Odyssey 2010
The Speaker and Language Recognition Workshop
28 June – 1 July 2010, Brno, Czech Republic

Exploiting variety-dependent Phones in Portuguese Variety Identification

Oscar Koller[12], Alberto Abad[1], Isabel Trancoso[13]

[1] L^{2}F - Spoken Language Systems Lab, INESC-ID Lisboa, Portugal
[2] Berlin University of Technology, Germany
[3] IST Lisboa, Portugal
oscar@l2f.inesc-id.pt

Abstract

This paper presents a new approach of building a language identification system using a specialized Phone Recognition system followed by Language Modeling (PRLM) to differentiate Portuguese varieties spoken in African Countries from European Portuguese. The system is designed to focus on exploiting the phonotactic information of a single discriminatively trained tokenizer for the specific pair of target varieties. In contrast to other PRLM-based methods, the single tokenizer already combines distinctive knowledge about the differences between both target varieties. This knowledge is introduced into a dedicated multiple-stream Multi-Layer Perceptron (MLP) phone recognizer by training mono-phoneme models for two varieties as contrasting phoneme-like classes within a single tokenizer. Significant improvements in terms of identification rate and computational cost were achieved compared to a conventional single tokenizer PRLM-based systems and to the combination of up to five parallel PRLM identifiers. The method is also applied to other varieties of Portuguese yielding similar results.

Variety identification; Portuguese varieties

1. Introduction

With around 178 Million L1-speakers, Portuguese is the seventh most spoken language in the world [1]. About five percent of the Portuguese speakers live in Portugal and consequently speak Portuguese with European accent. Automatic captioning of broadcast news (BN) - L^{2}F's core technology - faces heavy difficulties in the presence of different accents. For instance, the word error rate (WER) of L^{2}F's baseline European Portuguese (EP) recognizer degrades from under 20% with EP speech to nearly 30% with African Portuguese (AP) varieties. In order to overcome the challenges imposed by the presence of multiple varieties of Portuguese in BN data, variety dependent recognition systems and efficient variety identification modules are needed.

L^{2}F's PostPORT project (Porting Speech Technologies to other varieties of PORTuguese) focuses on these needs. At INESC-ID, we have been working for several years on Large Vocabulary Continuous Speech Recognition (LVCSR) using hybrid recognizers, combining Artificial Neural Networks and Hidden Markov models (ANN/HMM), the so-called connectionist paradigm. Our first LVCSR system was initially developed for European Portuguese. It was recently ported to the Brazilian Portuguese (BP) variety [2]. However, porting it to AP varieties caused severe difficulties, due to the limited amount of manually transcribed data available for training.

As manual transcriptions are expensive goods, the best alternative is unsupervised training of the acoustic models with automatically transcribed data, which has proven to have significantly lowered the WER in our EP recognizer. However, in the case of AP we face the problem of having only access to broadcast news shows, that include speakers from different countries, speaking a broad variety of different accents of Portuguese. Unsupervised training cannot be realized without having a variety verification module to solely select speech with heavily accented AP. Moreover variety specific speech recognition needs a module choosing the appropriate variety dependent speech recognizer. Hence, the need for variety identification systems arises twice to fullfill the PostPORT project.

This paper deals with a promising implementation of such a system. We present a new approach of building a specialized language identifier, based on Phone Recognition and Language Modeling (PRLM), to differentiate Portuguese varieties spoken in African Countries with Portuguese as the official language in a highly accurate and efficient manner from European Portuguese.

The motivation to consider AP as a broad class and not all African varieties separately comes from our previous work [3], where a human benchmark revealed that identifying African varieties in BN is much harder than identifying accents of everyday's people on the street, possibly due to higher level of education and contact with EP in BN. Additionally, the reduced amount of available data of some of the varieties was also a reason for considering this broad classification incorporating all Portuguese varieties spoken in the PALOP Countries (African Countries with Portuguese as Official Language), namely Angola, Cape Verde, Guinea-Bissau, Mozambique and São Tomé and Príncipe.

In the following section 2 we introduce a conventional PRLM system, that serves as baseline. Our proposed approach is presented in Section 3. The experimental results are shown in Section 4. In Section 5, the approach is extended to other Portuguese varieties. A discussion of all results will be found in Section 6, before presenting final conclusions in Section 7.

