Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations

Arthur S. House, and Edward P. Neuburg

Citation: The Journal of the Acoustical Society of America **62**, 708 (1977); doi: 10.1121/1.381582
View online: https://doi.org/10.1121/1.381582
View Table of Contents: http://asa.scitation.org/toc/jas/62/3
Published by the Acoustical Society of America

Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations

Arthur S. House[a)]

Institute for Defense Analyses, Princeton, New Jersey 08540

Edward P. Neuburg

Department of Defense, 9800 Savage Road, Fort Meade, Maryland 20755
(Received 30 August 1976; revised 1 May 1977)

A procedure is described for possible use in automatic identification of the language being spoken. The method is to assume that the gross linguistic classes of a language are characterized by certain statistical constraints, namely, they are probabilistic functions of a Markov chain; and that these gross classes can be identified automatically with usable accuracy. The present discussion is in the form of a feasibility experiment and deals only with artificial data (data not derived from acoustic signals). The results indicate that statistical models, as described, can discriminate amongst languages.

PACS numbers: 43.70.Gr, 43.70.Qa, 43.70.Sc, 43.70.Ve

INTRODUCTION

We believe that the language of a given utterance can be identified automatically—that is, by analytic procedures without the intervention of a human listener—on the basis of current knowledge of (1) the structure of natural language, (2) statistical procedures, and (3) the availability of large computers. We suggest that the procedure can be divided, conceptually at least, into four parts as follows: (1) a mathematical model of a spoken language, having some set P of parameters, is chosen; (2) for each language L of interest, actual parameter values $P(L)$ are assigned to the model; (3) for testing utterances, another set of parameters Q is chosen, and for an incoming utterance U actual parameter values $Q(U)$ are calculated; and (4) a function $F(P, Q)$ is chosen, and $F[P(U), Q(L)]$ is evaluated for each L so that the L which optimizes F may be selected. With minor modification of terminology this schema applies to a number of related tasks, such as talker identification, word verification, etc.

It seems clear that there are a number of philosophies underlying language-recognition schemes. At one end of the continuum, speech is modeled by parameters whose values are easy to estimate from samples of speech (Hanley *et al.*, 1966; Atkinson, 1968; Zalewski and Majewski, 1971). The parameters may be various measures of energy, of fundamental frequency contours, linear predictive coefficients, etc. At the other end of the continuum, speech is modeled as a sequence of linguistic elements (Schwartz and Makhoul, 1975). Work done anywhere along this continuum must seek independence from talker characteristics, channel characteristics, noise, etc. It seems evident, to us at least, that we must avoid both ends of this continuum—work aimed directly at the acoustic characteristics of the speech signal has not succeeded in isolating the variability attributable to talker, speech contexts, and channel noise, for example, and work aimed at extracting phonemic categories is far from satisfactory. It should not surprise anyone when we say we have sought a sensible middle ground.

It is clear also that, given a phonetic representation of a script, procedures for modeling and identifying languages can be formulated based on the set of phonemes in the character strings, the first-order statistics of the characters, various higher-order statistical measures taken over the characters, etc. (Kucera and Monroe, 1968; Hultzén *et al.*, 1964; Denes, 1963; Menzerath, 1950). These statistical approaches to the problem have not been appealing for at least two rather obvious reasons: (1) Automatic identification of the phones or phonemes of an utterance with high accuracy under conditions approaching those of practical interest is beyond the current state of the art in automatic speech recognition, as stated above, and, (2) the statistical procedures generally require a great deal of computation.

Our choice of a middle ground is based on a belief that (1) it will prove possible to identify *gross* linguistic categories with a precision great enough for our procedures, and that (2) the sequence of gross categories in languages is sufficient to characterize them for identification. [The question of precision will not be treated here, but will make up the major part of a later paper; some preliminary results have been reported earlier (House *et al.*, 1975).] The characterization of languages by categories grosser than phonemes is, in a strong sense, a description of the syllabic structure of a language in terms of consonant categories and nuclei (O'Connor and Trim, 1953). The important consideration is to recognize that such descriptions have two aspects, the linear consonant/vowel assignments and the privileges of occurrence of particular classes during concatenation of consonants. We propose to face up to the second problem—that of computation—directly, and will make use of statistical procedures and computer facilities available to us; fortunately they are adequate to the tasks we will undertake.

