Generating Language Distance Metrics by Language Recognition Using Acoustic Features

Le Sun [a1], Roland Hu [b1], Huimin Yu [c1], and Timothy J. Sluckin[d2]

[1]*Department of Information Science and Electronic Engineering, Zhejiang University, China*
[2]*Mathematical Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom*

October 15 2016

Abstract

A language recognition system is used to build quantitative measure of language distance. The OpenEAR toolkit is used to extract more than 6,000 features per speech sample. The features consist of 56 low level descriptors (LLDs) and their Delta and Delta Delta values, the corresponding 39 functionals. The language model training component is based on the Gentle AdaBoost algorithm. When tested on a group of 10 principally Indo-European languages, the language recognition system performs comparatively to other language recognizers. The UPGMA tree built from the interlanguage distances identifies the major subgroups of Indo-European. Genetic algorithms are also implemented to generate the language map on the 2D plane. Although some errors remain, the obtained language tree and map are indicators of language relationships. We discuss errors in our system and more generally perspectives for the use of sound file classifiers in historical linguistics.

Index Terms: linguistic distance, language recognition, Gentle AdaBoost

---

[a]lesun@zju.edu.cn

[b]haoji_hu@zju.edu.cn

[c]yhm2005@zju.edu.cn

[d]t.j.sluckin@soton.ac.uk

1

1 Introduction

Computational language recognition, i.e. the automatic identification of language(s) spoken in speech samples, dates back to the 1970s. There are several successful strategies. Zissman's classic approach [1], implemented in the parallel-phone recognition language modeling (PPRLM), classifying languages using phone tokenization of speech combined with a phonotactic analysis of the output. An alternative approach relies on Gaussian mixture models (GMMs), using the acoustic characteristics of the speech utterance [2]. More recent alternative acoustic-based systems (e.g. GMMs [3, 4] and SVMs [5, 6]) perform comparably. The field possesses both numerous commercial applications and great intrinsic interest; over the period 1996-2015, NIST has held seven language recognition evaluations [7] involving both commercial and academic organizations.

Linguists often study language relationships using language distance metrics [8, 9]. This paper uses language recognition algorithms to generate such metrics. World languages are divided into families, known as *phyla* (singular *phylum*) [10]. Within families the relationships can be represented as trees. Linguists use many techniques to construct the trees. For example, the so-called *Swadesh word list* [11, 12] summarises word/meaning shifts between languages, and a metric (the *Levenshtein distance*) is constructed by determining the number of non-cognates that exist in root words in the Swadesh word list. High interlanguage distance implies very diffferent languages. We note that other philologically plausible possible metrics exist (see e.g. [13, 14]). A conceptually different approach [15] defines a language distance from the difficulty a speaker of one of the languages finds in learning the other.

Here by building a language-recognition-based language metric, we are able, least conceptually, to build a possible phylogeny. There are then algorithms which can be used to construct a linguistic phylogeny (family tree), the simplest of which is the so-called UPGMA algorithm [16].We shall remark further at the end of the paper on various issues in theoretical linguistics raised by this research programme.

In the view of historical linguistics, the trees represent a real language evolution history [10, 17]. The standard historical-linguistics technique for establishing common origin is the Comparative Method. This requires the researcher to use well-understood mechanisms of language change, to reconstruct words and grammar in hypothetical ancestor languages (see, e.g. [18]). Distinct language phyla can be identified (e.g. Indo-European, Sino-Tibetan etc.). The trees in each are clear, but between them there is little

connection.

But building metrics only on written languages is incomplete. The spoken language is more basic. Since acoustics could provide much more information than written languages, it is reasonable to build language metrics by considering acoustic influences. Two unrelated or only distantly related languages sound very different to naive listeners unfamiliar with either. For with closely related dialects, observers can sometimes use subtle pronunciation, prosody or vocabulary differences to geographically identify speaker. The human ear, however, has difficulty in quantifying these differences. But computational language recognizers, using measurable features, may automatically measure sound differences between languages or dialects.

