A Review of Research in Automatic Language Identification

Yeshwant K. Muthusamy

Oregon Graduate Institute
Department of Computer Science
and Engineering
19600 N.W. von Neumann Drive
Beaverton, OR 97006-1999 USA

Technical Report No. CS/E 92-009

May, 1992

A Review of Research in Automatic Language Identification

Yeshwant K. Muthusamy

Center for Spoken Language Understanding
Oregon Graduate Institute
19600 NW Von Neumann Drive
Beaverton, OR 97006-1999 USA

Technical Report No. CS/E 92-009

May 1992

1 Introduction

The problem of automatic language identification—identifying the language being spoken by an unknown talker from a short excerpt of speech—is a challenging and important one, of interest to linguists and computer speech researchers. This document reviews the studies done so far in this area.

2 Review

Given below are brief descriptions of the major studies in automatic language identification to date. Table 1 summarizes the salient features of these studies.

2.1 The Texas Instruments Effort

One of the most sustained efforts in automatic language identification was carried out between 1973 and 1980 at Texas Instruments (TI), documented in a series of four reports [LD74, LD75, LD78, Leo80].

The basic philosophy underlying the TI approach was that languages differ by the frequency of occurrence of certain reference sounds or sound sequences. The sounds or sound sequences characteristic of a language occur more often in that language than in any other language under consideration. Therefore, the likelihoods of the languages, given these sequences, could be computed and used to make decisions in reasonably short times.

Study 1. The first study [LD74] concentrated on single reference sounds. The data consisted of read text from 100 adult male speakers of 5 languages, referred to simply as L_1, L_2, L_3, L_4, and L_5. The training data consisted of 90-second segments of speech from each of 10 speakers of each of the five languages. The test data consisted of 90-second segments from: 10 speakers of L_1, L_3, and L_5; 6 speakers of L_2; and 14 speakers of L_4.

The first step in this approach was an automatic segmentation of the digitized speech based on a measure of dynamic spectral change called "transitionitivity". Reference files of sound segments potentially useful for language discrimination were automatically generated from the training data, using an "intersegment distance" measure. This technique allowed a segment to be added to a file only if it was sufficiently different from each segment already in the file. These reference files were then pruned to eliminate sounds that did not demonstrate sufficient language specificity. The frequency of occurrence of the remaining reference segments in the files was determined and the time average log-likelihood of the languages computed (i.e. for each language L and reference segment R, the probability or likelihood that language L was spoken, given that segment R has occurred). In one experiment, decision functions were computed for each pair of languages for each of the 50 test speakers. The decision strategy was to choose the language with the smallest negative average log likelihood. Pairwise identification accuracy of the 10 language pairs ranged from 60% to 100%. Overall accuracy was 64% with a nearest neighbor decision rule using the pairwise identification results. The identification decision was made using 60 seconds of speech.

Study 2. The second phase of the study [LD75] used the same data as above, but used *sequences* of several phoneme-like segments for classification. Another improvement was the use of "time-frequency scanning" to accept or reject hypothesized occurrences of component sound segments. Two measures were introduced to help prune the file of reference sequences: (i) an information-theoretic measure called "entropy threshold" that guided the selection

