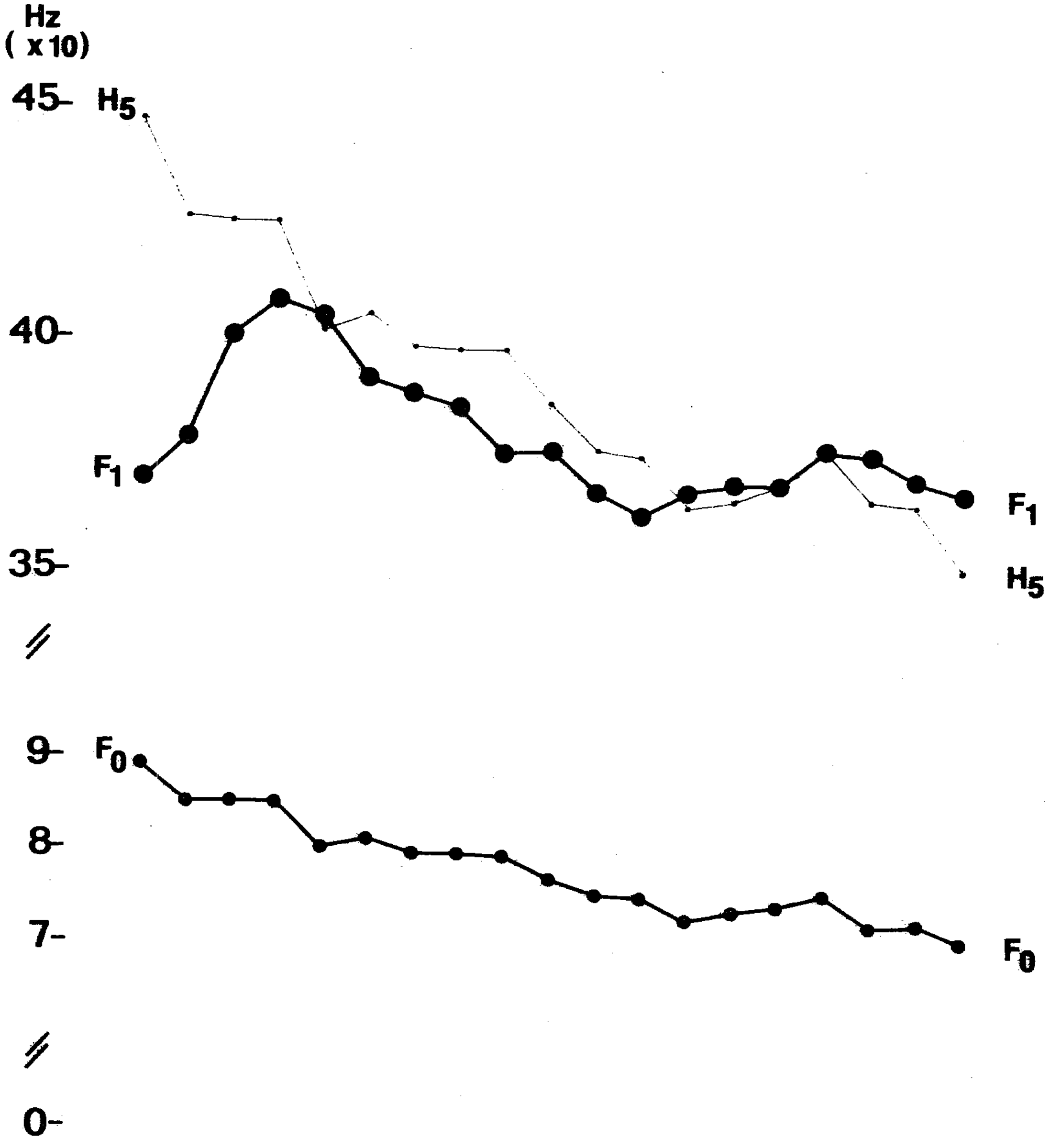


Figure 1. Vowel [e] spoken on a low tone by Speaker 1, showing variations in formant frequency due changes in F_0

It should also be noticed that it is often impossible for the LPC analysis to extract certain formant frequencies for certain vowels. For all three speakers, the second formant frequency for the vowels [a], [u] and [ɔ] often could not be tracked by the LPC. Ladefoged (1967), analyzing formant frequencies using a sound spectrograph, encountered similar difficulties. He found that 'both [ɔ] and [u] often (39 per cent of the time) could not be analysed in terms of three formants, because it was impossible to locate the position of formant two; and the complete analysis of [u] was usually (84 per cent of the time) impossible. In the case of these vowels as with [a], it would seem unlikely that any formant tracking machine would be any more successful in locating the formant frequencies' (p. 90). To solve the problem, the second difference of the Linear Prediction Spectrum (LPC2D) was used. The LPC2D was able to locate the second formant frequencies for most of the tokens, except for the vowel [u] produced with the low tone by Speaker 2, where it was successful on four (out of 10) tokens only. The LPC2D has a window size of 51.2 msec and the values of formant frequencies were obtained every 25.6 msec. Due to the different formant frequency analysis methods, the vowels [i], [e] (by LPC) are expected to occupy slightly different areas in the acoustic vowel space in relation to the vowels [a], [u] and [ɔ] (by LPC2D). This difference however is not particularly relevant for our study, since we are concerned with the locations of each individual vowel produced with different tones.