2. Baseline Variety Identification System

There are several approaches to tackle the problem of automatic language -or more specifically variety- identification that one can find in the literature. Most common approaches include acoustic, phonotactic or even prosodic based methods [4][5][6]. Phonotactic methods have been usually considered one of the best performing approaches. In general, variety identification

deals with much more subtle differences than language identification. Two varieties share a much closer phoneme set. Identification systems often employ several systems in parallel to ensure sufficient performance.

We want to focus on phonotactic systems, following the PPRLM (Parallel Phone Recognition and Language Modeling) approach [7]. In the next subsections we explain the general phonotactic identification approach, introducing a baseline system joining five common phonetic classifier for EP, BP, Spanish, English and AP.

2.1. Baseline PPRLM System

The key aspect of PRLM systems are robust phonetic classifier that generally need to be trained with word-level or phonetic level transcriptions. The tokenization of the input speech data in both training and testing sets is done with the neural networks that are part of our hybrid recognition (AUDIMUS). This type of recognizer is generally composed by one or more phoneme classification networks, particularly MultiLayer Perceptrons (MLP). We use these phonetic recognizers to generate token sequences. The same classifier, fed with speech of different varieties, produces different sequences of tokens. In training mode, we use them to train variety-specific statistical language models. The posterior probability of belonging to a certain variety can be estimated in test mode by comparing the token sequences for a given speech frame (and its context) with that variety's trained statistical model.

2.2. Feature Extraction and Corpora used to train the Phonetic Classifier

A basic system combines three MLP outputs trained with Perceptual Linear Prediction features (PLP, 13 static + first derivative), PLP with log-RelAtive SpecTrAl speech processing features (RASTA, 13 static + first derivative) and Modulation SpectroGram features (MSG, 28 static).

L^2F's EP classifier was trained with 378 hours of training data, of which 332 were automatically transcribed, using word confidence measures [8]. The BP classifier was trained with more than 46 hours of BN data from the Record channel transmitted by cable TV in Portugal, out of which 13 had been manually transcribed (PostPORT BP training corpus [2]). The Spanish classifier used a total of 164 hours BN data, 17 hours manually transcribed. Finally, the English system was trained with the HUB-4 96 and HUB-4 97 data sets, that contain around 142 hours of TV and Radio Broadcast data. The AP classifier is not yet considered to make part of our LVCSR system, as its performance is still not satisfactory. It is an adaptation of L^2F's current EP recognizer trained with three epochs of nearly 6 hours manually transcribed AP BN data (PostPORT AP training corpus). It is hence our only baseline classifier that has been trained with data from two different varieties.

The size of the neural networks of each tokenizer differs due to the different amounts of training data. The context windows of the MLP networks trained with PLP and RASTA features is fixed to 13, while a context of 15 frames was considered more appropriate for MSG features. The EP networks have two hidden layers of 500 weights and an output layer of 39 weights. The BP networks have also two hidden layers of 500 units and an output layer of 40 units. The size of the Spanish network is 500 weights for the two hidden layers and 30 weights for the output layer. The English networks use two hidden layers of 1000 weights and an output of 41 units. The AP networks consist of two hidden layers with 2000 weights each and have 39 output weights. Note, that the size of the output layer corresponds to the number of phonetic units of each language, plus silence (no additional sub-phonetic or context-dependent units have been considered [9]). The phonetic classifers are summerized in Tab. 1.

	Train Data [hours]		Layer Size	
Classifier	total	manual	Hidden	Output
EP	378	46	500	39
BP	46	13	500	40
ES	164	17	500	30
EN	142	142	1000	41
AP	378+6	46+6	2000	39

Table 1: *Overview of employed baseline classifier.*

2.3. Phonotactic modeling

For every phonetic tokenizer, the phonotactics of each target language are modeled with a 3-gram back-off model, that is smoothened using Witten-Bell discounting. For that purpose the SRILM toolkit is used [10]. We model the token sequences for each target variety with every tokenizer separately. To model the EP variety, 1283 segments of the ALERT corpus are used, being about 279 minutes of spoken audio from manually transcribed Broadcast news. The AP model is trained with 1424 segments, adding up to 240 minutes, of the PostPORT AP training corpus. See Tab. 2 for details.

In both training and test, the raw phonotactic sequence obtained by each tokenizer is filtered, in order to avoid spurious phone recognitions. Concretely, phones that appear only once in the middle of long sequences of identical phones are deleted and only transitions between phones are considered in the language model.