STUDY	LANGUAGES	TYPE OF DATA	SPEAKERS	APPROACH	RESULTS
Texas Instruments (1973-80)	7 (not specified)	Read Speech	100 adult males (50 train 50 test)	Detection of "Reference Sounds" and estimation of log likelihoods of the languages	62% (no rejection) 100% (68% rejection)
House and Neuberg (1977)	American English, Chinese, Greek, Korean, Urdu, Japanese, Russian and Swahili (8)	Phonetic trans. of text (no real speech)	-	HMMs trained on sequences of broad category labels	Near-perfect discrimination (no % specified)
Li and Edwards (1980)	2 Asian & 3 Indo-European (not specified)	Read Speech	150 (50 train 50 evaluate 50 test)	Segment-based and syllable-based Markov models	80%
Cimarusti and Ives (1982)	American English, Czech, Farsi, Korean, German, Mandarin, Russian & Vietnamese (8)	Read Speech	40 (train and test sets unspecified)	Acoustic features and a polynomial decision function	84%
Ives (1986)	American English, Czech, Farsi, Korean, German, Mandarin, Russian & Vietnamese (8)	Spoken Speech (100 to 5000 Hz)	122 adult males (train and test sets unspecified)	Expert System Production Rules	92%
Foil (1986)	3 (not specified)	Speech from radio (SNR 5 dB)	Not specified	Processing Pitch & Energy Contours Formant-clustering algorithm	39% 64 % (11% rejection)
Goodman et. al (1989)	Four different sets of languages (not specified)	Speech from radio (SNR 9 dB)	Not specified	Improved Formant Clustering Algorithm	Reduced Foil's error rate in half (no % specified)
Sugiyama (1991)	20 languages (CCITT SG-XII CD-ROM)	Spoken speech (Avg. SNR 49.2 dB)	76 Males 77 Females	Standard VQ VQ histogram Algorithm	65% 80%
Savic et. al (1991)	English, Hindi, Mandarin & Spanish	Read Speech (0 - 4.5 kHz)	Not specified	HMMs & Pitch Contour Analysis	Not specified
Muthusamy et. al (1992)	English, Japanese, Mandarin & Tamil	Conversational Speech (0 - 8 kHz)	40 Males 40 Females	Broad phonetic category segment-based features & neural networks	79.5% (5.7s of speech) 89.5% (17.1s of speech)

Table 1: Studies in Automatic Language Identification

of reference sequences with sufficient language specificity, and (ii) an acceptance level for hypothesized sequences, that rejected sequences that did not occur often enough to merit inclusion.

Classification was based on the summed logarithms of the language likelihood estimates, given the occurrences of the reference sound sequences. Experiments were performed with sound sequences of different lengths. It was found that sequences of length 4 performed best on the training data: 88% correct classification of the 5 languages with an entropy threshold of 2.3 and acceptance level of 12.5%. A decision rule using sequences of length 5 in combination with sequences of length 1 yielded 70% accuracy on the test data, with the same threshold values as above.

Study 3. The third study [LD78] used an interactive approach to the generation of reference sounds: manual selection of reference sounds was followed by automatic isolation of the representative occurrences of these sounds from the speech data. The isolated sounds were then manually verified before further processing. This approach produced achieved comparatively better results than the previous two studies.

Study 4. In the final study [Leo80], the interactive approach to reference sound generation was extended to allow more accuracy in specifying reference sounds and more flexibility in the allowed types of reference sounds. Another improvement was the introduction of a criterion for rejection, i.e. not classifying an utterance when the basis for such a decision is not sufficiently strong.

The data of the first three studies was augmented with speech from 17 speakers of language L_7 (L_6 was reserved for English) and 14 speakers of language L_8. There were now 66 speakers in the training set and 65 speakers in the test set.

The speech was digitized and the characteristic sound sequences determined using the improved interactive reference sound generation program. The initial reference file had 94 sounds from the 7 languages. The training data was processed to automatically detect and count occurrences of these sounds to compute parameters of a decision function. After applying pruning techniques based on various thresholds, a file of 80 reference sounds was produced. The decision function for a language was defined to be the negative of the sum of the log-likelihoods for all detected reference sounds. The test data was then processed to detect and count reference sounds to evaluate decision function values (one for each possible language). The language with the minimum decision function value was chosen. If the difference between the smallest and next smallest decision function value was below a certain threshold, the speaker was rejected (the rejection criterion).

With the 7-language 65-speaker test set consisting of 80 reference sounds, 62% accuracy was attained when no rejects were allowed, and 100% accuracy was achieved with a rejection rate of 68%. With the original 5-language 50-speaker test set consisting of 54 reference sounds, the corresponding figures were 72% and 100% with only a 56% reject rate.