3. RESULTS

The results of analysis for Speakers 1, 2 and 3 are shown in Tables I, II and III respectively. They show the values of F1, F2 and F3 for each vowel associated with the high or the low tone. Also shown are the average fundamental frequency for the high or the low tone for each vowel. There are five values for each category (F_0, F1, F2 or F3), representing five different tokens.

Tables Ia, IIa and IIIa show the means and the standard deviations of the values of F1, F2 and F3 for each vowel associated with the high or the low tone for Speakers 1, 2 and 3 respectivley. They also show the means of the values of fundamental frequency and the results of grouped data t-tests for formant frequencies associated with the tones for each vowel.

The level of significance of the results of grouped data t-tests for formant frequencies associated with the high or the low tone for all three speakers are summarized in Table IV.

Figures 2, 3 and 4 show vowel ellipses associated with the high tone and the low tone. For each vowel cluster of 5 tokens, associated either with the high tone or with the low tone, an ellipse was drawn along axes which are oriented along the principal components, with the radii being two stan-

Table I. Values of F_0, F1, F2 and F3 for vowels associated with the high and the low tones for Speaker 1

	[i]		[e]		[a]		[ɔ]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
F_0	132	885	120	77	119	78	129	83	136	83
	134	78	117	75	123	78	136	81	135	88
	139	78	125	82	120	79	132	84	135	83
	138	81	137	76	124	79	133	89	134	85
	126	83	123	77	118	79	137	85	145	81
F1	277	278	383	379	881	889	533	546	293	290
	273	264	396	375	892	906	518	530	284	282
	275	269	399	378	868	905	532	526	283	291
	274	267	397	377	888	974	508	507	283	293
	263	275	389	376	842	1030	531	523	293	310
F2	2571	2385	2316	2231	1159	1182	787	682	829	602
	2728	2476	2374	2180	1160	1185	835	657	757	581
	2607	2467	2368	2274	1217	1209	731	648	821	582
	2608	2360	2345	2241	1226	1258	760	641	797	628
	2562	2460	2350	2271	1193	1246	820	677	825	609
F3	3294	3438	2744	2713	2351	2419	2591	2635	2606	2521
	3337	3469	2794	2740	2335	2446	2466	2634	2611	2486
	3265	3520	2827	2889	2309	2434	2280	2509	2620	2531
	3383	3350	2727	2869	2284	2421	2351	2463	2587	2441
	3347	3448	2906	2886	2374	2405	2496	2585	2654	2494

Table Ia. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 1

		F_0	F_1			F_2			F_3		
Vowel	Tone	Mean	Mean	SD	T-score	Mean	SD	T-score	Mean	SD	T-score
[i]	High	134	274	5.5	0.51	2616	66.0	4.90	3325	46.2	3.47
	Low	81	271	5.8		2430	55.2		3445	61.8	
[e]	High	124	393	6.6	5.16	2351	22.8	5.60	2799	71.5	0.39
	Low	77	377	1.6		2239	38.0		2819	1.4	
[a]	High	121	874	20.2	2.38	1191	31.2	1.20	2331	35.3	5.48
	Low	79	941	59.4		1216	34.7		2425	15.6	
[ɔ]	High	133	524	11.0	0.25	787	42.6	6.07	2437	122.9	1.98
	Low	84	526	14.0		661	17.9		2565	76.8	
[u]	High	137	289	5.2	0.78	806	30.0	12.80	2616	24.6	6.29
	Low	84	293	10.3		600	19.7		2495	35.2	

Table II. Values of F_0, F1, F2 and F3 for vowels associated with the high and the low tones for Speaker 2

	[i]		[e]		[a]		[ɔ]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
	159	108	145	100	145	106	152	103	153	105
	154	103	140	102	138	106	148	101	162	106
F_0	158	108	152	111	143	106	143	103	155	104
	153	106	144	108	138	104	152	113	159	105
	151	103	149	96	144	108	149	102	157	107
	303	275	406	379	963	946	450	393	310	313
	300	265	410	366	916	960	422	396	315	328
F1	302	289	408	387	969	976	423	417	318	318
	299	279	393	379	1007	974	431	387	321	324
	297	283	424	374	895	976	430	371	307	326
	2244	2307	2047	2197	1273	1242	794	770	753	537
	2176	2342	2036	2178	1233	1226	697	728	778	620
F2	2218	2309	2077	2307	1308	1267	739	781	777	573
	2223	2289	2139	2203	1254	1225	740	783	806	675
	2209	2249	2162	2329	1168	1267	774	786	748	-
	2890	2991	2732	2763	2405	2621	2846	2853	2491	2503
	2942	2984	2709	2680	2468	2619	2790	2814	2515	2598
F3	3020	3042	2714	2750	2559	2698	2721	2764	2571	2512
	2902	3015	2799	2744	2445	2618	2846	2840	2607	2626
	3031	2950	2732	2798	2554	2640	2780	2815	2451	2456