In this note we will not deal with the actual analysis and categorization procedures, but will describe a feasibility study to test our model. The speech elements, therefore, will not be extracted from the speech wave form, but from a presumably correct sequence of finer categories, that is, a broad phonetic transcription.

I. THE MODEL

(1) A speech utterance U consists of a succession of symbols chosen from a set whose elements are

A = stop consonant,

B = fricative consonant,

C = nonvocalic sonorant,

D = vowel,

and

E = silence.

(2) The elements of the set are produced as a probabilistic function of a Markov chain. A k-state model is specified completely by a k-by-k transition matrix

$$\{a_{ij}\},$$

where a_{ij} is the probability that the chain will transit to state j, given that presently it is in state i; and by k probability distributions

$$\{b_{js}\},$$

where b_{js} is the probability that the symbol s will be uttered if the underlying chain is in state j.

We note that the string of output symbols need not be a Markov chain of any order, so that the speech itself is not assumed to be Markovian. (Later we will discuss a model in which speech is Markovian.)

II. THE PARAMETERS OF THE MODEL

The selection of language-dependent parameters for the model will be accomplished by means of a statistical procedure. First we obtain a string of symbols (see the description of materials below) long enough, hopefully, to reflect the statistical properties of a language. Given an utterance

$$U = s(1), s(2), \ldots, s(n),$$

and a set of actual parameter values,

$$V = (\{a_{ij}\}, \{b_{js}\}),$$

we can calculate the probability $P\{U \mid V\}$. The set V of parameter values that maximize this probability is the so-called maximum-likelihood set. If the properties of the utterance in language L are typical of that language, then the values V are the set of parameters $P(L)$ we are seeking. There are a number of machine algorithms for arriving at the maximum-likelihood values (Baum *et al.*, 1970; Davidon, 1968; Box, 1966; Fletcher and Powell, 1963). For our tests we have used the algorithm developed by Baum *et al.* (1970). All of these methods require substantial computing power, a burden we accept because we have the use of a large and fast computer. Furthermore, in theory at least, the hill-climbing procedure needs only be done once for a given language.

III. UTTERANCE PARAMETERS

Our ultimate aim is that the parameters $Q(U)$ of the utterance and $P(L)$ of the model will be symbols generated by an automatic segmentation procedure. For the feasibility study we are describing, however, we have reduced published phonetic transcriptions of various language materials into four-character (A, B, C, D) or five-character (A, B, C, D, E) texts. The silences E that were introduced corresponded to breath pauses and related stoppages that were reasonable for the textual materials. Materials consisted of a short fable given in a large number of languages, each in a broad phonetic transcription (Anon, 1949); tales from Indian cultures (Boas, 1911); and English phonetic readings (Abercrombie, 1964; Snydover and House, 1949). The parameters used for this experiment, therefore, probably are at least as accurate as the best automatically derived parameters we could expect to obtain.

IV. DISCRIMINATION

The function used to decide the language represented by an utterance (or in the present case, by a transcription) is the same function used in building the models. Given the sequence U, we calculate for each L the probability $P\{U \mid L\}$ and we say that U represents that language L for which this probability is a maximum. For computation we actually use the natural logarithm of this probability of utterance U against language L. If this logarithm is divided by the number of symbols in U, we obtain the "expected score per character," a measure we use when comparing texts of different lengths. In the discussion of results, below, the "expected score per character" will be referred to as the *score*.

We have used the same set of parameters in our putative model and in our analysis of utterances. It should be noted that this is not an essential part of the procedures under discussion; it merely simplifies the choice of a discrimination function.

V. RESULTS

As an initial step, eight phonetic texts of the same fable, each in a different language, were reduced to four-character alphabets and these samples were used to form two-state, three-state, four-state, and five-state statistical models of each language. The languages and the length of the fables after reduction were as follows: American English (354), Russian (554), Greek (451), Hindustani [Urdu] (428), Chinese [Peking] (467), Korean (399), Japanese (414), and Swahili (390). The selection of these languages for the initial work represents an attempt to include among our samples a diversity of modern European and non-European languages, as well as to include languages that differ considerably in syllabic structure, in terms of consonant–vowel shape and in privileges of occurrence of particular phonemes within the structure of a syllable.