2.4. Calibration and fusion

Calibration of each individual PRLM and the right weighting for the fusion of all five parallel systems is needed. Linear logistic regression fusion and calibration is done with the FoCal Multiclass Toolkit [11]. We perform a single crossfold calibration for all segments of different lengths. The final calibration and fusion weights correspond to the mean of five independent calibrations, each using random 20% of the test set. The test set is composed by segments from the EP ALERT and the PostPORT AP testing corpus. We selected the data to ensure the same speakers do not appear in train and test simultaneously. As for EP test data, we use 412 segments, an equal of 99 minutes. The AP test data contains 610 segments with a total of 89 minutes. Details can be found in Tab. 3.

3. Single mono-phonemic PRLM using a specialized phonetic tokenizer

Our proposed system also follows the PRLM approach. We focus on the phonetic classifier and try to build a system, that performs well with a single, but highly specialized, tokenizer. Since two varieties have phoneme sets that are very similar, but not identical, we aim at training a classifier that incorporates the differences between AP and EP on a phonetic level. To better characterize these differences, we divide all occurring

Train Data	AP	EP	$\sum$
duration [min.]	238.8	279.1	517.9
segments	1424	1283	2707
∅ dur./segm. [s]	10.1	13.1	11.5
<3s [%]	16.9	0.1	8.9
3-10s [%]	42.3	49.6	45.8
10-30s [%]	38.7	44.1	41.3
>30s [%]	2.2	6.3	4.1

Table 2: *Data to train the language models (train mode).*

Test Data	AP	EP	$\sum$
duration [min.]	88.8	99.0	187.8
segments	610	412	1022
∅ dur./segm. [s]	8.7	14.4	11.0
<3s [%]	23.3	0.2	14.0
3-10s [%]	43.1	42.5	42.9
10-30s [%]	32.8	50.0	39.7
>30s [%]	1.0	7.5	36.2

Table 3: *Data for evaluation of identification (test mode).*

phonemes in our varieties into the following two groups [12]:

1. mono-phonemes: phonemes whose acoustic realizations in one language overlap little or not at all with those in another language (e.g. the English /r/ vs. the German /r/).
2. poly-phonemes: phonemes whose acoustic realizations are similar enough across the languages to be equated (e.g. /sh/ in English and German)

Phonotactic approaches are able to benefit from both types of phonemes in order to classify speech. With the help of statistical modeling differently occurring poly-phonemes carry important information through the sequences they appear in. If mono-phonemes are found in speech they could, at least in theory, instantly help to differentiate our varieties.

To incorporate this idea, our phonetic classifier needs to be able to identify mono-phonemes of each variety. We therefore train our classifier network with data from both AP and EP simultaneously. Each of the varieties contains a set of 38 phonemes plus silence. If we considered every phoneme out of these sets to be a mono-phoneme, we could train a phonetic tokenizer generating 76 different output tokens plus silence, corresponding to the phonetic realization of each phoneme in its variety. However, since our varieties resemble each other this is not convenient, as it would prevent the phonetic classifier to converge and the resulting system would be poorly trained. We rather need to find out which phonemes actually have mono-phonemic properties, instead of training a phone-recognizer with 76 output phonemes.

3.1. Determining mono-phonemes

Determining the set of phonemes, which is unique for a certain language, given its neighboring variety is not straightforward. Linguistic knowledge about the varieties' phonological characteristics is crucial, but often not available, not sufficiently detailed or controversial. We use a computational method instead to find variety dependent unique phonemes. Binary MLPs are trained to discriminate between the same pairs of aligned phoneme classes. It is worth noting that the selected phonemes

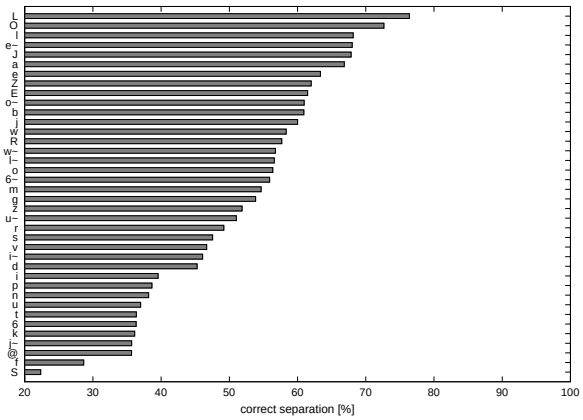

Figure 1: *Correct separation of AP and EP by each SAMPA phoneme using a binary classifier.*

are no mono-phonemes from a strict linguistic point of view, but rather mono-phoneme-like units. For the sake of simplification the term 'mono-phonemes' is nevertheless used in the scope of this paper to refer to these units.