Table IIa. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 2

Vowel	Tone	F_0	F_1			F_2			F_3		
		Mean	Mean	SD	T-score	Mean	SD	T-score	Mean	SD	T-score
[i]	High	155	300	2.4	5.28	2214	24.8	4.53	2957	65.5	1.18
	Low	106	278	9.0		2299	34.0		2996	34.5	
[e]	High	146	408	11.1	5.18	2092	55.9	3.76	2737	36.0	0.39
	Low	103	377	7.7		2243	69.7		2747	42.9	
[a]	High	142	950	44.6	0.82	1247	52.1	0.07	2486	68.0	4.50
	Low	106	967	12.1		1245	20.9		2639	34.0	
[ɔ]	High	149	431	11.3	4.28	749	37.2	1.05	2800	56.9	0.57
	Low	104	392	16.6		770	24.0		2817	34.1	
[u]	High	157	314	5.7	2.02	772	23.2	5.95	2527	62.4	0.28
	Low	105	322	6.2		601	59.8		2539	70.7	

Table III. Values of F_0, F1, F2,and F3 for vowels associated with the high and the low tones for Speaker 3

	[i]		[e]		[a]		[ɔ]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
F_0	176	114	167	120	162	122	167	123	178	123
	184	112	165	121	169	115	175	110	175	122
	185	107	168	111	164	112	167	114	176	121
	178	108	166	116	166	116	170	117	178	114
	180	113	168	117	169	120	170	114	175	120
F1	294	246	400	395	570	655	505	357	349	304
	309	255	389	382	502	675	516	426	343	317
	328	252	397	402	463	684	499	408	344	327
	316	276	411	377	466	668	506	436	349	318
	329	252	395	396	489	655	496	379	338	301
F2	2416	2307	2325	2221	1311	1216	858	552	628	547
	2428	2137	2441	2205	1231	1269	921	619	672	565
	2480	2247	2427	2273	1302	1252	952	577	551	583
	2444	2142	2252	2216	1322	1263	942	625	628	569
	2496	2144	2459	2226	1246	1247	976	568	540	564
F3	3520	3491	3125	3053	2847	2539	2756	2892	2676	2959
	3406	3397	3111	3311	2702	2416	2736	2934	2734	2811
	3468	3416	3142	3239	2830	2508	2873	2793	2639	2879
	3429	3520	3210	3388	2755	2523	2922	2696	2676	2776
	3416	3401	3285	3167	2784	2453	2918	2711	2735	2765

Table IIIa. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 3

Vowel	Tone	F_0	F_1			F_2			F_3		
		Mean	Mean	SD	T-score	Mean	SD	T-score	Mean	SD	T-score
[i]	High	181	315	14.4	7.11	2453	34.1	6.80	3448	46.8	0.08
	Low	110	256	11.5		2195	77.5		3445	56.6	
[e]	High	167	410	29.0	1.46	2395	55.4	6.10	3174	72.4	0.86
	Low	117	390	10.5		2228	26.2		3231	129.3	
[a]	High	166	498	43.4	8.38	1282	41.0	1.61	2783	58.4	8.49
	Low	117	667	12.7		1249	20.6		2488	51.6	
[ɔ]	High	170	505	9.3	6.81	930	44.7	13.86	2841	89.1	0.58
	Low	115	401	32.9		588	32.2		2805	106.2	
[u]	High	176	345	4.6	5.97	604	56.3	1.48	2692	41.6	3.59
	Low	116	313	10.7		566	12.9		2838	80.9	