Each generated model has associated with it a score expressed as the natural logarithm of the probability of occurrence. In Table I are shown the scores associated with the various models generated from the eight texts identified above. Study of these scores suggests that increasing the number of states in the underlying model from two to four (or five) increased the score significantly, except in the case of the Japanese text. The Japanese text was included in our sample particularly because of its simple syllable structure, which in most standard dialects allows only three types of syllables, $\#V\#$, $\#CVN\#$, and $\#CV\#$ [or more concisely,

TABLE I. Scores associated with various models of eight language samples. See text for definition of scores; minus signs are omitted from scores.

Samples used in modeling	Number of states in models			
	2	3	4	5
English	1.177	1.130	1.120	1.099
Russian	1.253	1.190	1.166	1.146
Greek	0.981	0.958	0.946	0.928
Hindustani	1.129	1.098	1.093	1.070
Pekingese	1.215	1.082	1.032	1.042
Korean	1.034	0.972	0.937	0.912
Japanese	0.713	0.708	0.706	0.680
Swahili	0.927	0.894	0.858	0.829

#(C) $V(N)$#]. Since the last type of syllable is the most frequently occurring one in Japanese, the phonologic pattern in the language is predominantly an alternation of V and C (that is, vowel and consonant), making it particularly suited to a two-state model. In contrast to the score change of 0.007 in going from a two-state to a four-state model found for Japanese, American English showed a change of 0.057. It is interesting to note that the latter language is characterized by a more-complex set of permissible syllables [that is, #(C)(C)(C) V(C)(C)(C)(C)#] and may be better modeled with a number of states greater than two. The relation of the syllabic complexity of a language to the complexity of its statistical models, however, is a question that requires further study.

TABLE II. Scores obtained using various models of eight modified language samples. Minus signs are omitted from all scores. See discussion in text.

Scored sample	Sample used in generating models							
	E	*R*	*G*	*H*	*P*	*K*	*J*	*S*
				Two-state models				
English	1.177	1.200	1.256	1.208	1.393	1.223	1.655	1.342
Russian	1.278	1.253	1.422	1.334	1.447	1.362	1.981	1.518
Greek	1.043	1.101	1.098	1.011	1.295	1.037	1.128	1.030
Hindustani	1.167	1.197	1.162	1.129	1.309	1.175	1.402	1.200
Pekingese	1.496	1.435	1.461	1.328	1.215	1.535	1.767	1.454
Korean	1.073	1.125	1.092	1.071	1.341	1.034	1.278	1.098
Japanese	0.933	1.034	0.811	0.861	1.227	0.832	0.713	0.772
Swahili	1.050	1.096	0.982	0.988	1.261	0.996	1.022	0.927
				Three-state models				
English	1.130	1.205	1.319	1.185	2.355	1.253	2.704	1.748
Russian	1.304	1.190	1.521	1.375	2.316	1.509	3.285	2.346
Greek	1.105	1.200	0.958	1.027	1.500	1.123	1.741	1.123
Hindustani	1.153	1.227	1.183	1.098	1.652	1.189	2.309	1.335
Pekingese	1.455	1.499	1.543	1.322	1.082	1.500	2.870	1.768
Korean	1.064	1.214	1.111	1.046	1.855	0.972	1.984	1.220
Japanese	1.035	1.209	0.827	0.919	1.273	0.897	0.708	0.804
Swahili	1.092	1.236	0.997	0.993	1.312	0.994	1.418	0.894
				Four-state models				
English	1.120	1.220	1.326	1.196	2.369	1.622	2.695	1.922
Russian	1.316	1.166	1.643	1.390	2.486	2.346	3.286	2.377
Greek	1.181	1.200	0.946	1.027	1.564	1.134	1.737	1.559
Hindustani	1.258	1.249	1.193	1.093	1.677	1.261	2.300	1.723
Pekingese	2.078	1.503	1.533	1.374	1.032	1.805	2.860	2.334
Korean	1.077	1.242	1.122	1.041	1.841	0.937	1.971	1.622
Japanese	1.202	1.184	0.858	0.926	1.314	0.900	0.706	0.998
Swahili	1.178	1.179	1.089	1.007	1.321	0.996	1.412	0.858
				Five-state models				
English	1.099	1.220	1.735	1.223	1.865	1.669	2.785	1.960
Russian	1.402	1.146	2.282	1.646	2.666	2.545	3.410	2.674
Greek	1.234	1.225	0.928	1.032	1.805	1.511	1.853	1.321
Hindustani	1.369	1.255	1.434	1.070	1.665	1.426	2.404	1.643
Pekingese	2.410	1.526	2.028	1.300	1.042	1.759	3.328	2.433
Korean	1.085	1.250	1.368	1.097	1.309	0.912	2.011	1.464
Japanese	1.332	1.206	0.845	0.909	1.130	0.905	0.680	0.941
Swahili	1.401	1.178	1.021	0.991	1.172	1.042	1.554	0.829

