



Language Recognition using Phonotactic-based Shifted Delta Coefficients and Multiple Phone Recognizers

Luis Fernando D'Haro¹, Ricardo Cordoba¹, Christian Salamea^{1,2}, Javier Ferreiros¹

¹Speech Technology Group. Dpto. de Ing. Electrónica. Universidad Politécnica de Madrid. Ciudad Universitaria s/n 28040, Madrid, Spain

²Universidad Politécnica Salesiana del Ecuador
Calle Vieja 12-30 y Elia Liut, Casilla 26, Cuenca, Ecuador
{lfdharo, cordoba, csalamea, jfl}@die.upm.es

Abstract

A new language recognition technique based on the application of the philosophy of the Shifted Delta Coefficients (SDC) to phone log-likelihood ratio features (PLLR) is described. The new methodology allows the incorporation of long-span phonetic information at a frame-by-frame level while dealing with the temporal length of each phone unit. The proposed features are used to train an i-vector based system and tested on the Albayzin LRE 2012 dataset. The results show a relative improvement of 33.3% in C_{avg} in comparison with different state-of-the-art acoustic i-vector based systems. On the other hand, the integration of parallel phone ASR systems where each one is used to generate multiple PLLR coefficients which are stacked together and then projected into a reduced dimension are also presented. Finally, the paper shows how the incorporation of state information from the phone ASR contributes to provide additional improvements and how the fusion with the other acoustic and phonotactic systems provides an important improvement of 25.8% over the system presented during the competition.

Index Terms: language recognition, SDC, Phone-Log Likelihood Ratios, parallel phone recognizers

1. Introduction

In this paper, we will describe new enhancements done to the final system we presented to the Albayzin LRE 2012 evaluation [1]. The main change has been the incorporation of a new kind of phonotactic subsystem that uses Phone Log-Likelihood Ratios features (PLLR) which has proved to improve both language [2] and speaker recognition systems [3]. Later, in [4] we described how these features can be extended to provide a better performance thanks to the use of likelihood ratios at a phone state level, instead of a phone level; In this paper we will show how the incorporation of long-span phonotactic information to new coefficients based on the Shifted Delta Cepstra (SDC) philosophy [5], which we called Shifted Delta PLLR Coefficients (SDPC) contributes to improve the system accuracy and it is compatible with the use of state-based information. Moreover, inspired on PPRM systems [6][7], we will show results incorporating multiple phone ASR on these SDPC coefficients. Finally, since in both techniques the resulting input vector has a high dimensionality we have applied a dimensionality reduction technique to deal with the course of dimensionality problem. The paper is organized as follows: Section 2 describes the evaluation database and each subsystem included in the baseline system, while section 3 describes the proposed extensions to the phonotactic system. Sections 4 and 5 show the fusion results, the conclusions and future work respectively.

2. Baseline System and Data Description

For the experiments, we have used only the database provided during the Albayzin LRE 2012 evaluation [11]. The audio used for this evaluation was very challenging since it was collected from YouTube videos, with different length durations, channel conditions, number of speakers, and several kinds of noises. All audio files were sampled at 16 KHz. During the evaluation, four different conditions were proposed depending on the languages to recognize, the availability of training data (plenty or empty), and the possibility of recognizing out-of-set languages or not (open vs closed). This paper shows results only on the main condition (plenty-closed), where the target languages were: Spanish, Catalan, Basque, Galician, Portuguese, and English; the total number of files for the train set was 5115, 458 for dev., and 941 for eval. During the evaluation two metrics - F_{act} and C_{avg} - were used; we will provide results on both for comparison with the evaluation results. The F_{act} metric is useful since it provides information about how discriminative a system is and how well calibrated it is.