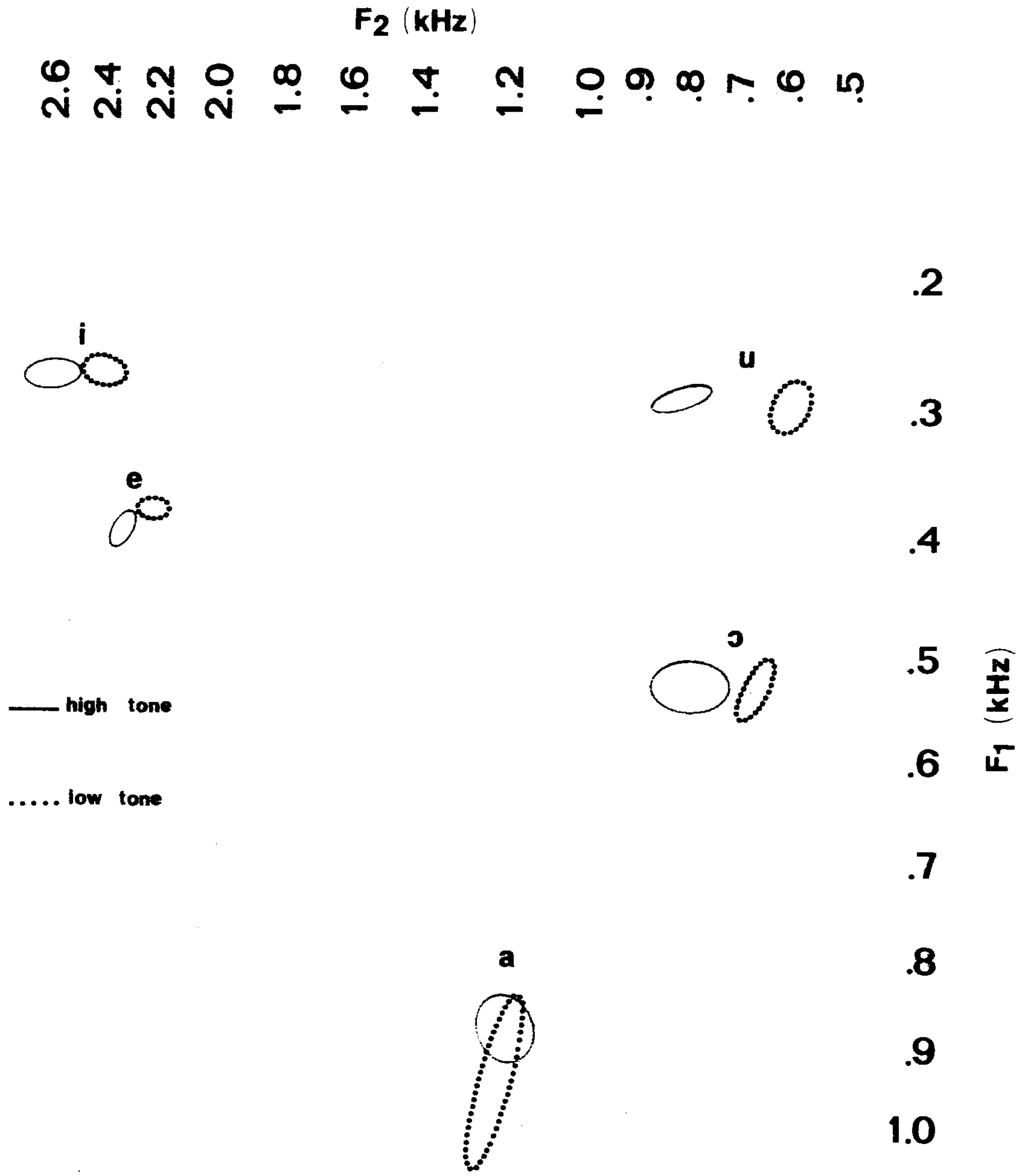


Figure 2. Vowel ellipses associated with the high and the low tones for Speaker 1

LO

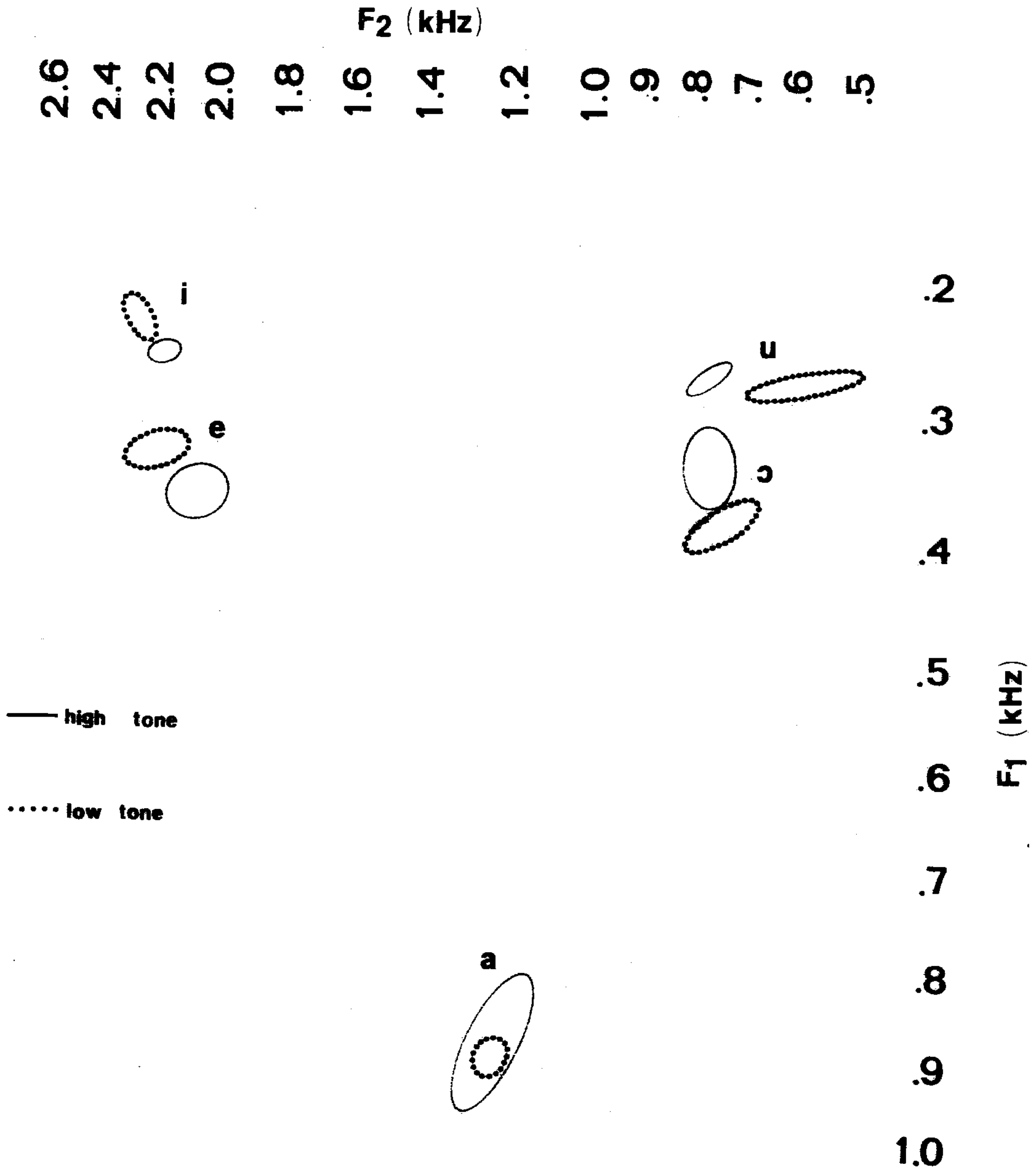


Figure 3. Vowel ellipses associated with the high and the low tones for Speaker 2