While it is clear that significant contributions to the field of automatic language identification have been made by the TI effort, the extensibility of their general approach is open to question. Improved results were obtained in the latter two studies, in which automatic determination of reference sounds was replaced by an interactive process that required considerable human input. However, such manual determination of the reference sounds in the languages under consideration mandates the researchers' *a priori* knowledge of, and familiarity with the languages. This could severely limit addition of languages to the identification system. This weakness is apparent in the fourth study [Leo80] in which there is a degradation in performance (from 72% to 62%) with the addition of the two languages L_7 and L_8. The author attributes it to a lack of familiarity with the two languages resulting in

selection of inappropriate reference sounds for these languages. Also, the database consisted of adult male speakers only. It is not clear how the systems would perform with female or young speakers.

2.2 House and Neuberg

House and Neuberg [HN77] demonstrated the feasibility of using acoustic features derived from broad phonetic categories of speech to identify languages. They reasoned that, since accurate phoneme recognition is beyond the current state-of-the-art, the information provided by the broad phonetic categories (stop, fricative, vowel, silence) should be examined. They assumed that the sequence of broad phonetic categories of a language were produced by a Markov model, and that the parameters of the model could be estimated for a given language from sufficient training data.

The data for this study consisted of manually generated phonetic transcriptions of text from each of the following eight languages: American English, Chinese, Greek, Japanese, Korean, Russian, Swahili and Urdu.

Hidden Markov models were trained on sequences of broad phonetic category labels derived from these phonetic transcriptions, and perfect discrimination of the eight languages was obtained.

It should be noted that this study did not make use of real speech, only phonetic transcriptions of *text*. With actual speech, manual segmentation is not feasible for large amounts of data. Also, it is not clear whether the differences between the scores for the different languages are statistically significant.

2.3 Li and Edwards

The Markovian techniques suggested by House and Neuberg were further developed by Li and Edwards and applied to actual speech data. Their work [LE80] represents some of the earliest efforts to develop statistical inference techniques to discriminate among real languages. They used a broad segmentation scheme to classify data into six acoustic-phonetic classes:

1. syllabic nuclei (vowels and syllabic nasals, etc.)
2. non-vowel sonorants (nasals, liquids, semivowels, and voiced stops and fricatives in intervocalic environments)
3. vocal murmur (voice detection)
4. voiced frication (voiced sibilants, fricatives, etc.)
5. voiceless frication (voiceless sibilants, fricatives and aspiration of stops)
6. silence and low energy segments (plosive gaps, /f, h/ etc.)

Based on these broad segmental classes, two statistical models for automatic language identification were developed: one based on segments and one based on syllables. The segmental models were implemented as either zero, first or second order Markov models and characterized segmental sequences in the languages.

The syllable model was divided into two types, one based on inter-syllable-nuclei sequences and one based on intra-syllable-nucleus segment sequences. The inter-syllable zero-order Markov model described segment sequences between two syllabic nuclei, which can be roughly paraphrased as characterizing possible consonant clusters in the languages. The intra-syllable model represented a syllable as a nucleus preceded or followed by up to two

segments (not including a neighboring syllabic nucleus), and approximated the internal structure of a syllable without requiring detection of specific syllable boundaries. The intra-syllable model was implemented as both zero and first order Markov models.

The database consisted of read speech from 20 talkers of five languages, two Asian and three Indo-European. The two Asian languages were monosyllabic tonal languages with relatively simple consonant-vowel (CV) or CVC word structure. The three European languages represented two different language families, and were distinguished from the Asian languages by greater word length and more complex consonant clusters.

The training database consisted of 200 minutes of speech (four minutes each from ten talkers for each of five languages) collected in a reading mode, for a total of about 42,000 syllables and 150,000 segments. The test data was 100 minutes of read speech (two minutes each from ten talkers for five languages).