TABLE III. Analysis of scores in Table II showing range (R) of scores obtained for each model when tested against all samples, and the difference (D) between the score for the appropriate model and the score for the closest competitor.

Sample used in modeling		Number of states in models			
		2	3	4	5
English	R	0.478	1.574	1.575	1.686
	D	0.023	0.055	0.076	0.121
Russian	R	0.728	2.095	2.210	2.264
	D	0.025	0.114	0.150	0.250
Greek	R	0.314	0.783	0.791	0.925
	D	0.030	0.069	0.081	0.104
Hindustani	R	0.273	1.211	1.207	1.334
	D	0.033	0.055	0.100	0.185
Pekingese	R	0.552	1.788	1.828	2.286
	D	0.113	0.240	0.342	0.258
Korean	R	0.307	1.012	1.034	1.099
	D	0.037	0.074	0.104	0.173
Japanese	R	0.514	0.565	0.608	0.652
	D	0.059	0.096	0.152	0.225
Swahili	R	0.334	0.524	0.554	0.725
	D	0.055	0.099	0.138	0.162

The lengths of the texts used in this portion of our investigations were estimated to be only marginally adequate for generating models, and, consequently, were used in their entirety, leaving no succeeding text against which the model for a given language could be tested. Each text, therefore, was tested against every model (including its own) in an attempt to establish the power of the procedures to distinguish among languages. This method of evaluation is open to criticism, of course, since testing against a model made from the text being tested will give an optimal answer, but if no other text (that is, language) scores better the results are encouraging, even if not conclusive.

The results of the tests are given in Table II. The tabulated data indicate that when one of our (reduced) texts is tested against a set of models derived from (reduced) texts, the best score is always produced by the model derived from the text in question. Furthermore, as the number of states in the models increases, there is a general tendency for (1) the spread of scores to increase, and (2) the difference between the score for the "correct" language and its nearest competing score to increase. These tendencies are shown in a more compact form in Table III, where it can be seen, for example, that when these eight texts are scored against a four-state model of English, the closest competitor obtained a score that is 0.076 worse than that of the English text itself. This can be interpreted as meaning that we expect n characters of an English text to score $\exp(0.076\times n)$ better against the English model than against any of the other four-state models tested—that is, the probability that 100 characters from the English sample are, in fact, English is about 2000 times as

TABLE IV. Scores obtained against independently generated four-state models. For each sample, the models were generated using the number of characters shown in the column headings, while the tested samples consisted of the succeeding number of characters indicated in the row headings. Minus signs are omitted from scores.

	Language used in model				
Sample tested	Hupa (900)	Haida (800)	Kwakiutl (450)	Fox (500)	Teton (500)
Hupa (600)	1.145	1.378	1.709	1.923	1.624
Haida (700)	1.248	1.119	1.500	1.871	1.511
Kwakiutl (500)	1.166	1.436	1.111	1.402	1.622
Fox (500)	1.121	1.183	1.091	0.869	0.928
Teton (480)	1.163	1.208	1.154	1.315	0.941

TABLE VI. Scores obtained for various language samples, using four-state models, as indicated. The number of characters in each text is shown, parenthetically. Minus signs are omitted from scores.

Scored sample language	Sample used in generating models						
	E	*R*	*F*	*S*	*G*	*C*	*J*
English (354)	1.120	1.220	1.241	1.331	1.317	2.269	2.695
Russian (554)	1.316	1.166	1.356	1.817	1.506	2.486	3.286
French (359)	1.216	1.178	1.035	1.047	1.316	1.681	1.826
Spanish (426)	1.180	1.218	1.167	1.061	1.296	1.871	2.450
German (521)	1.145	1.239	1.290	1.235	0.970	1.791	2.855
Chinese (467)	2.078	1.503	1.462	1.611	2.032	1.032	2.860
Japanese (414)	1.202	1.184	0.882	0.965	1.245	1.314	0.706

great as the probability that they are Hindustani, the nearest competitor.