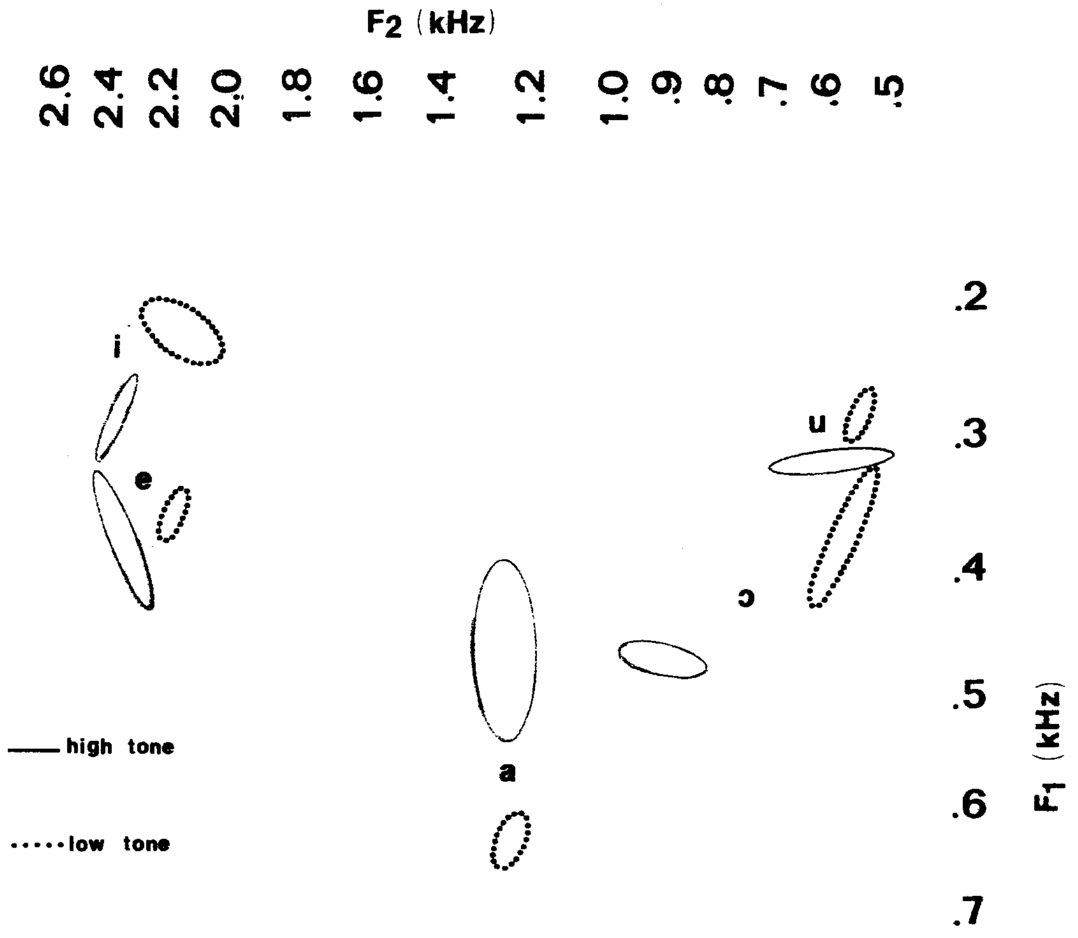


Figure 4. Vowel ellipses associated with the high and the low tones for Speaker 3