Next, we will describe each of the subsystems which were used and fused for creating the final system presented in this paper and used for the Albayzin 2012 LRE evaluation [1]. For the evaluation, our primary system consisted mainly in the fusion of three different i-vector based systems: 1) an acoustic system trained on MFCC-SDC features, 2) an acoustic system based on RPLP-SDC features, and 3) a phonotactic system trained on trigram posteriorgram counts. In [8] it was shown that one of the main advantages of our system was the use of RPLP (Revised PLP) features which allowed the incorporation of noise-robust features, and the use of a phonotactic i-vector based system trained on non-sparse n-gram counts estimated using the posterior probabilities output of a phoneme recognizer and the i-vectors trained using subspace multinomial models [9]. Finally, the fusion of the scores of these three different sub-systems allowed the integration of various levels of perceptual cues providing the best results as it is demonstrated in [10].

2.1. Acoustic Systems

2.1.1. MFCC-SDC i-vector system

For this subsystem, for each audio file we extracted 12 MFCC coefficients (including C0) from 24 Mel filter banks plus the energy for each frame. As Voice Activity Detector (VAD) we used the output from the BUT Hungarian phone recognizer [16] suppressing all segments marked as silence or noise in the output. Then, a RASTA filter was applied to reduce noise variations in each frequency sub-band followed by a short-term

Cepstral Mean and Variance Normalization (CMNV) normalization. After that, every 10 ms speech frame was mapped to a 56-dimensional feature vector generated from the concatenation of SDC features using the 7-1-3-7 configuration. Then, as in [4], we included feature warping [12] after removing all non-speech frames using the toolkit available at [13] and training i-vectors of 400 dimensions and 512 Gaussians which we found as optimum configuration. For this system, the use of the short-term CMNV and feature warping allowed us to improve the C_{avg} and F_{act} in 10.4% and 7.2% relative respectively in comparison with the evaluation system.

2.1.2. RPLP-SDC i-vector System

Proposed in [14] and [15], the Revised PLP (RPLP) features can be seen as a hybrid approach between the MFCC and PLP features, combining the best of both and providing as result noise-robust features. In [8] we showed that these features highly contributed to improve the final system and performed better than the MFCC subsystem in spite of using the same configuration in terms of number of Gaussians, i-vector dimension, SDC configuration, etc. As for the MFCC, we also tried the use of short-term CMNV and Feature Warping, but they did not provide any improvement. Therefore we kept the same configuration used for the evaluation.

2.2. Phonotactic systems

For this section, the phonotactic information was provided using the phone decoders developed by the Brno University of Technology (BUT). The advantage of these phone decoders is that they use a three-state model per phone, which means that three posterior probabilities per unit are given at each frame.

2.2.1. Posterior-gram based i-vector System

Proposed in [9], this subsystem uses posteriorgram counts extracted from the posterior probabilities extracted from the phone recognizer instead of the traditional soft-counts extracted from phoneme lattices. In comparison with other techniques based on soft-counts, the new features make use of the frame-by-frame posterior probabilities provided by the phone recognizer and do not lack of sparse counts, and avoid the use of the pruning techniques used when creating lattices. On the other hand, [17] and [18] describe different systems and algorithms for LID where the posterior probabilities information has been also successfully used. In our approach, the high-dimensional vectors of counts are reduced to low-dimensional units for which we adapted the commonly used technique i-vectors. The reduction is based on subspace multinomial modeling (SMM, [19] and [20]) and is designed to work in the total-variability space. For training the i-vectors, we followed the algorithm reported in [21] with slight modifications in the EM step. Finally, for our experiments, we considered a set of 1089 multinomial models when using trigrams counts generated from the output of the Hungarian recognizer.