These results raised a number of questions: Can similar results be obtained with a less-powerful statistical method—digraphic modeling and scoring, for example—and will a more complicated model improve our results? Since the original group of language samples was selected to contain diversity of structure and location, will similar results obtain if more-closely related languages are tested? Will these procedures be successful when the text being scored is different from the text used to construct the model? How does the significance of the score depend on sample size? Each of these questions was considered, in part at least, and answers to them are attempted below.

VI. INDEPENDENT TEXTS FOR MODELING AND SCORING

A number of texts of American Indian languages were generated specifically with this question, and similar issues, in mind. They were the most available transcribed extended texts available to us; they suffer from a rather nonstandard phonetic representation, but since our reduction procedures aim at only four gross classes, the reduced texts are reasonably accurate representations. The languages and the number of characters in each text were as follows: Hupa (2043), Haida (1593), Kwakiutl (962), Fox (Algonquian) (1020), and Teton (Siouan) (993).

Consequently, the initial portion of each American Indian four-character text was used to generate a four-state model and different portions of the texts were scored against these models. The results of these tests are shown in Table IV, where the number of characters used in modeling and in scoring is indicated for each language. These data show clearly that, given the string of characters in four categories as input, the procedures differentiated successfully among the five texts (that is, languages).

TABLE V. Scores obtained by testing various portions of an English text against one model. The first-order, five-state model used in the testing was derived from the first 500 characters of the text. Minus signs are omitted from scores.

Starting point in text	Size of sample tested					
	100	200	300	400	500	600
1	1.177	1.182	1.198	1.213	1.194	1.215
550	1.340	1.272	1.296	1.256	1.250	1.240
1150	1.173	1.182	1.181	1.169	1.183	1.200
1750	1.292	1.295	1.277	1.263	1.272	1.267
2350	1.250	1.264	1.268	1.253	1.252	1.261
2950	1.240	1.336	1.344	1.347	1.352	1.332
3550	1.236	1.223	1.244	1.228	1.237	1.283
4150	1.192	1.200	1.234	1.236	1.231	1.241
4750	1.337	1.300	1.299	1.285	1.279	1.298
5350	1.304	1.281	1.264	1.283	1.279	1.278
Mean	1.254	1.253	1.260	1.253	1.253	1.261
S.D.	0.059	0.051	0.046	0.045	0.045	0.037

VII. SCORE STABILITY AND SAMPLE SIZE

The question of score significance, of course, must be answered, but it is beyond the scope of this feasibility study. We have initiated some work leading toward an answer by making a series of tests on a relatively lengthy sample of English. The sample was constructed from a number of published phonetic texts, totaling about 6000 characters (Abercrombie, 1964; Snydover and House, 1949). Before we put the sample together we tested the texts against English models, and our procedures provided no evidence that these texts were from different languages. Consequently the whole sample was reduced to a five-character alphabet, the first 500 characters of which were used to form a model. Nonoverlapping sequences of various lengths were then tested against the model; the results are summarized in Table V.

The tabulated scores show that as the size of the tested sample increases the effect of its location decreases. In general these scores indicate, for English at least, that a test sample of about 300 characters may

TABLE VII. Scores obtained by testing four-state digraphic models. The number in parentheses is the difference between the score for the winning model and the closest competitor. Minus signs are omitted from scores.

Scored sample	Sample used in generating models							
	E	*R*	*G*	*H*	*P*	*K*	*J*	*S*
English	1.142 (0.107)	1.249	1.322	1.343	2.284	1.776	3.222	2.014
Russian	1.349	1.216 (0.133)	1.468	1.561	2.333	2.629	3.773	2.745
Greek	1.067	1.111	0.960 (0.107)	1.168	1.452	1.588	1.867	1.541
Hindustani	1.163	1.209	1.219	1.091 (0.072)	1.609	1.668	2.488	1.739
Pekingese	1.491	1.460	1.555	1.576	1.085 (0.375)	1.747	2.797	1.699
Korean	1.087	1.199	1.208	1.204	1.731	0.948 (0.139)	1.951	1.581
Japanese	0.957	1.036	0.832	0.876	1.177	0.868	0.706 (0.126)	0.865
Swahili	1.044	1.129	0.992	1.011	1.243	1.008	1.473	0.877 (0.115)

be necessary to ensure a representative sample. It seems also that increasing sample size much beyond 300 will not add significantly to the power of the method.