TSA1

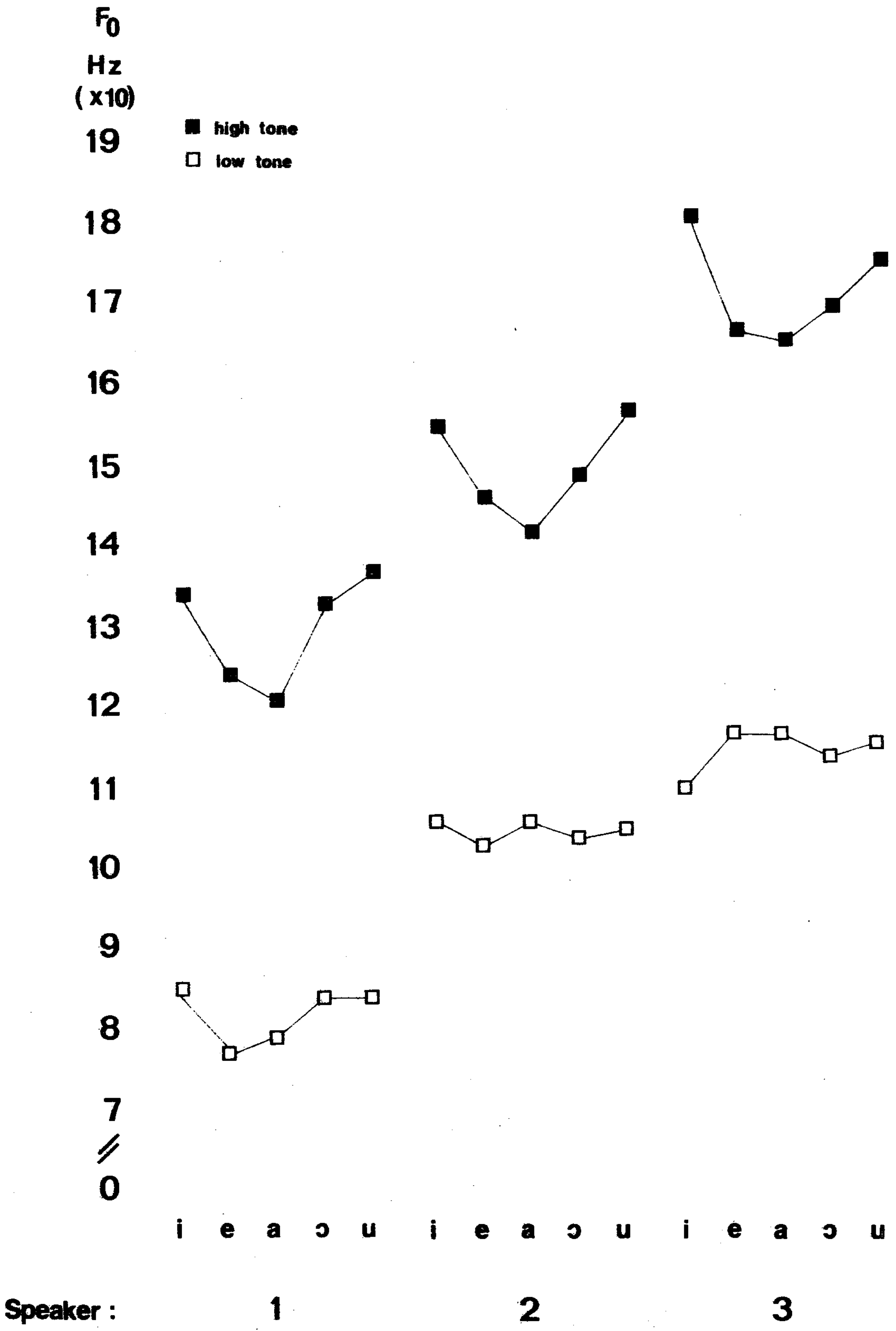


Figure 5. Average F_0 values of each tone for each vowel for Speakers 1, 2 and 3

Table IV. Level of significance (two-tailed test) for Speakers 1, 2 and 3

Speaker	Vowel	F1 (High/Low)	F2 (High/Low)	F3 (High/Low)
1	[i]	n.s.	0.01 (H>L)	0.01 (H<L)
	[e]	0.001 (H>L)	0.001 (H>L)	n.s.
	[a]	0.05 (H<L)	n.s.	0.001 (H<L)
	[ɔ]	n.s.	0.001 (H>L)	0.10 (H<L)
	[u]	n.s.	0.001 (H>L)	0.001 (H>L)
2	[i]	0.001 (H>L)	0.01 (H<L)	n.s.
	[e]	0.001 (H>L)	0.01 (H<L)	n.s.
	[a]	n.s.	n.s.	0.01 (H<L)
	[ɔ]	0.01 (H>L)	n.s.	n.s.
	[u]	0.10 (H<L)	0.001 (H>L)	n.s.
3	[i]	0.001 (H>L)	0.001 (H>L)	n.s.
	[e]	n.s.	0.001 (H>L)	n.s.
	[a]	0.001 (H>L)	n.s.	0.001 (H>L)
	[ɔ]	0.001 (H>L)	0.001 (H>L)	n.s.
	[u]	0.001 (H>L)	n.s.	0.01 (H<L)

(H: High tone '>': greater
L: Low tone '<': smaller n.s.: nonsignificant)

dard deviations (Davis, 1977). We can see in Figures 2, 3 and 4 for all three speakers the vowel ellipses associated with different tones do not overlap, except for vowel [a] for Speakers 2 and 3; that is, vowels [i], [e], [ɔ] and [u] associated with the high or with the low tone occupy distinct areas in the acoustic vowel space for all three speakers.

Some observations in terms of the relationship between vowel formant frequencies and tones across speakers can be made as follows:

1. For Speakers 1 and 2, there are significant differences in the same direction in F1 at the 0.001 level for vowel [e]; for Speakers 1 and 3, there are significant differences, but in the opposite directions, in F1 at the 0.05 or 0.001 level for vowel [a].

2. For Speakers 2 and 3, there are significant differences in the same direction in F1 at the 0.10 or 0.001 level for vowels [i] and [ɔ]; for the same speakers, there are significant differences, but in the opposite directions, in F1 at the 0.01 or 0.001 level for vowel [u].

3. Considering all three speakers, for the vowels [i] and [e], there are significant differences in F2 at the 0.01 or 0.001 level, but in each case for one speaker the differences are in the opposite directions to the other

two.

4. For all three speakers, there are no significant differences in F2 for vowel [a]. There are significant differences in F3 at the 0.01 level, but again they are not in the same direction for all three speakers.

5. There are significant differences in the same direction in F2 at the 0.001 level for vowel [u] for Speakers 1 and 2, and for vowel [ɔ] for Speakers 1 and 3, also in the same direction.

6. There are significant differences in the opposite directions in F3 at the 0.01 or 0.001 level for vowel [u] for Speakers 1 and 3.

7. These results show that vowels are in fact affected by tonal difference, although not in a systematic fashion for different speakers.

