



Basque-Spanish language identification using phone-based methods

Víctor G. Gujarrubia, M. Inés Torres

Departamento de Electricidad y Electrónica
 Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain
 {vgga,manes}@we.lc.ehu.es

Abstract

This paper presents initial experiments in language identification for Spanish and Basque, which are both official languages in the Basque Country in the North of Spain. We focus on three methods based on Hidden Markov Models (HMMs): parallel phone decoding, with no phonotactic knowledge, phone decoder scored by a phonotactic model and single phone decoder scored by a phonotactic model, with phonotactic knowledge. Results for the three techniques are presented, along with others obtained using a neural network classifier. Significant accuracy is achieved when better phonotactic knowledge is used. The use of a neural network classifier results in a slightly improvement and, in overall, similar results are achieved for both languages, with accuracies around 98 %.

Index Terms: language identification, phone decoding, neural networks.

1. Introduction

This paper is focus on language identification (LID) systems. Several are the ways of performing this task, like those exploiting prosodic cues [1, 2], although most of them are based on speech recognition approximations like phone-based approaches [3, 4, 5], Gaussian mixture models [3, 4, 6] or large vocabulary continuous speech recognition approaches [7, 8].

The aim of this paper is to apply phone-based techniques to Spanish-Basque identification. Basque is a minority language which is the joint official language along with Spanish for a community of 2,5 million people living in the Basque Country in the North of Spain. It is considered to be one of the oldest European languages and includes some interesting and infrequent typological characteristics. On the other hand, Basque is more and more present in contemporary life, so interest in developing bilingual industrial applications related to human language technologies has increased. However, speech recognition and understanding systems need a robust Spanish-Basque language identification.

The set of Basque phones is not very different from the Spanish one. The two languages share the same vowel triangle (only five vowels). However, Basque includes larger sets of fricative and affricate sounds. Thus, language identification using phone-based systems is presumably a difficult task.

The rest of the paper is organised as follows: Section 2 describes each of the methods used in this study, Section 3 contains information about the speech databases used in the experiments,

This work was partly sponsored by the Industry Department of the Basque Government and by the University of the Basque Country under grants INTEK CN02AD02 and 9/UPV 00224.310-15900/2004 respectively.

Section 4 presents the results obtained using the different LID approaches, and Section 5 discusses the conclusions and suggest future directions for extending the work presented.

2. Language identification methods

To perform language identification, three phone-based approaches are applied. The first makes no use of any phonotactic knowledge. Two methods based on the phone distributions of each language are then presented. These methods were also combined in several ways using a neural network classifier.

2.1. Parallel acoustic phonetic decoding (PD)

For each language being studied, an unconstrained phone recognizer is applied and the language which gives the most probable acoustic sequence is chosen as the identified language, so no phonotactic knowledge is being used. In our case, the phone recognizer is based on HMMs and performs a Viterbi decoding. The network used to perform the Viterbi decoding for each of the languages did not allow transitions from a phone to the same one and the rest of transitions could be performed without any cost, this is, with a probability of 1.

2.2. Using phonotactic knowledge

In this case, two approaches including phonotactics are used. In both methods, phonotactics are introduced as language models computed from phone sequences obtained from some training corpus.

2.2.1. Phone decoder scored by a phonotactic model (PD+PhM)

This is an extension of the previous method. Instead of using the likelihood of the best phone path to obtain the hypothetical language, a language-dependent phone language model is used to score these phone paths. The language with the higher score is selected as the language of the utterance. If X_l is the best phone path and Ph_l the language model for each of language, then the selected language is obtained as

$$L = \arg \max_l P(X_l|Ph_l) \quad (1)$$

2.2.2. Single phone decoder scored by a phonotactic model (SPD+PhM)

This is similar to the PD+PhM, but instead of performing one decoding per language, only one decoding is performed. For the other language, the decoded phone sequence is mapped and thus,



the PD+PhM method can be applied. In our case, as Basque contains a higher number of phones, the Basque acoustic models are used to perform the phone decoding and the Spanish sequence is obtained mapping the decoded phone sequence.

3. Speech corpora

The experiments reported in this paper were performed using several speech databases.

To train the basic acoustic models for Basque, a phonetically balanced database called EHU-DB16 [9, 10] was used. This database comprises 9394 sentences uttered by 25 speakers and contains around 340000 phones. For Spanish, the phonetic corpus of the Albayzin database [11] was used. It comprises 4800 sentences uttered by 29 speakers, and is also phonetically balanced.

The evaluation set consisted of a weather forecast database recorded for both Spanish and Basque. This database contains 500 different pairs of sentences uttered by 36 speakers. The 500 pairs were divided into blocks of 50 sentences each, and each speaker uttered the sentences in one of these blocks. A total of around 3 hours for Spanish and 3.5 for Basque were recorded, resulting in more than 8000 words for each language. For the experiments showed in Section 4, this database was divided in a training and a testing subset. The training subset was composed of the speakers uttering the initial 300 pairs, resulting in a total of 22 speakers, whereas the rest, 14 speakers, were used for testing purpose.