L^2F's baseline EP Automatic Speech Recognition (ASR) system [13] is used to align the training and development data of both varieties. To train a binary classifier, that allows us to see if an AP representation of a certain phoneme differs from its EP counterpart, we keep solely those two phonemes in our training data to train a binary classifier. All other phonemes are removed. The chosen two phonemes are given distinct output classes. After training, the successful separation of both classes can be verified using development data. This is a fast process, which enables the training of binary classifier to be performed for all 38 pairs of phonemes. In this way we can determine, which phoneme classes are different enough to be successfully distinguished and hence contain mono-phonemic characteristics. The performance of all binary classifier is shown in Fig. 1.

We choose the eight best performing phoneme classes as mono-phoneme units, namely [L], [O], [I], [e~], [J], [a], [e] and [Z] using Portuguese SAMPA (Speech Assessment Methods Phonetic Alphabet). After preliminary experiments with various numbers of mono-phonemes this seemed to be a good trade-off between network complexity, training data and classification performance. Eight mono-phonemes were further chosen in correspondence with the 'apbp' mono-phonemic system, which is the weakest among those analyzed in this work (see 5.2 for details). This leads to a phonetic recognizer with 30 poly-phonemes and two times eight mono-phonemes plus silence, thus 47 outputs, which seems reasonable considering the 14 hours of available training data.

It is interesting that some of the chosen mono-phonemes, namely [L] and [J], are the least frequently occurring phonemes in our train data. It needs to be verified that this fact did not influence the process of selection. A further linguistic interpretation of the results is part of our future work plans.

3.2. Phonotactic Core and Corpora

As our baseline system in section 2.2, the mono-phonemic phone-recognizer combines three MLP outputs trained with PLP, RASTA and MSG features. It is based on a single phone-recognizer with subsequent statistical modeling of the phoneme

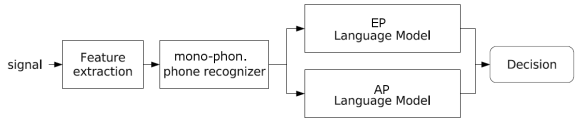

Figure 2: *Block diagram of our specialized PRLM system.*

occurrences, as shown in Fig. 2. To train the mono-phonemic neural network, both AP and EP data are taken. L^{2}F PostPoRT's AP training and development corpus are used, with 430 minutes for training and 49.8 minutes for development of manually transcribed emissions of broadcast news. Moreover a part of the EP ALERT corpus is used to equally balance AP and EP data.

To train our phonotactic models, as well as for performance evaluation, we use the same corpora as for training and assessment of our baseline, introduced in section 2.3.

4. Variety Identification Results

In Fig. 3 the results of previously introduced systems are given. The presence of AP or EP in a given audio segment is verified. Being a binary decision, this is equivalent to an identification of AP, which is appropriate for our task of selecting uniquely AP data for unsupervised training. Results achieved by the single baseline tokenizers using EP, Spanish (ES), BP, English (EN) and AP classifier are presented. Moreover we tested a fusion of the AP and the EP tokenizer, a fusion of all five systems and the specialized mono-phonemic classifier alone. The mono-phonemic system is labeled with lower case letters, whereas all baseline results are named using capital letters. The first and brightest bars in Fig. 3 show the percentage of audio segments that contain an AP speaker, but are not classified as AP (AP_{Pmiss}). The second and darkest bars show the false classification of EP speakers as AP (AP_{Pfa}). The third bars display the average detection cost. The values AP_{Pmiss} correspond to EP_{Pfa} and EP_{Pmiss} equals AP_{Pfa}, as we are dealing with binary identification. It becomes obvious that the mono-phonemic approach outperforms all other single or combined systems. We achieve a relative reduction of the average detection cost of around 60% compared to our baseline systems, with values for $AP_{Pmiss} = 3.6\%$, $EP_{Pmiss} = 5.6\%$ and an average detection cost of 4.6%. Moreover, as we are dealing with a single PRLM, the processing cost is reduced drastically compared to the parallel systems.

5. Application to other Varieties

In order to confirm relevance for variety recognition in general our mono-phonemic approach needs to produce comparable results with other varieties. Hence, we now apply the procedure to systems recognizing EP versus BP, AP versus BP. We also introduce a single PRLM differentiating all three varieties. Training and testing of the BP variety's statistic models is done with data from L^{2}F PostPORT's BP training corpus. See Tab. 4 for details.