2.2.2. Phone Log-Likelihood Ratios i-vector System

In [2] and [3] it was shown that the Phone Log-Likelihood Ratio (PLL) features can be successfully used for language and speaker recognition tasks. Its success is probably due to the simplicity of its calculation and because they can be easily integrated with the i-vector framework where the PLL can be seen as an alternative to the acoustic MFCC-SDC features. On

the other hand, as proved in [3], other alternative features as frame-level posteriors or phone log-posteriors (which are usually provided by phone recognizers) are not suitable for tasks where the features are assumed to be Gaussian-distributed. In contrast, the transformation from log posteriors into log-likelihood ratios (LLR) provides final distributions that are nearly Gaussian. In order to calculate the PLLR features [22], the acoustic posterior probability of a phone unit m at each frame f , is calculated by summing up the posteriors of its corresponding states. Then, the log-likelihood ratios at each frame f can be computed from posterior probabilities using equation (1) where it is assumed a classification task with flat priors.

$$LLR_f^m = \log \frac{p(x_f / m)}{\frac{1}{M-1} \sum_{\forall n \neq m} p(x_f / n)} \quad m = 1, \dots, M \quad (1)$$

Finally, the resulting M log-likelihood ratios per frame are stacked together to create the Phone Log-Likelihood Ratio (PLL) features. For our system, these features were created using the open-source toolkit available in [23]. After that, an i-vector system was trained on the PLLR features.

2.3. Classifier and calibration back-end

As classifier in all subsystems, we used a Multiclass logistic regression, and for calibration and fusion, a Gaussian Back-end followed by a Discriminative Multi-Class Logistic Regression. Previously, the input i-vectors were conditioned by within-class covariance normalization [24] and length normalized.

3. Proposed Systems

3.1. Baseline PLLR features: phone-PLL

As in [22], before computing the PLLR features, the three non-phonetic units of the BUT phone recognizers, i.e.: int, pau, and spk, were fused into a single non-phonetic unit. Then, a unified posterior probability is computed for each phone model by adding the posterior probabilities of all the states in the corresponding phone model. Finally, the log-likelihood ratios were computed using (1). This way, for the Hungarian phone recognizer we have 59 PLLR features, 50 for Russian, and 43 for Czech. Here, the use of first order deltas provided us a relative improvement of 3.4% in C_{avg} and the use of different kind of phone mappings did not provide improvements for any of the phone recognizers. Therefore, our baselines are given using the complete phone set for all the recognizers, including the delta features, using i-vectors of 400 dimensions, and UBMs with 512 Gaussians.

3.2. State-based PLLR

One important modification we proposed in [4] for the baseline PLLR features was to use the likelihood ratio of each individual state as a feature instead of summing up the posteriors probabilities of the corresponding phone-states. Our motivation was to take advantage of the information encoded in the transitions between phones as well as between states, which also provides discriminative information between languages. Thanks to these state-based PLLR we found significant improvements over phone-PLL in all cases (between 9.6% and 14.3% relative in C_{avg}). The caveat is the dimensionality problem: since each phone has three states per phone, the final PLLR vector for each frame is of dimension

177 for the Hungarian phone recognizer, of 129 for the Czech, and 150 for the Russian decoder. We dealt with this problem using dimensionality reduction techniques such as PCA, LDA or HLDA, being PCA the one that provided the best results.

3.3. Extending the context with SDPC Coefficients

To compensate the potential drawbacks resulting from using the short-term phone information from the PLLR features and to extend the context of the parameters we first considered applying the first order deltas. As an alternative we propose to use what we call Shifted Delta PLLR Coefficients (SDPC) which borrows the windowing concepts of SDC parameters for the acoustic vectors to capture longer-term statistics, but applied to our PLLR vectors. One drawback of this technique is that the final dimensionality of the vectors is too high. For example, for the Hungarian recognizer and one of our optimal SDC configurations in most cases: using $d=1$ (time advance and delay for the delta computation), $P=5$ (time shift between consecutive blocks) and $k=3$ (number of blocks) will result in a vector of dimension $177 \times 4 = 708$ features. Therefore, applying dimensionality reduction is a must. In our case, we used PCA as it provided the best results.

3.4. Handling multiple recognizers

Inspired by the good results obtained on LID when using parallel phone recognizers in comparison with the use of a single phone recognizer, we considered several options in order to handle the 3 phone recognizers that we have available.

