GRAPHEME AND MULTILINGUAL POSTERIOR FEATURES FOR UNDER-RESOURCED SPEECH RECOGNITION: A STUDY ON SCOTTISH GAELIC

Ramya Rasipuram[1,2]*, Peter Bell*[3] *and Mathew Magimai.-Doss*[1]

[1] Idiap Research Institute, CH-1920 Martigny, Switzerland
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[3] Centre for Speech Technology Research, University of Edinburgh, Edinburgh EH8 9AB, UK
{ramya.rasipuram,mathew}@idiap.ch, peter.bell@ed.ac.uk

ABSTRACT

Standard automatic speech recognition (ASR) systems use phonemes as subword units. Thus, one of the primary resource required to build a good ASR system is a well developed phoneme pronunciation lexicon. However, under-resourced languages typically lack such lexical resources. In this paper, we investigate recently proposed grapheme-based ASR in the framework of Kullback-Leibler divergence based hidden Markov model (KL-HMM) for under-resourced languages, particularly Scottish Gaelic which has no lexical resources. More specifically, we study the use of grapheme and multilingual phoneme class conditional probabilities (posterior features) as feature observations in KL-HMM. ASR studies conducted show that the proposed approach yields better system compared to the conventional HMM/GMM approach using cepstral features. Furthermore, grapheme posterior features estimated using both auxiliary data and Gaelic data yield the best system.

Index Terms— Automatic speech recognition, Kullback-Leibler divergence based hidden Markov model, grapheme, phoneme, posterior feature, under-resourced speech recognition, Scottish Gaelic

1. INTRODUCTION

Recently, there is a growing interest to use graphemes as subword units for speech recognition [1, 2], [3, Chapter 4], [4, 5], especially for under-resourced languages where well developed phoneme sets and phoneme pronunciation dictionaries are usually not available [6, 7, 8, 9, 10]. Under-resourced languages also typically lack acoustic resources. Therefore, research in this domain has focussed on the efficient development of multilingual and crosslingual grapheme-based ASR approaches that can leverage from resources available in other languages. In [7, 6], the use of multilingual grapheme models for rapid bootstrapping of acoustic models to new languages was studied. In [7], polyphone decision tree based tying for porting decision tree to a new language was applied for grapheme models. More specifically, porting of multilingual grapheme models to German was studied and was found to be beneficial compared to monolingual grapheme models when limited adaptation data is available. In [6], data driven mapping of grapheme subword units across languages was studied when the alphabet set between the source and target languages was disjunct. In [8], grapheme-based acoustic modeling was investigated and compared with phoneme-based acoustic modeling for under-resourced language Vietnamese. For context-independent grapheme modeling, word boundary detection based initialization of grapheme acoustic models was proposed. However, it was found that grapheme-based system could not reach the performance of phoneme based system.

A novel grapheme-based ASR in the framework of Kullback-Leibler divergence based HMM [11], which jointly models grapheme and phoneme information, was recently proposed in [5]. More specifically, in this approach, first the relationship between the acoustic feature (e.g., cepstral features) and phoneme is modeled through a posterior feature (phoneme class conditional probabilities) estimator (more precisely, multilayer perceptron). Then, soft/probabilistic correspondence between phonemes and graphemes is modeled/learned through the state multinomial/categorical distribution of KL-HMM system.

In this paper, we investigate the potential of the approach for low-resourced and minority languages, in particular Scottish Gaelic, where language resources are very sparse. With only 60,000 speakers, Gaelic represents a genuinely under-resourced minority language, and its endangered status makes low-cost speech technology particularly important for language conservation efforts.

We study two novel approaches. The first approach exploits the auxiliary acoustic and phonetic resources available in other languages to develop a grapheme-based ASR system for Scottish Gaelic. More specifically, the KL-HMM system models the relation between Gaelic graphemes and multilingual phonemes by using multilingual phoneme posterior features (extracted by MLP trained on auxiliary data) as feature observation. The second approach exploits the flexibility to choose the posterior feature space representation in KL-HMM system. More precisely, we investigate the use of grapheme posterior features as feature observations. In that regard, we investigate two different ways to estimate grapheme posterior features: 1) using an MLP trained on Scottish Gaelic corpus and 2) using hierarchical MLP [12] trained on both auxiliary resources and Scottish Gaelic corpus. We study these KL-HMM based approaches along with the traditional HMM/GMM approach (using cepstral features). ASR studies show that KL-HMM systems outperform the HMM/GMM system and the KL-HMM system modeling grapheme posterior features yields the best performance.