In all cases we first train binary classifier to determine the mono-phonemes to be chosen. Then we train the phonetic classifier, the statistical language models and finally calibration and fusion are performed.

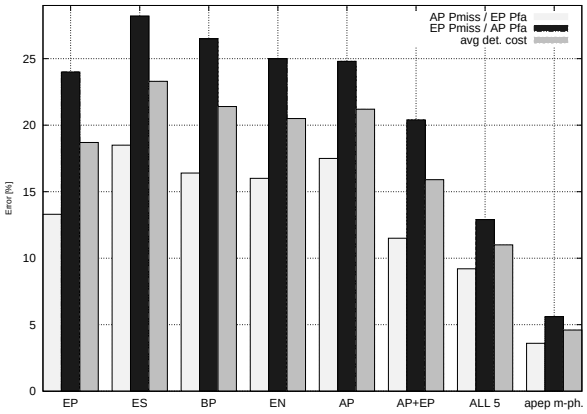

Figure 3: *Variety Identification of AP vs EP .*

	Train BP	Test BP
duration [min.]	256.1	80.2
segments	1434	462
∅ dur./segm. [s]	10.7	10.4
<3s [%]	12.8	18.0
3-10s [%]	44.4	40.0
10-30s [%]	38.7	38.1
>30s [%]	4.1	4.1

Table 4: *BP Data for training and evaluation of the statistical models.*

5.1. European Portuguese and Brazilian Portuguese

For the purpose of training the phonetic classifier, we use around 400 minutes of PostPORT's manually transcribed BP training and 90 minutes from the development corpus. We use the same EP corpus, as mentioned in section 3.2. For alignment we work with L^{2}F's ported BP ASR system [2], which produces more accurate results. It already incorporates a slightly different phoneme-set than the EP system, with [tS], [dZ] and [x] as phonemes that appear uniquely in BP and without [@] and [l~], that are EP-specific. To keep the same recognizer layout with 47 outputs as in our AP/EP mono-phonemic system (see Section 3.1), we choose additional five mono-phonemes through a binary classification, namely [o~], [j], [R], [u~] and [6~]. The result of all binary classifier can be seen in Fig. 4.

Results can be seen in Fig. 5. The first bars show BP segments that have not been classified as such (BP_{Pmiss}). This corresponds to EP_{Pfa}. Bars with the darkest color represent EP audio segments that have not been classified as EP (EP_{Pmiss}). In our case of a binary identification this corresponds to segments falsely classified as BP (BP_{Pfa}). The colors used to fill the bars and the varieties they refer to correspond in all result figures (Fig. 3, Fig. 5 and Fig. 7). We achieve values for $BP_{Pmiss} = 4.9\%$, $EP_{Pmiss} = 7.2\%$ and an average detection cost of 6.1%. The mono-phonemic reaches similar results as the fusion of all five baseline systems, although it performs slightly worse, which is due to the selection of mono-phonemes, as experiments with a different choice showed. The fusion of BP and EP baseline also performs well. We can see that the average detection cost of our single baseline PRLMs are about half as high as in the AP vs. EP identification in Fig. 3. The 'EP' and the 'AP' single baseline identifier perform much

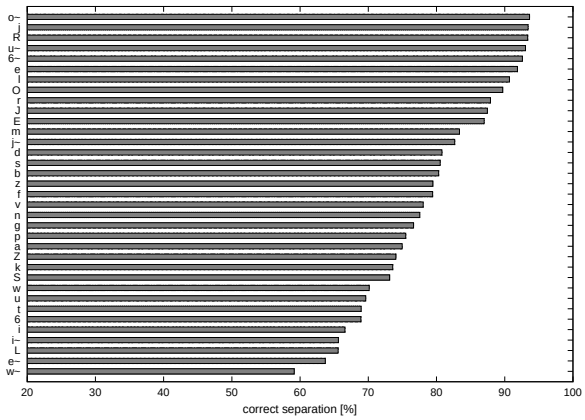

Figure 4: *Correct separation of BP and EP by each SAMPA phoneme using a binary classifier.*

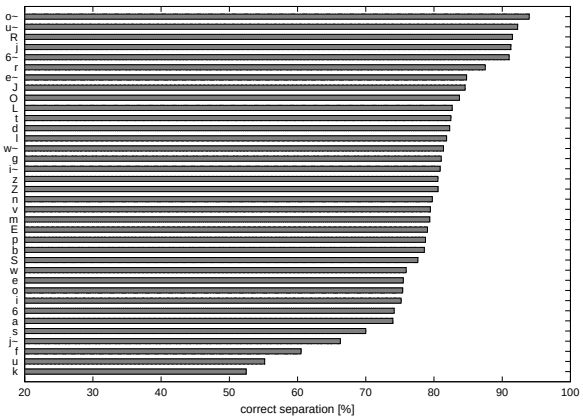

Figure 6: *Correct separation of AP and BP by each SAMPA phoneme using a binary classifier.*