4. Experimental results

4.1. Conditions of the Experiments

To perform the different experiments, the databases were parametrized into 12 Mel-frequency cepstral coefficients with delta and acceleration coefficients, energy and delta-energy. Thus, four acoustic representations were defined. The length of the analysis window was 25 ms and the window shift was 10 ms.

Each phone-like unit was modeled by a typical left to right non-skipping self-loop three state HMM, with 32 Gaussian mixtures per state and acoustic representation. A total of 35 context-independent phones were used for Basque and 24 for Spanish.

For the PD+PhM and SPD+PhM techniques a language model is needed to score the recognized phone sequence. In these experiments, instead of n-gram models, a *k*-testable in the strict sense (*k*-TTS) model [12] was used, since the *k*-TTS models keep the syntactic constraint of the language. Different *k* values were used: from *k* = 2 (bigram) to *k* = 5 (5-gram).

For each of the languages, several text corpora were phonetically transcribed and used to train a generic phone language model.

4.2. Results

To carry out Spanish-Basque identification experiments, a complete utterance was presented to the system and the different approaches in Section 2 were applied.

4.2.1. Base-system results

Table 1 shows the base-system results, that is, using only the PD, PD+PhM and SPD+PhM methods of Section 2. Results considering only the test subset of the evaluation database are presented. For the PD+PhM approach a recognized-string length normalization was used, since this approximation yielded the best results.

Table 1: Base-system results. Percentage of correctly classified utterances using the phone-based methods. Results for different *k* values are given.

Method		Spanish	Basque
PD		62.86	71.00
PD+PhM	k=2	98.00	72.43
	k=3	97.57	92.86
	k=4	97.29	95.14
	k=5	97.57	96.14
SPD+PhM	k=2	59.86	98.00
	k=3	69.57	99.14
	k=4	77.00	99.14
	k=5	81.86	99.00

Poor performance is achieved using the PD technique and generic HMMs, since the two languages are acoustically similar and Spanish sounds are contained in the Basque ones. Basque sounds could be seen as a more specific variant of the Spanish ones. As suggested in [7], the channel conditions could also be affecting the results and be benefiting Basque identification, since the recording conditions for the test database and the Basque database EHU-DB16 are more similar.

As might be expected, the use of the phonotactic constraint yields a great improvement. The SPD+PhM method does not perform well in overall, since poor results are achieved for Spanish, even if it works perfectly for Basque, probably because Basque is used for the phone decoding in this approach. However, using the PD+PhM method good results are achieved. The use of better language models, even if they make the system run slowly, increases the overall system performance since more accurate information is being used.

4.2.2. Neural-network classifier

To improve the results obtained using the phone-based methods, a neural-network classifier was used.

The classifier was composed of a feed-forward backpropagation network with ten neurons in the hidden layer and one in the output layer, both using radial basis transfer functions.

The training subset of the evaluation database was used to train the neural network. The probabilities used to decide between the languages in the previous subsection were used in this case as the input of the neural network, while the corresponding language was used as the output. The classification experiments were carried out using the test subset.

Table 2 shows the results obtained when using the network combined with the phone methods individually.

In this case, the performance is, in overall, slightly better, especially for the SPD+PhM. The PD approach still performs bad and the PD+PhM technique improves for lower *k* and remains practically unchanged for higher *k*.

Further experiments were carried out using the neural network classifier, but combining more than one technique.

The first experiment consisted of combining the PD+PhM and SPD+PhM methods, resulting in vectors of three elements as input for the neural network. The three probabilities used were the two obtained from the PD+PhM technique and the Spanish probability



Table 2: Percentage of correctly classified utterances using the phone-based methods combined with the neural network classifier individually.

Method		Spanish	Basque
PD		73.43	64.43
PD+PhM	k=2	92.86	91.29
	k=3	97.86	94.14
	k=4	97.86	94.43
	k=5	98.14	95.86
SPD+PhM	k=2	79.57	91.86
	k=3	94.86	93.00
	k=4	96.14	92.14
	k=5	94.86	93.00

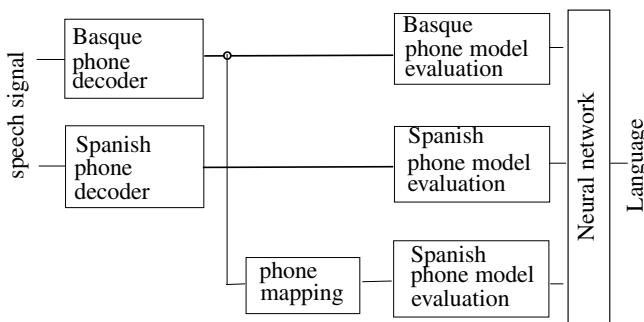


Figure 1: Scheme used to combine the PD+PhM and SPD+PhM methods when using the neural network classifier.

from the SPD+PhM. The Basque probability from the SPD+PhM method was not consider since it was equal to the obtained from the PD+PhM method. The Figure 1 shows the scheme representing this approach.