Figure 5 displays the average fundamental frequency values of the high tone and of the low tone for the five vowels produced by Speakers 1, 2 and 3. The filled squares represent the fundamental frequency values of the high tone and the open squares represent the fundamental frequency values of the low tone. Each square is the average of five tokens. The fundamental frequency value for each token is the average of the values of every 10 msec intervals. The value for each individual token is listed in Tables I, II and III. As far as the high tone is concerned, the values of the fundamental frequency correlate with vowel height. As for the low tone, the values of fundamental frequency and vowel height do not exhibit such a correlation for all three speakers.

4. DISCUSSION

It is nicely attested that the production of tone involves larynx movement, that a tone of higher fundamental frequency requires a higher larynx position (for a comprehensive review of the literature, see Ohala, 1973), and that a raised larynx shortens the length of a speaker's vocal tract from the glottis to the lips. Since the formant frequencies are to a first approximation inversely proportional to the length of a speaker's vocal tract (Fant, 1973), the acoustic consequences of different larynx height should be in general to increase or decrease all formant frequencies. Thus, by lowering the larynx in a simulated articulatory model, Lindblom and Sundberg (1971a) showed that all formant frequencies were lowered. Similarly, Riordan (1977) demonstrated that in computer simulated vowels, i, u and y, there was a consistent increase of all formant frequencies as the larynx was raised. However, our results do not show any consistent increase in the formant frequencies for vowels associated with the high tone for all three speakers. Moreover, some formant frequencies for certain vowels associated with the low tone are greater than those associated with the high tone. For instances, F1 and F3 of vowel [a], F3 of vowel [i] and F3 of vowel [ɔ] for Speaker 1; F1 of vowel [u], F2 of vowels [i] and [e] and F3 of vowel [a] for Speaker 2; and F3 of vowel [u] for Speaker 3 (See Table IV). This seems to indicate that shifts in other articulators, besides the rise of larynx, have taken place in the production of a vowel of a higher fundamental frequency, and these shifts have contributed to the modification of the formant frequencies in addition to the modifications due to larynx movement. In fact, in an X-ray investigation, Parmenter et. al. (1933) showed that articulatory shifts accompanying an upward change in pitch of one octave involve not only the larynx alone, but also the pharynx width, the epiglottis,the hyoid bone, the tongue, the jaw and the lips. These articulatory shifts seem to be the result of interaction between the anatomical effect of the larynx movement and the articulatory configuration required for the production of the vowels. Anatomically, all major articulators are connected tissues and muscles. Larynx movement will certainly affect the positions of other articulators which in turn affect the formant frequencies. Vowel perception may be another contributing factor to the modification of formant frequencies. Ainsworth (1976), in his study of fundamental frequency as the second determinant of vowel quality, reported that compensation in F1 and F2 was needed to retain phonemic identity of vowels when F_0 was submitted to large changes (from 120 Hz to 240 or 260 Hz). Although the average fundamental frequency difference between the high tone and the low tone is approximately 50 Hz for all five vowels for all three speakers, compensation in formant frequencies might be expected to occur in order to preserve the phonetic quality of the vowels. As a result, as Parmenter et. al. also pointed out, when the pitch is raised or lowered changes may be made in the size and shape of the supraglottal cavities in order to maintain the best resonance for the overtones which characterize the quality of the vowel. Our case seems to be more complicated as the low tone in Taiwanese is in fact a low falling tone (Zee, 1976). The exact way in which falling tone affects the identification of vowels is yet to be explored. It is possible to determine under laboratory conditions both the anatomical effect on the positioning of the articulators due to the larynx movement, and the perceptual effect on vowel quality due to pitch changes. However, it is much less straightforward to determine the interplay between them that eventually yields the vocal tract shape and size used by a tone language speaker producing a vowel with different tones. The formant frequency pattern associated with the high tone differs from the formant frequency pattern associated with the low tone in a way that is unique for all three speakers. This may be partly due to the differences in the auditory system and in the vocal tract shape and size among the speakers. But it may also be due to the possibility that each speaker uses different articulatory control and perceptual processing to produce a high tone or a low tone for a vowel.

5. SUMMARY OF RESULTS AND CONCLUSION

1. The formant frequencies associated with the high tone are not nece-

ssarily greater than the formant frequencies associated with the low tone for all three speakers.

2. For all three speakers, vowels [i], [e], [ɔ] and [u] associated with the high tone occupy areas in the acoustic vowel space that are distinct from those occupied by vowels associated with the low tone.

3. The way the formant frequency pattern associated with the high tone differs from the formant frequency pattern associated with the low tone is unique for all three speakers.

4. The value of the fundamental frequency of the high tone correlates with vowel height for all three speakers. No such correlation is found in vowels associated with the low tone for all three speakers.

5. Conclusion: vowels are in fact affected by tonal difference, although not in a systematic fashion for different vowels and for different speakers.

Acknowledgements

The assistance of the following people is gratefully acknowledged: Peter Ladefoged, Jean-Marie Hombert, Steve Greenberg, Ian Maddieson, Lloyd Rice, Jim Wright, and Lo I-Toh.

REFERENCES

Ainsworth, W. A. (1975). Intrinsic and extrinsic factors in vowel judgements. In *Auditory Analysis and Perception of Speech* (Fant & Tatham, eds.), N.Y., Academic Press, pp. 103-111.