VIII. RELATED LANGUAGES

Transcribed versions of the original fable in French, Spanish (Castilian), and German were obtained and converted into four-character texts. Statistical models were generated for these new texts and the texts, along with those of American English, Russian, Chinese, and Japanese, were scored as before against these seven models. In this group of languages, five are modern European languages—some with close relatives, (viz., English and German, French and Castilian) and two are unrelated Oriental languages. The results are shown in Table VI, and suggest that the procedures are capable of identifying languages on the basis of "syllabic structure," even when the languages under test are closely related, although there is some indication that larger samples of text may be needed for accuracy.

IX. DIGRAPHIC MODELING

A digraphic four-state model was computed for each of the eight texts in question and scores were obtained as before. That is, each text was assumed to be a Markov chain (in contrast to our assumption that a text is a function of a Markov chain). The results are presented in Table VII, along with the differences between the scores for the appropriate models and their closest competitors. These entries can be compared to the four-state results in Table II and the differences in Table III. In general, these data suggest that digraphic modeling and scoring procedures may be adequate, at least when based on manually generated, highly accurate materials. However, we feel that our more general model potentially is more useful. In experiments on language structure using larger numbers of output categories, it has given better results, both in the accuracy and the linguistic significance of the models (Neuburg, 1971). More work is needed to resolve this issue, particularly for practical automatic uses.

X. CONCLUSIONS

When a stream of connected text is labeled discretely as a sequence of characters representing four gross phonetic categories, the procedures we have described appear to be capable of differentiating among languages. These exercises have been done primarily to establish the feasibility of such a decision procedure, given a reasonably accurate sequence. These successes suggest strongly that the problem of providing such an input sequence automatically, as well as the influence on the decision procedure of errors in the input sequence, is worth investigating. Furthermore, since it seems reasonable to expect that automatic extraction procedures can provide the gross characterizations used in our decision processes, a successful automatic language-identification scheme may be feasible.

APPENDIX: A TYPICAL HIDDEN MARKOV MODEL FOR REDUCED PHONETIC TEXT

Source: English

Order: 1

States: 5

Parameters: 40

Text length: 500

Initial portion of text:

DBDBDCDADCACDADBCABDCAC
CDCDBADBDADAADCBDACDAC...

Score for text: −596.9

Score per character: −1.194

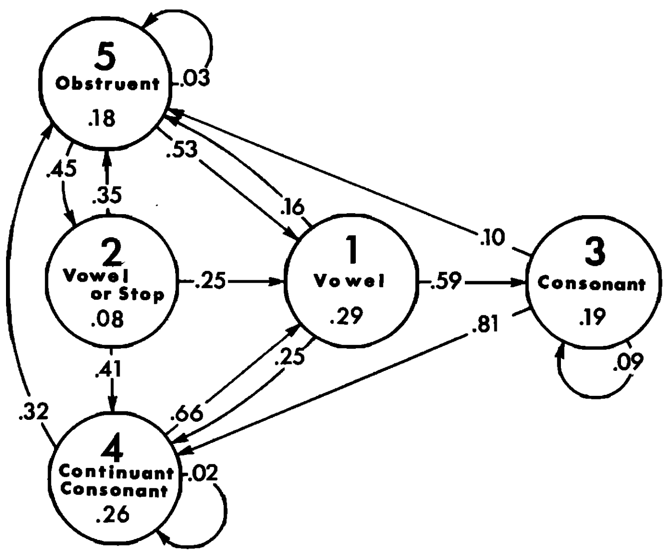

FIG. 1. Diagrammatic representation of the transition matrix for a model of English. The states are labeled with the numerical designations used in the transition matrix given in the Appendix. The stationary probability associated with each state is shown within its circle, along with linguistic labels descriptive of the categories assigned to it; the linguistic labels are not necessarily mutually exclusive. The output lines from each state are labeled with the probability of transiting to another state. The states are arranged arbitrarily to suggest the syllabic structure of English.