In this paper we make some steps to build language metrics using acoustic cues, and also comment on implications for theoretical linguistics. We select acoustic features from a large set (see Hassan et al. [19]). The Gentle AdaBoost algorithm constructs language classifiers using on these features. We build a language classifier for each language pair and use the recognition rate as a distance measure for languages. Lower recognition rates imply more similar languages. The resulting UPGMA language trees are generally but not completely consistent with linguistic approaches, validating our hypothesis that acoustic cues may benefit linguistic analyses. We note other work in the general same spirit (e.g. [20, 21]) but differing much in detail.

Our contribution is twofold. We first propose a language recognition method based on acoustic feature selection by the Gentle AdaBoost algorithm; the recognition accuracy is comparable with current approaches. We then build language distance metrics based on recognition accuracy of the proposed algorithm and show that such metrics are consistent with existing linguistic approaches. Thus, we propose a new analytical tool that may give linguists insight into language phylogeny.

This paper is organized as follows. In 2 we describe our language recognition system. We discuss also the features used in our experiments, and also give a brief review of the traditional AdaBoost algorithm ("TA") as well as the Gentle AdaBoost ("GA") algorithmic strategy. 3 discusses the performance of our language recognition system, as well as the results of the linguistic distance analysis. In 4 we make some brief concluding remarks to put our work in context.

2 Language Recognition Algorithm

The language recognition method has two steps: (a) a feature extraction method is implemented on each speech file to obtain a feature vector; (b) the AdaBoost algorithm is implemented to build language classifiers by using the extracted feature vectors.

2.1 Feature Extraction

Feature extraction uses the OpenEAR toolkit [22]. The feature set consists of 56 low level descriptors (LLDs). Input audio files are divided into 20 millisecond frames, with 10 millisecond overlap. Descriptors are taken for each frame. We also include the Delta value (the value of the current frame minus that of the previous frame), and the Delta Delta value for these descriptors. There are thus in all $56 \times 3 = 168$ descriptors. For each descriptor, we extract 39 functionals, yielding a feature vector with $168 \times 39 = 6,552$ dimensions for each audio file. Descriptor and functional details are given in Table 1. Features are normalized into the range of 0 to 1.

2.2 Gentle AdaBoost for Language Recognition

Given the feature vectors, GA is used to build language classifiers. At this stage a classifier need only distinguish language pairs. Suppose training samples $\{(\boldsymbol{x}_i, y_i)\}, i \in (1, 2, ...m)$. Each $\boldsymbol{x_i}$ represents a feature vector; $y_i \in \{-1, 1\}$ represents the class information; $y_i = \pm 1$ imply $\boldsymbol{x}_i$ belongs to languages (1) and (2), respectively.

From the extracted features we now generate a series of weak classifiers. A weak classifier $h(\boldsymbol{x}, f, p, \theta)$ consists of a feature f, a threshold θ and a polarity $p = \pm 1$ indicating the direction of the inequality:

$$h(\boldsymbol{x}, f, p, \theta) = \begin{cases} 1 & \text{if } pf(\boldsymbol{x}) < p\theta \\ -1 & \text{otherwise} \end{cases} \tag{1}$$

For each feature f, the threshold θ and polarity p are chosen so that the classifier based on eq.(1) obtains the minimum error rate on the training data. This procedure yields 6,552 single-feature related weak classifiers $h_i(\boldsymbol{x}); (i = 1, 2....6552)$.

TA, proposed by Freund and Shapire [23], generates a strong classifier from these weak classifiers. During initialization, we set $\boldsymbol{D}_1(i) = \frac{1}{m}$, where $i = 1, ..., m$. Then we repeat the following training procedure for T times.