3.4.1. Independent Phone Recognizers (Indep-ASR)

Our first approach was to process the 3 recognizers as independent modules and fuse them with the Gaussian Backend. Table 1 shows the obtained results we published in [4]. The relative improvements in parenthesis are comparisons with the best results without using SDPC. For phone-based PLLR, the optimum results are obtained using a PCA projection of dimension 30 and then SDPC with parameters $d=1$, $P=5$, $k=5$, with a final vector of dimension 180. For state-based PLLR, the optimum was a PCA of dimension 35 with parameters $d=1$, $P=5$, $k=3$, resulting in a final vector of dimension 140. Unfortunately, this approach is very time-consuming and requires almost 3 times more space.

3.4.2. Multiple Phone Recognizers Stacked (Multi-ASR)

Our second approach was to stack the output of the three different phone recognizers and apply different merging approaches followed by a PCA reduction. Here, first we tried to merge the PLLR features from all recognizers into a single feature vector and apply a common PCA projection. This way we have a single PLLR system and the CPU and memory requirements are divided by 3 approximately (depending on the final dimensions of the feature vectors). The second idea was to apply a different PCA projection over the output of each phone recognizer and then stack the resulting projected vectors into one. Different experiments, not included in the paper, showed that the first idea provided the best results. Therefore we will use this approach from now on. Table 1 shows the performance over the last approach comparing the use of information at phone or state level, and comparing the use of SDPC or deltas. The relative improvement of this system over using just one recognizer is between 5.3 and 5.6%. In the table we show the optimum dimensions. In this case, for PCA + first

order deltas the optimum is 55 for Phone-based and 90 for State-based, with a total dimension of 110 and 180 respectively. Here the state approach obtains better results since it can take advantage of more information. On the other hand, when using the SDPC features, a similar effect occurs for the state-based system where the optimum configuration is PCA= 70, SDPC=1-6-3 (total dimension 280); and for the phone-based is PCA= 60, SDPC=1-5-3 (total dimension 240). Again, we can see that the state-based approach is consistently better than the phone-based approach, and the multi-ASR approach provides better results than any of the single ASR.

	Single ASR			Multi-ASR	
	Hu	Cz	Ru	PCA + Δ s	PCA+SDPC
Phone-based	6.74	7.86	7.11	7.04	6.38
	14.15	15.16	14.90	14.45	13.41
State-based	6.57	7.33	7.00	6.66	6.22
	13.97	14.70	14.75	14.04	13.19

Table 1. C_{avg} (top) and F_{act} (down) evaluation results using Delta or SDPC over independent or multiple phone ASR

3.5. Phonotactic-Based SDPC Coefficients (P-SDPC)

In this section we want to describe a different approach for the Shifted Delta PLLR Coefficients (SDPC) consisting of modifying the P value (time shift between consecutive blocks) according to the phone boundaries provided by the recognizer segmentations. There are two reasons for this: 1) to cover a wider context, and 2) to model the differences in phone log-likelihoods between adjacent phonemes. These differences are not taken into account on a frame-by-frame analysis where the PLLR values have slow changes, while between phone boundaries these changes are higher and could contain additional information. Therefore, we can consider this approach as a way to model something similar to a language model of log-likelihoods.

Figure 1 shows an example of the technique using different colors for the frames assigned to each state or phone. The procedure for calculating the P-SDPC for the current frame is to select the past and future frames considering the phone boundaries and the relative position of the current frame to the phone it belongs to. The figure shows how for the current frame (number 3 in blue, which is the central frame for the current phone) the system selects as relative frames in the adjacent phones the ones numbered as 5 and 8, and for the next P-SDPC coefficient the frames 3, 4, and 6.