The identification procedure consisted of moving a variable length analysis window through the training data and the independent test data. The analysis window was x segments (for the segment based model) or y syllables (for the syllable based models) where x and y were varied to cover an analysis period from 15 seconds to two minutes long. Each model was tested over a selected analysis window with each language accumulating a conditional probability of being the language tested. For each window, an accumulated weighted vote was obtained for each language based on the conditional probabilities. The window was then incremented by one element (segment or syllable) and the process repeated with new weighted votes accumulated until the data was exhausted for each talker. The language associated with the largest analysis-window vote for that talker was chosen as the correct language.

The results of these techniques varied considerably across the various models, reaching a maximum of about 80% correct identification using the inter-syllable model for an independent test of 50 talkers (ten per five languages). An analysis of the confusions among languages indicates that the techniques distinguish the two major types of languages very well, that is, the Asian languages from the Indo-European languages. This suggests that a two-stage algorithm might be useful in language identification. The first stage divides the languages into major types, and the second stage examines the languages within each type in more detail and makes focused decisions based on known characteristics of that language type.

2.4 Cimarusti and Ives

Cimarusti and Ives [CI82] conducted a feasibility study of an approach to automatic language identification that was not based on linguistic units such as acoustic-phonetic segments or syllables. This approach applied pattern analysis techniques to acoustic features extracted from the speech signal.

The data consisted of three minutes of read speech collected from five adult male speakers for each of the following eight languages: American English, Czech, Farsi, German, Korean, Mandarin, Russian and Vietnamese. The data was randomly divided into training and test sets.

Using a 30 ms moving analysis frame with a 30 ms increment, 100 features derived from LPC analysis (including autocorrelation coefficients, cepstral coefficients, filter coefficients, log area ratios and formant frequencies) were extracted. There were equal number of feature vectors in the training and test sets.

A potential function was generated for all features in the training set. Using an iterative pattern analysis program, the complexity of a polynomial decision function was systematically increased until all the vectors in the training set were separated into the properly identified languages (100% classification accuracy). When this "tuned" decision function

was applied to the evaluation test set, the overall classification accuracy was 84%. The individual language classification scores ranged from 76.8% (American English) to 93.4% (Korean).

It is not clear whether all of the 100 features contributed to the classification performance. Issues such as feature selection, and removal of redundant features were not examined. The absence of an independent, "uncorrupted" test set, and the fact that only 5 speakers from each language were used makes it likely that the system is not truly speaker-independent.

2.5 Ives

Using an extended database for the same languages as the previous study, Ives [Ive86] developed an expert system for real-time automatic language identification. The goal of this effort was to develop a set of rules which would minimize the time required for classification.

The database used consisted of a total of 50 hours of speech from 122 male speakers from each of the following eight languages: American English, Czech, Farsi, German, Korean, Mandarin, Russian and Vietnamese. Exactly 720 five-second patterns were randomly chosen from each of the 8 languages for analysis. Thus, a total of 5760 patterns were used. The training and test set subdivisions were not mentioned.

The classification logic was based on 50 distinguishing features. An empirical threshold algorithm converted these subjective distinguishing features into objective numerical boundaries or thresholds. These thresholds were used to design a minimum set of nine production rules. Application of this rule set to the test data resulted in classification scores ranging from 84% (Russian) to 99% (Vietnamese). The overall accuracy was 92%.

The training and test sets use in this study are not specified. Also, the database had only male speakers. Performance of this system on female speakers is not known.

2.6 Foil

Foil [Foi86] was perhaps the first researcher to address speech recorded from radio under noisy conditions (the typical signal-to-noise ratio was 5 dB). He imposed an additional constraint that language recognition be made using less than 10 seconds of speech.