4

At every iteration t, we choose a weak classifier h_t which minimizes the error rate with respect to $\boldsymbol{D}_t$.

$$h_t = \arg \min_{h_j} \varepsilon_j, \tag{2}$$

where

$$\varepsilon_j = \sum_{i=1}^{m} \boldsymbol{D}_t(i)[y_i \neq h_j(\boldsymbol{x}_i)]. \tag{3}$$

We then update the probability distribution of $\boldsymbol{D}_t(i)$ given the error rate ε_t:

$$\boldsymbol{D}_{t+1}(i) = \frac{\boldsymbol{D}_t(i) \exp[-\alpha_t y_i h_t(\boldsymbol{x}_i)]}{\boldsymbol{Z}_t} \tag{4}$$

where

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t}, \tag{5}$$

and $\boldsymbol{Z}_t$ is a normalization factor to make all $\boldsymbol{D}_{t+1}(i)$ sum to 1.

$$\boldsymbol{Z}_t = \sum_{i=1}^{m} \boldsymbol{D}_t(i) \exp[-\alpha_t y_i h_t(\boldsymbol{x}_i)]. \tag{6}$$

After T iterations, there are T weak classifiers, a linear combination of which provides the final strong classifier $\boldsymbol{H}(\boldsymbol{x})$:

$$\boldsymbol{H}(\boldsymbol{x}) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})\right) \tag{7}$$

Here α_t is produced during each iteration. It increases when the minimum error rate ε_t decreases. It can be seen that the more confidence the weak classifier has, the more effect it has on the final decision.

TA has the drawback of forcing the weak classifiers to have more influence on misclassified samples. This may lead to the problem that the final strong classifier performs very well on testing samples which are hard to classify, while performance on more easily classified samples may be poor. GA, proposed by Friedman et al. [24], alleviates this problem, by changing the binary output of h_t to continuous values over [-1,1] and replacing the probability distribution update step eq.(4) by:

$$\boldsymbol{D}_{t+1}(i) = \frac{\boldsymbol{D}_t(i) \exp(-y_i h_t(\boldsymbol{x}_i))}{\boldsymbol{Z}_t}, \tag{8}$$

while the final strong classifier generated by eq.(7) is replaced by

$$\boldsymbol{H}(\boldsymbol{x}) = \text{sign}[\sum_{t=1}^{T} h_t(\boldsymbol{x})]. \tag{9}$$

GA omits the variable α_t occurring in TA (see eq.(4)). All contributions from the selected weak classifiers are thus equivalent, causing it to be resistant to noisy data and outliers.

By using TA and GA, we can build classifiers for language pairs. These classifiers are used to produce the language distance metrics for experiments.

3   Experimental Results

Experiments are carried out on the Oregon Graduate Institute Multi-Language Telephone Speech (OGI-TS) corpora [25] which consist of fixed vocabulary utterances as well as fluent continuous speeches collected from 22 languages with around 90 minutes per language from different speakers. Each speaker contributes between one or two minutes of speech. Most utterances last less than 10 seconds, and there are few pieces of continuous speeches lasting more than 30 seconds. The short samples could in principle influence negatively the performance of the classifier.

Experiment 1 tests the proposed language recognition algorithm. We choose 4 languages (English, German, Spanish and Japanese). Each language has 100 different speakers. We randomly select 70 speakers for training and the remaining for testing. Both AdaBoost and Gentle AdaBoost algorithms are applied to these language pairs. The iteration time T is set to 200.

The recognition rates are shown in Table 2. In all cases GA performs better than TA. Note that for the Japanese vs. Spanish, the recognition rate is increased by $\sim 15\%$. Another observation is that the recognition rates are roughly reflections of language distances. For example, it correctly indicates that English is closer to German than to Spanish, and German is closer to Spanish than to Japanese. Admittedly, there are some cases in which recognition rates incorrectly reflect language distances. For example, English seems closer to Japanese than to Spanish. However, in most cases, recognition rate is a reliable distance measure for languages.