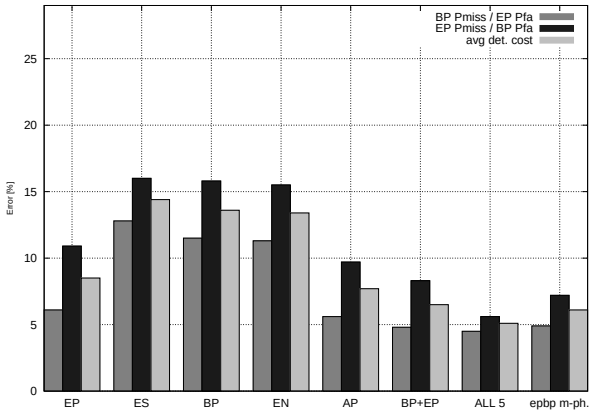

Figure 5: *Variety Identification of BP vs EP.*

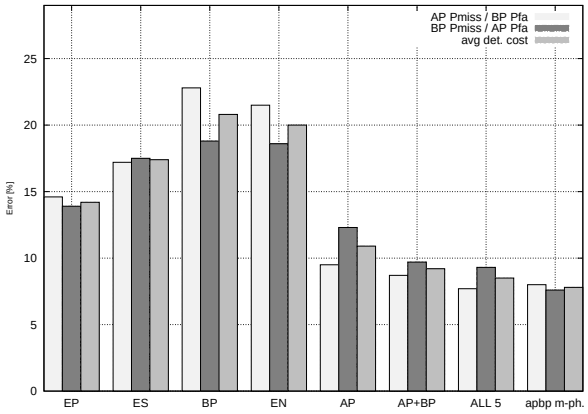

Figure 7: *Variety Identification of AP vs. BP.*

better than the Spanish ('ES'), 'BP' or the English ('EN'). Note that the AP phonetic classifier is an adapted version of the EP MLP and has thus been trained with EP and with AP data.

5.2. African Portuguese and Brazilian Portuguese

The mono-phonemic classifier to identify AP and BP is trained with the PostPORT AP and BP training and development corpora, with around 430 and 400 minutes respectively. The AP corpora are aligned using L^2F's EP ASR system, whereas we use the ported BP version for BP corpora. We identify five additional mono-phonemes, namely [o~], [u~], [R], [j] and [6~], achieving higher seperability than the rest, as can be seen in Fig. 6. This leads to a recognizer layout with 47 outputs as in the AP/EP and EP/BP mono-phonemic systems.

For the identification task AP versus BP the mono-phonemic system reaches $AP_{Pmiss} = 8.0\%$, $BP_{Pmiss} = 7.6\%$ and an average determination cost of 7.8%, as can be seen in Fig. 7. As a result, our approach shows a slightly better performance than the fusion of five baseline PPRLM and than the fusion of AP and BP ('AP+BP'). The single PRLM systems perform very differently. 'BP' and 'EN' achieve worst results, followed by the Spanish ('ES') and the 'EP' PRLM. The 'AP' classifier performs much better. However, as mentioned before, it was trained with data from two varieties.

5.3. African, Brazilian and European Portuguese in a single PRLM

We further want to analyze, if the performance gain resulting from a mono-phonemic classifier trained with two varieties can be reproduced in scenarios, where three varieties need to be distinguished. We hence train a phonetic tokenizer with AP, BP and EP data. The same corpora as in the other mono-phonemic systems is used. A total of 21 mono-phonemes distributed over the three varieties have been chosen, adding up to 57 different output tokens.

In Fig. 8 we show results of all baseline systems, a fusion of the three baseline systems 'EP+BP+AP', a fusion of all five baseline systems ('ALL 5'), each of the previous mono-phonemic classifier separately ('apep', 'apbp' and 'epbp'), a fusion of the three classifier trained with two varieties ('apep+apbp+epbp') and the triple mono-phonemic PRLM 'apepbp'. For each system, we display the average detection cost for a certain variety in the three shades light (AP), middle (BP) and dark grey (EP). It has to be noted, that the average detection cost for a specific variety includes the false alarms of the other two varieties. The mono-phonemic classifier trained with two varieties suffer heavy degradation by false alarms caused by the third variety, they have not been trained for.