The data used consisted of 10 hours of speech from each of three unspecified languages, each from a different major language group. One of them was Slavic, and another was tonal south-east Asian. The identity of the third group was not revealed. The training set consisted of 6 hours of speech, the development set had 1.5 hours of speech and the final evaluation set had 2.5 hours of speech. The number of speakers was not specified.

Two techniques were explored, each designed to capture the language information in the speech signal.

The first technique was based on the premise that prosodic features, such as rhythm and intonation patterns which vary from language to language, could be the basis of a powerful language identification technique. In one configuration, a classical quadratic classifier was applied to seven prosodic features extracted from pitch and energy contours in the speech signal. The recognition accuracy on the final test set, using an average of 5 seconds of speech for the identification decision, was 39%. This is only slightly better than chance, given the 3-way choice between the languages.

A second technique was designed to exploit the frequency of occurrence of characteristic sounds of a language by using formant frequency values and locations to represent the sounds. In this configuration, a k-means clustering algorithm determined the 10 best formant vector clusters for each language, and a vector-quantization distortion measure was

used as the basis for language decisions. The recognition accuracy on the final test set, using an average signal duration of only 4.5 seconds, was 64%, with a rejection rate of 11%.

Considering the extremely noisy data used, the results of this study are impressive. The inclusion of a development test set, that was used to provide feedback for the algorithm development process, seems to have helped in "fine-tuning" the features used. The relatively high success rate can also be attributed to the large volume of data used in the experiments.

2.7 Goodman et al.

Goodman et al. [GMW89] enhanced Foil's formant extraction technique for language identification by modifying and adding parameters, improving the classifier and reducing its channel sensitivity. A new formant peak-picking algorithm was devised that performed well even with very noisy speech. The original formant vector was augmented with log amplitude values at the formant frequencies, and time difference terms measuring the formant transitions between significant phonetic events in the language. An improved voiced/unvoiced decision algorithm significantly reduced the number of false voicing errors. A k-means clustering algorithm similar to the one used by Foil was used to determine the 60 best formant-vector clusters for each language. The decision strategy was improved by the use of a weighted Euclidean distance measure instead of a Euclidean distance measure.

The data consisted of a large (9.6 hours), noisy (signal-to-noise ratio: 9 dB), database of six languages, with 2.92 hours of speech in the training set, 2.78 hours in the development set and 3.9 hours in the final test set. The final evaluation was done on a larger database of four different language sets, including this six-language set, the original three-language set used by Foil, and two other geographical subsets.

The recognition results were superior to the earlier algorithm in all four language sets (percentage values not mentioned). The error rate on Foil's original three-language set was reduced by more than 50%. A significant result was the insensitivity of the recognition accuracy to the signal-to-noise ratio, indicating the robustness of the formant peak-picking algorithm.

2.8 Sugiyama

Sugiyama [Sug91] proposed two language identification algorithms that were based on vector quantization and used acoustic features of the speech signal such as LPC coefficients, autocorrelation coefficients and Δ cepstral coefficients.

The data was taken from a multilingual speech database distributed by NTT, Japan [IIK90]. It consists of 16 sentences uttered twice by 4 male and 4 female speakers in each of 20 languages. The duration of each sentence is about 8 seconds. Both the training and test sets had approximately the same amount of data: around 21 minutes.

The first algorithm was based on standard vector quantization (VQ). Each language, k, is characterized by its own VQ codebook, V_k, generated using the training sentences. In the recognition stage, input speech is quantized by V_k and accumulated quantization distortion, d_k, is computed. The language with the minimum accumulated distortion is the recognized language. Several measures of spectral distortion were experimented with. The recognition accuracy was 65% using 64 seconds of unknown speech.

In the second technique, a universal codebook $U = \{u_j\}$, is generated using all training data. Each language k is characterized by its occurrence probability histogram h_k. During recognition, each input sentence is quantized by U and its occurrence probability histogram, $h(u_j)$, is computed. The language which has the minimum distance between h_k and h is the recognized language. A Euclidean distortion measure is used to determine histogram

separation. With this technique, the overall recognition accuracy was 80% using 64 seconds of unknown speech.