$\{a_{ij}\}$ Transition matrix

	1	2	3	4	5
1	0.00000	0.00000	0.58664	0.25458	0.15879
2	0.24662	0.00028	0.00000	0.40530	0.34780
3	0.00038	0.00002	0.08790	0.81428	0.09742
4	0.66461	0.00000	0.00000	0.01577	0.31962
5	0.52697	0.44753	0.00000	0.00000	0.02550

P_i Stationary probabilities

1	2	3	4	5
0.28966	0.08092	0.18630	0.26237	0.18076

$\{b_{js}\}$ Output probabilities

	A	*B*	*C*	*D*	*E*
1	0.00000	0.00000	0.00000	1.00000	0.00000
2	0.14190	0.00000	0.00000	0.85810	0.00000
3	0.32809	0.19880	0.42079	0.05232	0.00000
4	0.52479	0.40640	0.00000	0.00006	0.06875
5	0.00000	0.26960	0.73040	0.00000	0.00000

a-Matrix entropy (uncertainty of next state given present state)

State	Entropy
1	1.3754
2	1.5594
3	0.8816
4	1.0121
5	1.1411

Weighted average: 1.1606

b-Matrix entropy (uncertainty of output symbol given present state)

State	Entropy
1	0.0000
2	0.5892
3	1.7390
4	1.2825
5	0.8409

Weighted average: 0.8601

Average entropies of state (x) and output (y):

$H(x)$	$H(y:x)$	$H(y)$	$H(x:y)$
1.1606	0.8601	1.7221	0.2986

[a]The order of authors is alphabetic.

Abercrombie, D. (1964). *English phonetic texts* (Faber and Faber, London).
Anon. (1949). *The principles of the international phonetic association* (London).
Atkinson, K. (1968). "Language identification from nonsegmental cues," J. Acoust. Soc. Am. **44**, 378(A).
Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat. **41**, 164–171.
Boas, F. (1911). *Handbook of American Indian languages* (Washington, D.C.).
Box, M. J. (1966). "A comparison of several current optimization methods and the use of transformations in constrained problems," Comput. J. **9**, 67–77.
Davidon, W. C. (1968). "Variance algorithm for minimization," Comput. J. **10**, 406–410.
Denes, P. B. (1963). "On the statistics of spoken English," J. Acoust. Soc. Am. **35**, 892–904.
Fletcher, R., and Powell, M. J. D. (1963). "A rapidly convergent descent method for minimization," Comput. J. **6**, 163–168.
Hanley, T., Snidecor, J., and Ringel, R. (1966). "Some acoustic differences among languages," Phonetica **14**, 97–107.
House, A. S., Neuburg, E. P., and Wohlford, R. E. (1975). "Preliminaries to the automatic recognition of speech: language identification," J. Acoust. Soc. Am. **57**, S34.
Hultzén, L. S., Allen, J. H D. Jr., and Miron, M. S. (1964). *Tables of transitional frequencies of English phonemes* (U. Illinois, Urbana).
Kucera, H., and Monroe, G. K. (1968). *A comparative quantitative phonology of Russian, Czech, and German* (American Elsevier, New York).
Menzerath, P. (1950). "Typology of languages," J. Acoust. Soc. Am. **22**, 698–701.
Neuburg, E. P. (1971). "Markov models for phonetic text," J. Acoust. Soc. Am. **50**, 116 (A).
O'Connor, J. D., and Trim, J. L. M. (1953). "Vowel, consonant and syllable—a phonological definition," Word **9**, 102–122.
Schwartz, R., and Makhoul, J. (1975). "Where the phonemes are: dealing with ambiguity in acoustic-phonetic recognition," IEEE Trans. Acoust., Speech Signal Process., **ASSP-23**, 50–53.
Snydover, A., and House, A. S. (1949). "Fulton Lewis, Jr. on housing projects," Q. J. Speech **35**, 227–228.
Zalewski, J., and Majewski, W. (1971). "Polish speech spectrum obtained from superposed samples and its comparison with spectra of other languages," Proc. Seventh I.C.A. (Budapest), **3**, 249–252, (Paper 24-C-22).