Experiment 2 extends the language recognition algorithm to multiple classes. We use the 1992 version of the OGI-TS corpora, with 10 languages: English, Mandarin, German, Spanish, Japanese, Korean, Farsi, Hindi, Tamil

and Vietnamese. We then build 45 classifiers by the Gentle AdaBoost algorithm for each pair of these languages. For testing, an audio file is put into these classifiers and obtain 45 recognition results. Then a voting scheme is implemented and the language which receives the biggest votes is selected as the final result. The proposed method is compared with other two methods (GMM and PPRLM) described in [2], and the recognition error rates are shown in Table 3. Our method performs better than GMM, but worse than PPRLM. We note that PPRLM is a phoneme-level method, i.e., it constructs classifiers for phoneme recognition and then uses the phoneme information for language recognition. However, we here only use acoustic features which can be easily extracted from language files. Thus, our method is a much simpler model than PPRLM, but with a comparable recognition error rate.

Experiment 3 implements the recognition rate as the distance measure between two languages; this distance measure is also compared with other distance measures in linguistics. Supposing that the recognition rate between language A and B is $R(A, B)$, then the distance between A and B is defined as:

$$D(A, B) = 2 \times [(R(A, B) - 0.5] \tag{10}$$

The above definition ensures that the distance is within the range of $[0, 1]$ and meanwhile it is zero in the case of random guess (recognition rate is 0.5).

We firstly compare the proposed distance measure in eq.(10) with Chiswick and Miller's study [26], in which they measure a language distance between English and other languages based on the difficulty level that Americans learn other languages. This comparison is shown in Table 4. The correlation coefficient of these two distance measures is 0.7427, which shows that the two distance measures share significant features.

We next compare the proposed distance metrics with currently accepted linguistic history for a subset of Indo-European languages with one outgroup member. We select eight languages, calculate their recognition rate distances, and then build languages trees using the UPGMA algorithm, as in e.g. [13]. The resulting tree is shown in Fig. 1.

Topologically, this tree is broadly similar to the widely accepted language tree [8, 9, 10, 13], shown in Fig. 2. The most serious anomalies are the apparent large distance between Italian and Spanish, forcing Italian outside a hypothetical Romance clade, and the apparent proximity of Czech and Hungarian, preventing the formation of a Slavonic clade in our tree. In this set of seven Indo-European languages and one other, Hungarian is the linguistic phylogenetic outlier. But it is possible that there is phonological diffusion
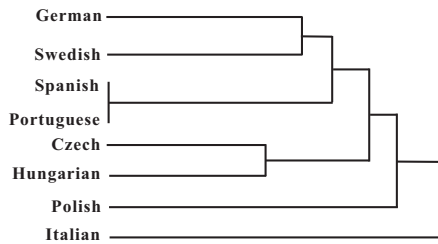
7

Figure 1: Language tree from UPGMA and recognition rate distances.

between Czech and Hungarian resulting from geographical proximity. We note also that in preliminary experiments using other classifiers, we have found Italian and Spanish close together.
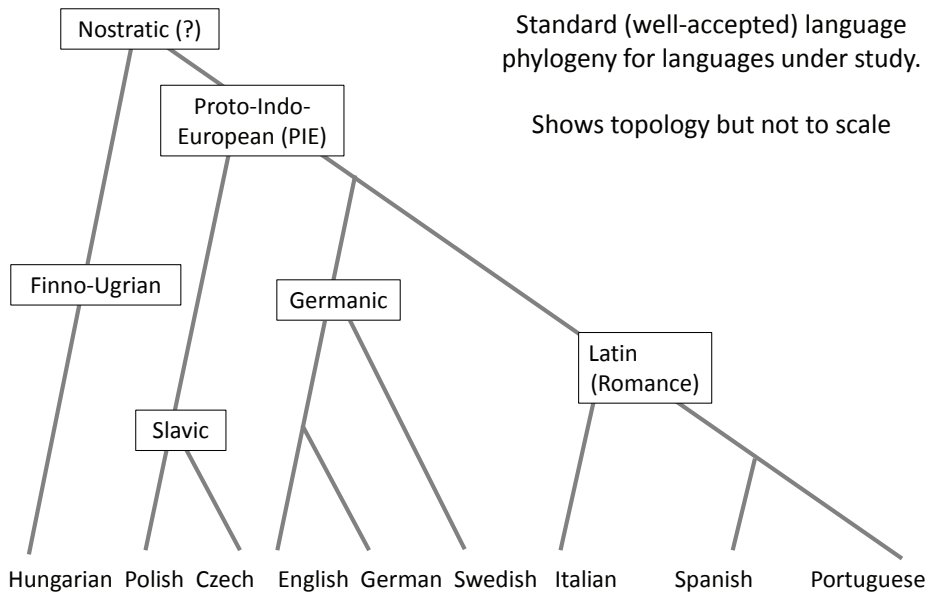