Fig. 8 shows that the three baseline systems 'ES', 'BP' and

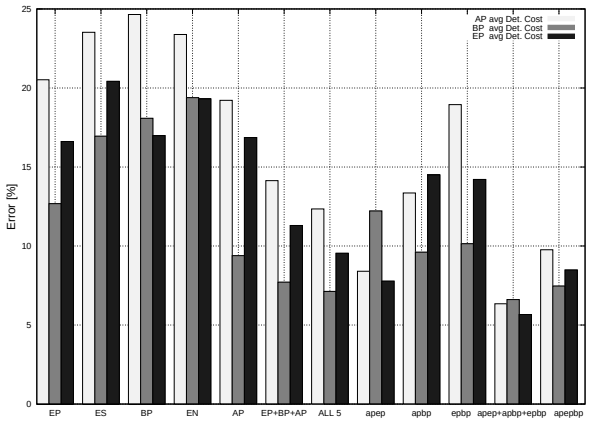

Figure 8: *Average Detection Cost of AP vs. EP vs. BP.*

'EN' do not perform well, with an average detection cost of around 20% for all varieties. On all baseline PRLMs, including the fusion of three and five systems, we observe that AP has a significant higher detection cost than BP, in some cases twice as high. This tendency partially also applies to the detection of EP. The mono-phonemic classifier 'apep' shows clearly that it was trained to differentiate well AP and EP, with very low detection costs in these varieties. The other two mono-phonemic classifier ('apbp' and 'epbp') do not perform that well, which is due to a high false alarm of the corresponding third variety, they have not been trained with. We further see in Fig. 8 that the fusion of all three mono-phonemic classifier 'apep+apbp+epbp' outperforms all other approaches. However, the single mono-phonemic 'apepbp' identifier achieves better performance than any other single classifier.

In Fig. 9 we see the Detection Error Tradeoff (DET) curves of all mono-phonemic approaches, the strongest single Baseline ('AP') and the fusion of all five baseline PRLMs as a averaged tradeoff between missed detections and false alarms over all three target varieties. As all curves are close to be straight lines, their underlying likelihood distributions are normal. The fusion of the three mono-phonemic systems outperforms all other approaches. However the triple mono-phonemic classifier ('apepbp') is clearly the best single PRLM-based classifier. The fusion of all five baseline systems clearly achieves better results than the 'apbp' and 'epbp' mono-phonemic systems. Again, this is due to influence of the third variety.

6. Discussion

The identification results show that out of the three pairs of varieties AP and EP are most difficult to distinguish. Our baseline performs very weak with these varieties. However, our mono-phonemic approach performs best here. It also achieves accurate results detecting BP versus EP, nevertheless our baseline is able to produce similar results, that are much stronger than its performance identifying AP versus EP. The same applies to AP versus BP, where especially the AP baseline performs very well. With these varieties the mono-phonemic system still produces good but not overwhelming rates. It is worth noting that we are comparing a single mono-phonemic PRLM to fusions of up to five parallel systems. As identification of AP and EP proved to be most difficult, we can conclude that mono-phonemic approaches produce particularly good results for very close varieties, where common systems fail. Moreover, with more distinct varieties there is not as much relative improvement possible, as the baseline already classifies well.

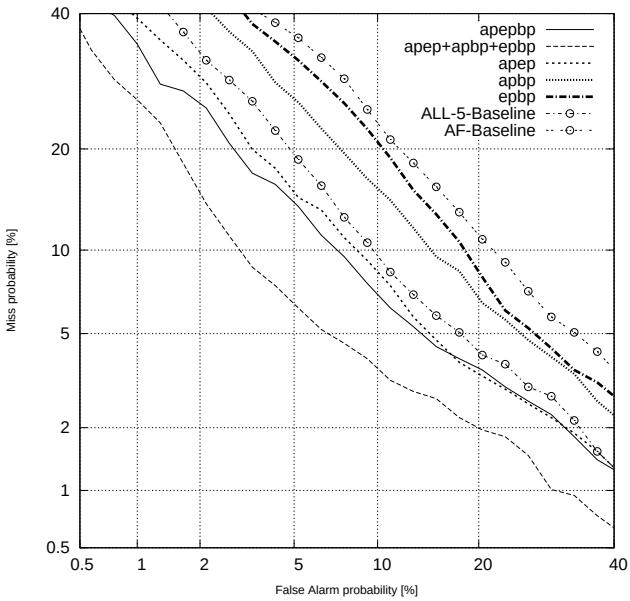

Figure 9: *DET-Curve for Identification of AP vs. BP vs EP*

Even the detection of three varieties justifies the use of a single mono-phonemic classifier. To achieve best results however, systems should be trained for two varieties and then employed in parallel.