We have considered two options for the units used for the block shift in this technique: the state boundaries and the phone boundaries (see Table 2). In the table, the values in parenthesis are the relative improvements w.r.t. the baseline. Here the baseline uses the Hungarian recognizer with the phone-based PLLR approach + PCA + SDPC. We can see that the phone boundaries provide similar results than the baseline, but the state boundaries improve slightly the results. The optimum configurations were $d=1$, $P=2$, $k=2$ for phone boundaries (clearly we have to use a low number of blocks because probably the shift is too long) and $d=1$, $P=3$, $k=3$ for state boundaries (now we can use more blocks), both using PCA=60. We can see that the results are slightly better than the baseline. Similar improvements are obtained using the Czech and Russian recognizers. Also, when fusing the 3 phone recognizers, the P-SDPC system obtains 5.89 in F_{act} and 4.21 in C_{avg} improving the baseline SDPC system in **8.0%** and **5.2%** respectively.

	Hungarian	
	$C_{avg}(\mathbf{Im})$	$F_{act}(\mathbf{Im})$
Baseline	6.74	14.15
Baseline + P-SDPC phone boundaries	6.72 (0.3)	14.25 (-0.7)
Baseline + P-SDPC state boundaries	6.60 (2.1)	13.73 (3.0)

Table 2. P-SDPC evaluation results for the Hung. ASR

4. Fusion Results

In Table 3 we can see a summary of results for the individual systems (first five lines) and the results fusing all our systems. All improvements from now on will be relative in the Fact metric. We can see that the PLLR features are even better than the acoustic systems, a 10% improvement with the states version. Besides, using only phonotactic information, i.e. the phonotactic subsystem + our PLLR proposal we get a better result than the acoustic + phonotactic system we presented for the Albayzin competition (6.3% improvement). Adding the PLLR to the acoustic systems provides 33.3% improvement (from 7.77 to 5.18), which is especially relevant. Finally, fusing all systems we get a reduction of 28.5% over the system presented to the competition.

Comparing the results between the approaches Multi-ASR and Indep-ASR we can see that they are similar, with a small improvement of the Multi-ASR (4.98 versus 5.18) when all modules are combined, but the advantage of the Multi-ASR approach is the reduction in memory and CPU requirements.