Figure 2: Topological representation of widely accepted language tree for languages in Fig.1 .

Experiment 4 draws a language relationship map using our language metrics. This is an alternative to visualising the distance data through a tree. With N languages there are $N(N-1)/2$ language distances d_{ij} $(1 \leq i < j \leq N)$. The language map assigns a 2D point $\mathbf{r}_i = (x_i, y_i)$ for each language i. The mapping procedure involves solving the following non-linear optimization problem:
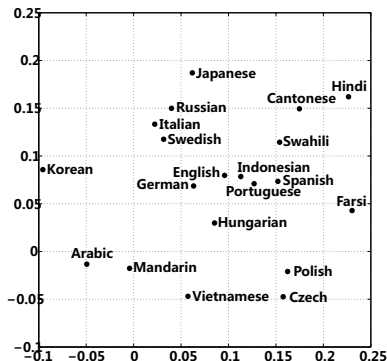
Figure 3: The language relationship map for 20 languages.

$$\text{Minimize:} \quad \sum_{i,j;1 \leq i < j \leq N} |\xi_{ij}|$$

$$\text{subject to} \quad d_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + \xi_{ij},$$

$$\text{where } 1 \leq i < j \leq N \tag{11}$$

The idea is that the data imperfectly represents an underlying clustering, and that the error functions ξ_{ij} represent the effect of hidden variables. We use a genetic algorithm method [27] to solve this problem for $N = 20$ languages.

The final language relationship map is shown in Fig. 3. The map correctly clusters many Western languages into one group, e.g., English, German, Portuguese, Spanish. The Slavic languages Polish and Czech are neighbours. However, other language relationships are poorer; e.g. (a) Italian is in the Western cluster close to Russian and Swedish, not in the Latin Group (Spanish and Portuguese); (b) Russian is far from Polish and Czech.

4   Conclusions

We propose here a new measure of linguistic distance relying on the output of a language recognition system, which involves (a) a new feature set to language recognition and (b) training of the classifier through the Gentle AdaBoost algorithm. System performance on the OGI-TS data corpora outperforms the GMM-based system. The linguistic distances bear some resemblance to previous linguistic studies.

The long-range goal is to extend the system to cast some light on deep language history using acoustic data, by analogy with the construction of trees from syntactic data [9]. We stress that this paper represents a first stage. The results give cause for some optimism, but there are still big gaps.

9

Each part of the problem presents major challenges and requires further study. Linguistically, we may ask: (a) which acoustic features are linguistically conservative and so can be used in order to build phylogenies, (b) how important it is to use phonemic and other data to add to our primarily prosodic data, (c) how to correct for variable speaking speeds in different individuals, and (d) what are the effects of linguistic diffusion which create false homologies (e.g. the Hungarian-Czech homology mentioned above). From a computer science viewpoint, we may: (a) seek to improve the language recognizers in order to identify regions of feature space associated with different languages, and (b) note that UPGMA is easy to apply but very unstable, and seek to employ more intelligent tree-building algorithms, e.g. Bayesian methods [28]. From a statistical point of view it is also important to study the stability of the learning system with respect to changing samples, and to investigate average trees. Finally, as mentioned earlier, the short OGI-TS speech utterance samples may also be a negative factor.