Table 3 shows the results achieved in this case.

Table 3: Percentage of correctly classified utterances when combining the PD+PhM and SPD+PhM methods using one neural network classifier with 3 inputs.

	Spanish	Basque
k=2	93.57	93.29
k=3	98.14	95.14
k=4	98.14	95.29
k=5	98.00	96.00

These results show a slightly improvement when comparing to the previous ones, specially for the PD+PhM. The inclusion of the Spanish probability of the SPD+PhM method gave more information to the classifier, which resulted in improvements for both languages.

The next experiment consisted of combining all the three techniques: PD, PD+PhD and SPD+PhD, resulting in vectors of five elements as input for the neural network. A scheme similar to the one showed in Figure 1 is used, but adding the probabilities ob-

tained from the PD method, that is, we combined the two probabilities from the PD method, the two from the PD+PhM method and the Spanish probability from the SPD+PhM method. The Basque probability from the SPD+PhM method was again not consider.

Table 4 shows the results achieved in this case.

Table 4: Percentage of correctly classified utterances when combining the PD, PD+PhM and SPD+PhM methods using one neural network classifier with 5 inputs.

	Spanish	Basque
k=2	94.29	92.57
k=3	97.57	96.00
k=4	97.00	96.86
k=5	97.29	97.14

In this case, the percentage of correctly classified utterances is smaller for Spanish, but higher for Basque, probably because the PD method did not work well for Spanish and a bit better for Basque. Nevertheless, even if the Spanish results are worse than the obtained with other approaches, in overall this approach works similar for both languages and high accuracies are achieved.

5. Conclusions and further work

Several LID systems for Spanish-Basque identification are presented. As might be expected, some of the experiments in the identification of Spanish-Basque resulted in major identification errors, due to the fact that Basque and Spanish share most of the sounds, but the inclusion of better phonotactic knowledge resulted in a great improvement. The use of a neural network classifier reported slightly improvements and high accuracies were achieved for both languages.

In the future, we intend to record the weather forecast database used for Spanish and Basque also in American English. The idea is to repeat some of these experiments, alone with new ones, to extend the study performed for Spanish and Basque also to English. We also intend to use other techniques and classifiers.

6. References

- [1] S. Itakahashi and L. Du, "Language identification based on speech fundamental frequency," in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH)*, Madrid, Spain, 1995, vol. 2, pp. 1359–1362.
- [2] L. Mary, K. Srinivasa Rao, and B. Yegnanarayana, "Neural network classifiers for language identification using phonotactic and prosodic features," in *Proceedings of International Conference on Intelligent Sensing and Information Processing*, Chennai, India, 2005, pp. 404–408.
- [3] M. A. Zissman and E. Singer, "Automatic language identification of telephone speech messages using phoneme recognition and n-gram modelling," in *Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Adelaide, Australia, 1994, vol. 1, pp. 305–308.
- [4] E. Singer, P. A. Torres-Carrasquillo, T. P. Gleason, W. M. Campbell, and D. A. Reynolds, "Acoustic, phonetic and



- discriminative approaches to automatic language identification,” in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland, 2003, pp. 1349–1352.
- [5] J. Navrátil and W. Zühlke, “An efficient phonotactic-acoustic system for language identification,” in *Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Seattle, USA, 1998, vol. 2, pp. 781–784.
- [6] E. Wong, J. Pelecanos, S. Myers, and S. Sridharan, “Language identification using efficient gaussian mixture model analysis,” in *Proceedings of the Australian International Conference on Speech Science and Technology*, 2000, pp. 78–83.
- [7] T. Schultz, I. Rogina, and A. Waibel, “Lvcsr-based language identification,” in *Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Atlanta, USA, 1996, pp. 781–784.
- [8] F. Metze, T. Kemp, T. Schaaf, T. Schultz, and H. Soltau, “Confidence measure based language identification,” in *Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000.
- [9] V. Guijarrubia, I. Torres, and L.J. Rodríguez, “Evaluation of a spoken phonetic database in basque language,” in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, 2004, vol. 6, pp. 2127–2130.
- [10] Karmele López de Ipiña, Inés Torres, and Lourdes Oñederra, “A speech database in basque language,” in *Workshop on Language Resources for European Minority Languages*, Granada, Spain, 1998.
- [11] A. Moreno, D. Poch, A. Bonafonte, E. Lleida, J. Llisterri, J. B. Mariño, and C. Nadeu, “Albayzin speech database: Design of the phonetic corpus,” in *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, Lisbon, 1993.
- [12] I. Torres and A. Varona, “k-tss language model in a speech recognition system,” *Computer Speech and Language*, pp. 15(2):127–149, 2001.