Davis, S.B. (1977). Computer evaluation of laryngeal pathology based on inverse filtering of speech. SCRL Monograph #13.

Ewan, W. G. and R. Krones (1972). Laryngeal height during VCV utterances: umbrometer measurements. Paper presented at the 2nd Califonria Linguitics Association Conference, Santa Cruz, California

Ewan, W. G. and R. Krones (1974). Measuring larynx movement using the thyroumbrometer. *J. of Phonetics*, 2:327-335.

Fant, G. (1973). *Speech Sound and Features*. Cambridge, Mass., MIT Press.

Ladefoged, P. (1967). *Three Areas of Experimental Phonetics*. London, Oxford Univ. Press.

Lindblom, B. E. F. and J. E. F. Sundberg (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *JASA*, 50: 1166-1179.

Ohala, J. (1973). The physiology of tone. In *Consonant Types and Tone* (L. Hyman, ed.), L.A., University of Southern California Press.

Parmenter, C. E., S. N. Trevino and C. A. Bevans (1933). The influence of a change in pitch on the articulation of a vowel. *Language*, 9.1:72-81.

Riordan, C. J. (1977). Control of vocal tract length in speech. *JASA*, 62: 998-1002.

Sundberg, J. E. F. and P.-E. Nordström (1976). Raised and lowered larynx - the effect on vowel formant frequencies. *STL-QRSP*, 2-3: 35-39.

Vanderslice, R. (1967). Larynx vs. lungs: cricothyrometer data refuting some recent claims concerning intonation and "archetypality". UCLA WPP #7: 69-79.

Zee, E. (1976). Duration and intensity as correlates of F_0. Paper presented at the 92nd ASA meeting, Autumn, 1976, San Diego, California.

Working Group on speech Errors: Slips of the Tongue, Ear, Pen, and Hand.

Victoria A. Fromkin

[XIIth International Congress of Linguistics, Vienna,
Austria, Sept. 1-2, 1978]

In 1895, the first major psycholinguistic analysis of speech errors, together with a corpus of over 8,000 such errors was published in Vienna. (Meringer, R. and K. Mayer, 1895) This was closely followed by the publication of Sigmund Freud's (1901) now classic psychological treatment of speech errors. The research reported on at the Working Group held at the XIIth International Congress of Linguists therefore had its origins in the city of Vienna and it is altogether fitting that the first meeting of scholars concerned with speech production and perception errors as the data base for the construction and testing of theories of language and language behavior should have been held there. During the 82 years that followed Meringer's book until the meeting of this Working Group a number of papers using speech errors as evidence for models of language and language use appeared. But it is only in the last ten years that speech error collection and analysis has involved a large number of research psychologists and linguists. The fact that the organizing committee of the XIIth International Congress of Linguists included this Working Group in the Congress Program attests to the fact that speech error research has finally 'come of age'. The Working Group was dedicated to the memory of Meringer and it now seems clear, although it did not at the time his books on speech errors were published, that he will have a lasting place as a contributor to our understanding of the nature of human language.

The Working Group on Speech Errors was organized and chaired by Victoria Fromkin. It met for one and a half days during the XIIth Congress. Prior to the Vienna meeting twenty-seven international scholars distributed to each other copies of their papers concerned with this area of research, papers which were discussed rather than read at the Working Group sessions. Twenty-three of the authors of these papers were present at the Congress and participated in the discussions. In addition many auditors attended the sessions.

In preparation for the discussion the active participants (see below for names) also received a list of questions which were to serve as the basis for the discussions held in Vienna. These questions and topics were as follows:

(1) An hisotircal view -- from Meringer and Freud (and Reverend Spooner) to Vienna, 1977.

(2) Why study speech errors?

(3) Speech errors and models of linguistic performance (production and/or perception models) -- what do the proposed models have in common? How do they differ?

(4) What do speech errors tell us about lexical organization, storage, and representation? About feature systems?

(5) What are the similarities and differences between slips of the tongue, ear, pen, and hands? (The slips of the hands refer to errors made by deaf sign language users.)

(6) What causes speech errors? What is the effect of 'non-linguistic' factors, e.g. environment, intrusive thoughts, 'repressed' ideas, in producing errors? What linguistic factors, e.g. phonological/phonetic similarity, semantic similarity, syntactic similarity, etc. of target and substitution, seem to produce errors more frequently than others? How can such factors and influences be accounted for in a model of performance?

(7) What are the similarities and differences between 'normal' and 'abnormal' errors?

(8) Methodology in collecting and analyzing errors? How to induce speech errors under controlled laboratory experimental conditions.

(9) Where do we go from here?

The questions all dealt with aspects of research covered at least to some extent in the papers which were distributed in advance. While no definitive answers to the questions were arrived at there was general agreement that speech error data provide, in many cases a better 'window into the mind' than does error-free speech. This is the goal of all the participants -- to use these errors as a means of constructing and testing linguistic and psycholinguistic hypotheses, rather than to study deviant speech and deviant comprehension per se. However, the participants differed in their specific goals. Those concerned with understanding error-free and deviant language use for clinical purposes were able to contribute to those primarily concerned with theoretical models and vice versa.

References

Freud, S. (1901) Zur Psychopathologie des Alltaglebens. Vienna.

Meringer, R. (1908) Aus dem Leben der Sprache. Berlin.

Meringer, R. and K. Mayer (1895) Verspechen und Verlesen, eine psychologisch linguistische Studie. Vienna.