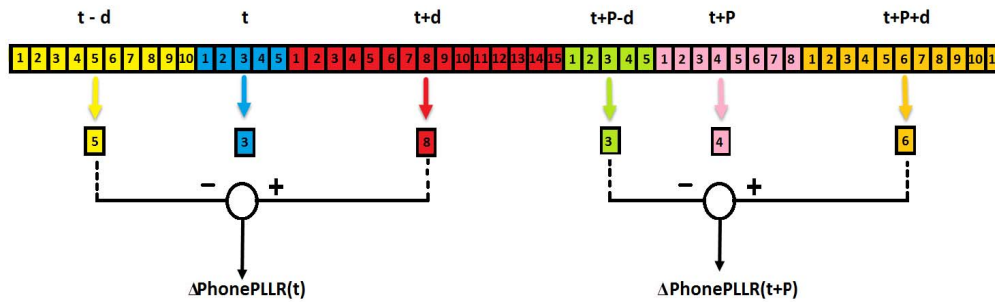


Figure 1. Example of the proposed Phonotactic SDPC features

Fusion	MFCC	RPLP	Posterior-gram	Independent PLLR-SDPC	Independent PLLR-P-SDPC	Multi-ASR State	Dev		Eval	
							C_{avg}	F_{act}	C_{avg}	F_{act}
MFCC-SDC							6.50	12.24	6.95	14.68
RPLP-SDC							6.54	12.34	7.36	14.73
Phonotactic							6.94	13.37	9.85	18.14
Multi-ASR phone							4.43	10.58	6.38	13.41
Multi-ASR states							4.57	10.35	6.22	13.19
Acoustic only	X	X					4.83	6.09	5.39	7.77
Competition	X	X	X				2.88	3.89	4.85	6.48
Phonotactics only + Indep. ASR			X	X			2.36	3.23	4.13	6.07
Phonotactics only + Multi-ASR			X			X	2.44	3.28	4.45	6.31
Acoustic + Indep. ASR	X	X		X			2.15	2.83	3.66	5.29
Acoustic + Multi-ASR	X	X				X	2.31	3.09	3.89	5.18
Acoustic + Indep. ASR P-SDPC	X	X			X		2.70	3.28	3.36	4.63
All + Indep. ASR	X	X	X	X			2.25	2.64	3.52	5.18
All + Multi-ASR	X	X	X			X	1.99	2.74	3.64	4.98
All + Multi-ASR + Indep P-SDPC	X	X	X		X	X	2.28	2.91	3.45	4.76
All	X	X	X	X		X	2.01	2.66	3.70	4.81

Table 3. Individual and fusion results for the different subsystems on the dev and evaluation sets

5. Conclusions and Future Work

We have described different improvements to a language recognition system. The main change is the incorporation of a phonotactic system based on the use of state-based phone log likelihood ratios, PCA as dimensionality reduction technique and SDPC features to increase the context. We compared two strategies: using independent phone recognizers and using the combined output from several recognizers (Multi-ASR). The second option is clearly better as results are similar or slightly better and the CPU and memory requirements are divided almost by 3. We also presented new coefficients, Phonotactic-Based SDPC Coefficients (P-SDPC), where the time shift between consecutive blocks is computed according to the phone boundaries provided by the recognizers. This approach provides additional improvements between 5.2 and 8% and is part of our best system in the fusion with a 4.63 in F_{act} and 3.36 in C_{avg} . In future work we will continue improving and adjusting the optimum topology for them. The overall results are impressive, as the inclusion of the best PLLR system (4.63 in F_{act}) provides a 40.4% improvement to the acoustic system using MFCC plus RPLP (7.77 in F_{act}).

6. Acknowledgements

This work has been supported by TIMPANO (TIN2011-28169-C05-03), INAPRA (MICINN, DPI2010-21247-C02-02), MA2VICMR (Comunidad Autónoma de Madrid, S2009/TIC-1542) and the European Simple4All (grant agreement 287678).