7. Conclusion

In this paper we presented an approach to exploit variety-dependent phonemes for the identification of AP and EP. We were able to reduce the average detection cost to less than 50% of the previous value, from 11% to 4.6%. Moreover our approach proves to be very efficient, employing just a single mono-phonemic tokenizer instead of five parallel system used before.

We also showed that our solution produces good results with other combinations of varieties, reaching an average determination cost of 6.1% on EP and BP and 7.8% identifying AP versus BP. Nevertheless future research needs to be done with more varieties to confirm general relevance and applicability to other languages.

Future work also includes experiments with more data for train, calibration and test of the language models. Moreover we need better understanding of how to choose the mono-phonemes, as further improvement can possibly be achieved with different combinations and different number of chosen phonemes. We limited our number of mono-phonemes due to the amount of data available to train the phonetic classifier. We therefore need to evaluate effects of more training data and consequently more mono-phonemes.

Finally, applying the mono-phonemic approach to varieties without available transcribed data, could be an interesting future investigation. This could be tried using automatic transcriptions from one of our speech recognition systems, analog to the "root phonetic recognizer approach" presented in [14].

8. Acknowledgments

This work was funded by FCT project PTDC/PLP/72404/2006.

9. References

[1] M. P. Lewis, *Ethnologue: Languages of the World, 16th Edition*, SIL International, 16th edition, May 2009.

[2] A. Abad, I. Trancoso, N. Neto, and M. C Viana, "Porting an european portuguese broadcast news recognition system to brazilian portuguese," *Interspeech 2009, ISCA, Brighton, UK*, Sept. 2009.

[3] J. Rouas, I. Trancoso, C. Viana, and M. Abreu, "Language and variety verification on broadcast news for portuguese," *Speech Commun.*, vol. 50, no. 11-12, pp. 965–979, 2008.

[4] Fabio Castaldo, Daniele Colibro, Sandro Cumani, Emanuele Dalmasso, Pietro Laface, and Claudio Vair, "Loquendo-Politecnico di torino system for the 2009 nist language recognition evaluation," *ICASSP 2010*, 2010.

[5] M. A Zissman and K. M Berkling, "Automatic language identification," *Speech Communication*, vol. 35, no. 1-2, pp. 115–124, 2001.

[6] Raymond W. M. Ng, Cheung-Chi Leung, Tan Lee, and Bin Ma, "Prosodic attribute model for spoken language identification," *ICASSP 2010*, 2010.

[7] M.A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech," *Speech and Audio Processing, IEEE Transactions on*, vol. 4, no. 1, pp. 31, 1996.

[8] H. Meinedo, M. Viveiros, and J. Neto, "Evaluation of a live broadcast news subtitling system for portuguese," in *Proc. Interspeech*, 2008.

[9] A. Abad and J. Neto, "Incorporating acoustical modelling of phone transitions in an hybrid ANN/HMM speech recognizer," *Proc. INTERSPEECH-08, Brisbane, Australia*, 2008.

[10] A. Stolcke, "SRILM-an extensible language modeling toolkit," in *Seventh International Conference on Spoken Language Processing*, 2002, vol. 3.

[11] N. Brümmer, "FoCal multi-class: Toolkit for evaluation, fusion and calibration of multi-class recognition Scores—Tutorial and user manual—," 2007.

[12] K. Berkling, T. Arai, and E. Barnard, "Analysis of Phoneme-Based features for language identification," *PROC ICASSP*, vol. 1, pp. 289—292, 1994.

[13] H. Meinedo, D. Caseiro, J. Neto, and I. Trancoso, "AUDIMUS.MEDIA : A broadcast news speech recognition system for the european portuguese language," in *Computational Processing of the Portuguese Language*, p. 196. 2003.

[14] A. Montero-Asenjo, D. T. Toledano, J. Gonzalez-Dominguez, J. Gonzalez-Rodriguez, and J. Ortega-Garcia, "Exploring PPRLM performance for nist 2005 language recognition evaluation," in *IEEE Odyssey 2006: The Speaker and Language Recognition Workshop, 2006*, 2006, p. 1–6.