5 Acknowledgements

References

[1] M. Zissman, "Comparision of four approaches to automatic language identification of telephone speech," *IEEE Transactions Speech and Audio Processing*, vol. 4, no. 1, pp. 31–44, January 1996.

[2] P. A. Torres-Carrasquillo, E. Singer, M. A. Kohler, R. J. Greene, D. A. Reynold, J. R. Deller, and Jr., "Approaches to language identification using Gaussian mixture models and shifted delta cepstral features," in *International Conference on Spoken Language Processing*, 2002, pp. 89–92.

[3] L. Burget, P. Matejka, and J. Cernocky, "Discriminative training techniques for acoustic language identification," in *Proceedings of the IEEE*

International Conference on Acoustics, Speech, and Signal Processing, 2006.

[4] V. Hubeika, L. Burget, P. Matejka, , and P. Schwarz, "Discriminative training and channel compensation for acoustic language recognition," in *Interspeech*, Brisbane, AU, 2008, pp. 301–304.

[5] W. M. Campbell, E. Singer, P. A. Torres-Carrasquillo, and D. A. Reynolds, "Language recognition with support vector machines," in *Odyssey 2004*, Toledo, Spain, 2004, pp. 285–288.

[6] W. Campbell, J. Campbell, D. Reynolds, E. Singer, and P. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Computer Speech and Language*, vol. 20, no. 2, p. 210229, 2006.

[7] A. F. Martin, C. S. Greenberg, J. M. Howard, G. R. Doddington, and J. J. Godfrey, "NIST language recognition evaluation past and future," in *Odyssey 2014*, Joensuu, Finland, 2014, pp. 145–151.

[8] F. Petroni and M. Serva, "Language distance and tree reconstruction," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 4, pp. 463–484, 2008.

[9] G. Longobardi, C. Guardiano, G. Silvestri, A. Boattini, and A. Ceolin, "Toward a syntactic phylogeny of modern indo-european languages," *Journal of Historical Linguistics*, vol. 3, no. 1, pp. 122–152, 2013.

[10] F. Bodmer, *The Loom of Language: a guide to foreign languages for the home student.* London: George Allen and Unwin Ltd, 1944.

[11] M. Swadesh, "Lexicostatistic dating of prehistoric ethnic contacts," *Proceedings American Philosophical Society*, vol. 96, pp. 452–463, 1952.

[12] ——, *The origin and diversification of language.* Transaction Publishers, 1971.

[13] D. Benedetto, E. Caglioti, and V. Loreto, "Language trees and zipping," *Physical Review Letters*, vol. 88, no. 4, pp. 1–4, 2002.

[14] H. B. Hamed, P. Darlu, and N. Vallée, "On cladistic reconstruction of linguistic trees through vowel data," *Journal of Quantitative Linguistics*, vol. 12, no. 1, pp. 79–109, 2005.

[15] L. Hart-Gonzalez and S. Lindemann, "Expected achievement in speaking proficiency," *School of Language Studies, Foreign Service Institute, US Department of State, Mimeographed*, 1993.

[16] "http://www.nmsr.org/upgma.htm,," accessed March 15 2016.

[17] D. W. Anthony, *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world.* Princeton University Press, 2010.

[18] B. D. Joseph and R. D. Janda, *The handbook of historical linguistics.* Wiley Online Library, 2003.

[19] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: compensating for covariate shift," *IEEE Transactions Audio, Speech and Language Processing*, vol. 21, no. 7, pp. 1458–1468, 2013.

[20] M. Dunn, A. Terrill, G. Reesink, R. A. Foley, and S. C. Levinson, "Structural phylogenetics and the reconstruction of ancient language history," *Science*, vol. 309, no. 5743, pp. 2072–2075, 2005.

[21] T. F. P. Group, "Phylogenetic inference for function-valued traits: speech sound evolution," *Trends in Ecology & Evolution*, vol. 27, no. 3, pp. 160–166, 2012.