7. References

- [1] L. F. D'Haro, R. Córdoba. 2012. "The GTH-LID System for the Albayzin LRE12 Evaluation". In Proc. Iberspeech 2012, pp. 528-539.
- [2] M. Diez, A. Varona, M. Penagarikano, Luis J. Rodriguez-Fuentes, G. Bordel. 2012. "On the use of Phone Log-Likelihood Ratios as Features in Spoken Language Recognition", IEEE Workshop on Spoken Language Technology (SLT); Miami, Florida, USA.
- [3] M. Diez, Luis J. Rodriguez-Fuentes, M. Penagarikano, A. Varona, G. Bordel. 2013. "Using Phone Log-Likelihood Ratios as Features for Speaker Recognition", Interspeech 2013; Lyon, France.
- [4] L. F. D'Haro, R. Cordoba, C. Salamea, J. D. Echeverry. 2014. "Extended Phone Log-Likelihood Ratio Features and Acoustic-based I-Vectors for Language Recognition". To appear in ICASSP 2014, Florence, Italy. May 4-9, 2014.
- [5] P.A. Torres-Carrasquillo, E. Singer, M.A. Kohler, R.J. Greene, D.A. Reynolds, and J.R. Deller Jr., 2002. "Approaches to language identification using Gaussian mixture models and shifted delta cepstral feature", in Proc. International Conferences on Spoken Language Processing (ICSLP), Sept. 2002, pp. 89-92.
- [6] M. A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech," IEEE Transactions on Speech and Audio Processing, vol. 4, 1996.
- [7] M. Penagarikano, A. Varona, L. J. Rodriguez-Fuentes, G. Bordel, 2011. "Improved Modeling of Cross-Decoder Phone Co-Occurrences in SVM-Based Phonotactic Language Recognition", IEEE Transactions on Audio Speech and Language Processing 19 (8), 2348.
- [8] L. F. D'Haro, R. Cordoba, M. A. Caraballo, J. M. Pardo. 2013. "Low-Resource Language Recognition using a Fusion of Phoneme Posteriorgram Counts, Acoustic and Glottal-based I-Vectors". ICASSP 2013, Vancouver, Canada. May 26-31, 2013.
- [9] L. F. D'Haro, O. Glembek, O. Plchot, P. Matejka, M. Souffar, R. Cordoba, J. Cernocký. "Phonotactic Language Recognition using i-vectors and Phoneme Posteriorgram Counts", in Proc. of Interspeech 2012.
- [10] H. Li; B. Ma; K A. Lee, "Spoken Language Recognition: From Fundamentals to Practice," Proceedings of the IEEE, vol. 101, no.5, pp. 1136-1159, May 2013.
- [11] L. J. Rodriguez-Fuentes, N. Brümmer, M. Penagarikano, A. Varona, G. Bordel, M. Diez, 2013. "The Albayzin 2012 Language Recognition Evaluation", Interspeech 2013; Lyon, France.
- [12] J. Pelecanos and S. Sridharan, 2001. "Feature warping for robust speaker verification", in Proc. ISCA Odyssey, Crete, Greece, Jun. 2001.
- [13] S. O. Sadjadi, M. Slaney, and L. Heck. 2013. "MSR Identity Toolbox vs 1.0", Microsoft Research. [Online] Available at <http://research.microsoft.com/>
- [14] J. Rajnoha, and P. Pollák. 2011. "ASR systems in Noisy Environment: Analysis and Solutions for Increasing Noise Robustness". Radio engineering, Vol. 20, No. 1, April 2011, pp. 74-84.
- [15] F. Hönl, G. Stemmer, C. Hacker, and F. Brugnara. 2005. "Revising Perceptual Linear Prediction (PLP)". In Eurospeech 2005, p. 2997-3000.
- [16] P. Schwarz, 2009. "Phoneme Recognition based on Long Temporal Context, PhD Thesis", Brno University of Technology, 2009. [Online] Available at <http://speech.fit.vutbr.cz/software/phoneme-recognizer-based-long-temporal-context>
- [17] K. Han and J. Pelecanos. 2012. "Frame-based phonotactic Language Identification". IEE Spoken Language Technology Workshop (SLT), 2012, pp. 303-306.
- [18] Leung, C.-C., Lee, T., Ma, B., Li, H. 2013. "Shifted-Delta MLP Features for Spoken Language Recognition", IEEE Signal Processing Letters, Vol. 20, Issue 1, pp. 15-18.
- [19] D. Povey, Lukas Burget et. al, 2011. "The Subspace Gaussian Mixture Model- a Structured Model for Speech Recognition", Computer Speech and Language, 25(2), pp. 404-439.
- [20] D. Martínez, O. Plchot, L. Burget, O. Glembek, and P. Matejka, "Language Recognition in iVectors Space", in Proc. of Interspeech 2011.
- [21] Kockmann, et al, 2010. "Prosodic speaker verification using subspace multinomial models with intersession compensation," in Proc. of Interspeech, Makuhari, Chiba, Japan, 2010.
- [22] M. Diez, Luis J. Rodriguez-Fuentes, M. Penagarikano, A. Varona, G. Bordel, 2013. "Dimensionality Reduction of Phone Log-Likelihood Ratio Features for Spoken Language Recognition", Interspeech 2013; Lyon, France, Aug. 25-29, 2013;
- [23] PLLR computation software. [Online]. Freely available at: <https://sites.google.com/site/gttsplrfeatures/home>
- [24] A. Hatch and A. Stolcke, 2006. "Generalized linear kernels for one-versus-all classification: application to speaker recognition", in Proc. ICASSP Vol 5, page V.