The results of this study are impressive, considering the large number of languages used. The individual language accuracies were not specified, so an analysis of the inter-language confusions is not possible.

2.9 Savic et al.

Savic et al. [SAG91] reported preliminary work on language identification using HMMs and pitch contours. The data consisted of 10 minutes of read speech in 4 languages: English, Hindi, Mandarin Chinese and Spanish. No quantitative results were specified.

2.10 Muthusamy et al.

Muthusamy et al. [MCG91, MC92] developed a 4-language high-quality speech language identification system using a combination of knowledge-based features and artificial neural networks. The fundamental assumption underlying this research was that each language has an unique acoustic signature that can be characterized by the acoustic, phonetic, and prosodic features of speech. Phonetic, or segmental features, include the the inventory of phonetic segments and their frequency of occurrence in speech. Prosodic information consists of the relative durations and amplitudes of sonorant (vowel-like) segments, their spacing in time, and patterns of pitch change within and across these segments.

The data for this research consisted of natural continuous speech recorded in a laboratory by 20 native speakers (10 male and 10 female) of American English, Mandarin Chinese, Japanese and Tamil. The speakers were each asked to speak a total of 20 utterances[1]: 15 conversational sentences of their choice, two questions of their choice, the days of the week, the months of the year and the numbers "0" through "10". The objective was to have a mix of unconstrained- and restricted-vocabulary speech. The average duration of the utterances was 5.4 seconds. The data was digitized at 16 kHz with 16-bit resolution. The training set consisted of 10 or 20 utterances from each of 14 speakers per language for a total of 930 utterances. The test set consisted of 10 or 20 utterances from each of 6 speakers per language for a total of 440 utterances.

The approach consisted of (a) neural network segmentation of the speech signal into seven broad phonetic categories: vowels, fricatives, stops, closures (silence and background noise), pre-vocalic sonorants, inter-vocalic sonorants, and post-vocalic sonorants; (b) design and computation of phonetic and prosodic features based on these broad category segments (e.g., inter-segment and intra-segment variation in pitch, duration statistics of the seven phonetic categories, frequency of occurrence of the seven categories, etc.), and (c) neural network classification of the languages using these features as input.

The segmentation phase consisted of a neural network that assigned 7 broad phonetic category scores to each 3 ms frame of the utterance. The frame-by-frame output of the network was converted into a time-aligned sequence of broad phonetic category labels using a Viterbi search with duration and bigram probabilities. The segmenter network was trained with 304 spectral and waveform features computed in the vicinity of each training frame. The segmentation output agreed with the hand-labels of the test utterances 85.1% of the time.

The language classification phase consisted of a second neural network that used features computed on the time-aligned broad phonetic category sequence to tell the languages apart. These features were designed to capture the phonetic and prosodic differences between the

[1] Five speakers in Japanese and one in Tamil provided only 10 utterances each.

languages. For example, an "inter- and intra-segment variation in pitch" feature was included specifically to distinguish Mandarin Chinese (a tonal language), from the other three languages which did not display such a large pitch variation within and across segments. Similarly, the presence of sequences of equal-length broad category segments in Japanese utterances led us to design an "inter-segment duration difference" feature. A total of 80 such phonetic and prosodic features were used.

On test utterances that were 5.7 seconds long on the average, the identification accuracy was 79.6%. When trained and tested on longer utterances (obtained by concatenating triples of the utterances in the training and test set), the performance rose to 89.5% on test utterances that were 17.1 seconds long on the average. Note that these performance figures were obtained by training and testing on both the fixed and free vocabulary utterances. The corresponding figures for testing on just the free vocabulary utterances were 79.5% and 88.5% respectively.

Despite the small number of speakers used, we are encouraged by the results obtained with this broad phonetic segment-based approach to automatic language identification.