[22] F. Eyben, M. Wöllmer, and B. Schuller, "OpenEAR— introducing the Munich open-source emotion and affect recognition toolkit," in *in Proc. Int. Conf. Affective Comput. Intel. Interact. (ACII'09)*, Amsterdam, The Netherlands, 2009, pp. 1–6.

[23] Y. Freund and R. Shapire, "A decision-theoretic generalization of on-line learning and application to boosting," in *Proceedings of the Second European Conference on Computaional Learning Theory*, 1995, pp. 23–27.

[24] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[25] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, "The OGI multi-language telephone speech corpus," in *Proceedings of the International Conference on Spoken Language Processing, ICSLP'96*, Alberta, Canada, 1992, pp. 1–9.

[26] B. R. Chiswick and P. W. Miller, "Linguistic distance: A quantitative measure of the distance between english and other languages," *Journal of Multilingual and Multicultural Development*, vol. 26, no. 1, pp. 1–11, 2005.

[27] J. A. Joines and C. R. Houck, "On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GA's," in *Proceedings of the IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, 1994, pp. 579–584.

[28] R. D. Gray and Q. D. Atkinson, "Language-tree divergence times support the Anatolian theory of indo-European origin," *Nature*, vol. 426, no. 6965, pp. 435–439, 2003.

Table 1: The 56 low level descriptors and 39 functionals. (a) The 56 low level descriptors (LLDs), (b) The 39 functionals.

Feature Group	Feature Descriptions
Pitch	Pitch (f_0) in Hz and its smoothed contour
Raw Signal	Zero-crossing rate
Voice Quality	Probability of voicing
Energy	Log Energy per frame, and energy in frequency bands 0-25- Hz, 0-650 Hz, 250-650 Hz, 1-4 KHz, respectively.
Mel-spectral	Energy in 26 mel-frequency bands
Cepstral	13 Mel-frequency cepstral coefficients
Spectral	Centroid, flux, relative positions of spectral max and min, spectral roll of points 90%, 75%, 50% and 25%

(a)

Functionals	Number
Relative positions of max/min values	2
Range (max - min)	1
Arithmetic and quadratic means	2
Quartile and inter-quartile ranges	6
95 and 98 percentile values	2
Zero crossings and mean crossing rate	2
Number of peaks and mean distance between peaks	2
Arithmetic mean of peaks	1
Overall arithmetic mean	1
Linear regression coefficients and error	4
Quadratic regression coefficients and error	5
Centroid of contour	1
Standard deviation, variance, kurtosis, skewness	4
Arithmetic, quadratic and absolute means	3
Arithmetic, quadratic and absolute means of non-zero values	3

(b)

Table 2: Recognition rates of 6 language pairs using Traditional Adaboost (TA) and Gentle AdaBoost (GA) algorithms.

Language pairs	AdaBoost	Gentle AdaBoost
Eng. vs Ger.	0.9362	0.9685
Eng. vs Spa.	0.9884	0.9919
Eng. vs Jap.	0.9685	0.9811
Jap. vs Spa.	0.7637	0.9120
Ger. vs Spa.	0.8252	0.8986
Ger. vs Jap.	0.8146	0.9076

Table 3: Recognition error rates of the GMM, PPRLM and the proposed methods.

	GMM	PPRLM	This work (GA)
Error Rate	29%	22%	25%

Table 4: The distance between English and other languages.

	Recognition Rate	The Proposed Distance	Distance as Defined in [26]
Swedish	0.9042	0.8084	0.3333
Italian	0.9111	0.8222	0.3636
Russian	0.9462	0.8924	0.4444
Farsi	0.9367	0.8734	0.5
Polish	0.9605	0.9210	0.5
Hindi	0.9550	0.9100	0.5714
Vietnamese	0.9551	0.9102	0.6667
Arabic	0.9618	0.9236	0.6667
Korean	0.9641	0.9282	1.0