3 Summary and Conclusions

There have been only about a dozen studies in automatic language identification over the past two decades. The data have spanned the range from phonetic transcriptions of text to telephone and radio speech. The number of languages has varied from three to twenty. The approaches to language identification have used "reference sounds" in each language, segment- and syllable-based Markov models, formant vectors, and acoustic, phonetic and prosodic features derived from broad phonetic categories. A variety of classification methods have been tried, including HMMs, expert systems, VQ, quadratic classifiers and artificial neural networks.

While the performance figures of each study might look impressive in isolation, meaningful comparisons across studies is virtually impossible, for the following reasons:

- Many of the studies represent classified or sensitive research, so experimental details (e.g., languages used, method of data collection) are often not described.
- There is no common, public-domain database (cf. TIMIT) with which to evaluate different approaches to automatic language identification.

We believe that basic research and the development of a public-domain, standardized, multi-language database are essential prerequisites to further advances in automatic language identification.

Acknowledgements

This research was supported in part by NSF grant no. IRI-9003110, a grant from Apple Computer, Inc., and by DARPA grant no. MDA 972-88-5-1004 to the Department of Computer Science & Engineering at the Oregon Graduate Institute. The author thanks Ronald Cole and Beatrice Oshika for their several useful comments and suggestions.

References

[CI82] D. Cimarusti and R. B. Ives. Development of an automatic identification system of spoken languages: Phase 1. In *Proceedings 1982 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Paris, France, May 1982.

[Foi86] J. T. Foil. Language identification using noisy speech. In *Proceedings 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, 1986.

[GMW89] F.J. Goodman, A.F. Martin, and R.E. Wohlford. Improved automatic language identification in noisy speech. In *Proceedings 1989 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, Scotland, May 1989.

[HN77] A. S. House and E. P. Neuberg. Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations. *Journal of the Acoustical Society of America*, 62(3):708–713, 1977.

[IIK90] H. Irii, K. Itoh, and N. Kitawaki. Multilingual speech database for evaluating quality of digitized speech. In *Proceedings 1990 International Conference on Spoken Language Processing*, pages 1025–1028, Kobe, Japan, 1990.

[Ive86] R. B. Ives. A minimal rule a.i. expert system for real-time classification of natural spoken languages. In *Proceedings 2nd Annual Artificial Intelligence and Advanced Computer Technology Conference*, Long Beach, CA, April-May 1986.

[LD74] R. G. Leonard and G. R. Doddington. Automatic language identification. Technical Report RADC-TR-74-200, Air Force Rome Air Development Center, August 1974.

[LD75] R. G. Leonard and G. R. Doddington. Automatic language identification. Technical Report RADC-TR-75-264, Air Force Rome Air Development Center, October 1975.

[LD78] R. G. Leonard and G. R. Doddington. Automatic language discrimination. Technical Report RADC-TR-78-5, Air Force Rome Air Development Center, January 1978.

[LE80] K.P. Li and T. J. Edwards. Statistical models for automatic language identification. In *Proceedings 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Denver, CO, April 1980.

[Leo80] R. G. Leonard. Language recognition test and evaluation. Technical Report RADC-TR-80-83, Air Force Rome Air Development Center, March 1980.

[MC92] Y. K. Muthusamy and R. A. Cole. A segment-based automatic language identification system. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, San Mateo, CA, 1992. Morgan Kaufmann Publishers.

[MCG91] Y. K. Muthusamy, R. A. Cole, and M. Gopalakrishnan. A segment-based approach to automatic language identification. In *Proceedings 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, May 1991.

[SAG91] M. Savic, E. Acosta, and S. K. Gupta. An automatic language identification system. In *Proceedings 1991 IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, May 1991.

[Sug91] M. Sugiyama. Automatic language recognition using acoustic features. In *Proceedings 1991 IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, May 1991.