The rest of the paper is organized as follows: Section 2 presents a brief overview of grapheme-based ASR using KL-HMM and motivates the two approaches. Section 3 presents an overview of the Scottish Gaelic with emphasis on alphabet and orthography. Section 4 presents the corpus, experimental setup and ASR results.

This work was partly supported by the Swiss NSF through the grants Flexible Grapheme-Based Automatic Speech Recognition (FlexASR) and the National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (www.im2.ch). Gaelic data collection was funded through the University of Edinburgh's iDEA Lab programme.

2. GRAPHEME-BASED ASR USING KL-HMM

Kullback-Leibler divergence based HMM (KL-HMM) is a recently proposed approach where a posteriori probabilities of phonemes are directly used as feature observation [11]. In grapheme-based ASR using KL-HMM, the HMM states represent grapheme subword units and the feature observations are *posterior features* i.e., a posteriori probabilities of phonemes [5]. Let $\mathbf{z}_t$ denote the phoneme posterior feature vector estimated at time frame t,

$$\begin{aligned}\mathbf{z}_t &= [z_t^1, \cdots, z_t^d, \cdots, z_t^D]^T \\ &= [P(p_1|\mathbf{x}_t), \cdots, P(p_d|\mathbf{x}_t), \cdots, P(p_D|\mathbf{x}_t)]^T\end{aligned}$$

where $\mathbf{x}_t$ is the acoustic feature (e.g., cepstral feature) at time frame t, $\{p_1, \cdots p_d, \cdots p_D\}$ is the phoneme set, D is the number of phonemes, and $P(p_d|\mathbf{x}_t)$ denotes the a posteriori probability of phoneme p_d given $\mathbf{x}_t$. In this work $\mathbf{z}_t$ is estimated by a well trained MLP.

Each HMM state i in the KL-HMM system is parameterized by a categorical distribution $\mathbf{y}_i = [y_i^1, \cdots, y_i^d, \cdots, y_i^D]^{\mathrm{T}}$. The local score at each HMM state is estimated as Kullback-Leibler (KL) divergence between $\mathbf{y}_i$ and $\mathbf{z}_t$, i.e.,

$$KL(\mathbf{y}_i, \mathbf{z}_t) = \sum_{d=1}^{D} y_i^d \log(\frac{y_i^d}{z_t^d}) \tag{1}$$

In this case, $\mathbf{y}_i$ serves as the reference distribution and $\mathbf{z}_t$ serves as the test distribution. KL-divergence being an asymmetric measure, there are also other ways to estimate the local score,

1. Reverse KL-divergence (RKL):

$$RKL(\mathbf{z}_t, \mathbf{y}_i) = \sum_{d=1}^{D} z_t^d \log(\frac{z_t^d}{y_i^d}) \tag{2}$$

2. Symmetric KL-divergence (SKL):

$$SKL(\mathbf{y}_i, \mathbf{z}_t) = KL(\mathbf{y}_i, \mathbf{z}_t) + RKL(\mathbf{z}_t, \mathbf{y}_i) \tag{3}$$

The HMM state parameters i.e., categorical distributions are estimated using Viterbi expectation maximization algorithm which minimizes a cost function based on one of the above local scores. During testing, decoding is performed using standard Viterbi decoder [11].

Until now, this approach has been studied on English, where the correspondence between graphemes and phonemes is weak/irregular [5, 13, 14]. These studies have revealed that,

- grapheme-based ASR system can yield performance similar to phoneme-based ASR system. More precisely, the proposed grapheme-based ASR approach can exploit the low complexity of KL-HMM to model longer subword contexts, which in turn helps to bridge the performance gap between grapheme and phoneme based ASR systems.
- the MLP can be trained on auxiliary data. In [5], it was found that the system benefits from MLP trained on large amount of out-of-domain data. In [13], where the aim was to recognize multi-accent non-native speech utterances with limited or no training data, it was found that the system using multilingual posterior features estimated using an MLP trained on multiple languages yields a better system compared to the system using monolingual posterior features.
- The categorical distribution of the HMM states capture probabilistic relationship between the graphemes and the phonemes [5]. This relationship can be exploited for grapheme-to-phoneme conversion [14].

In this paper, we investigate the approach on under-resourced minority language, particularly, Scottish Gaelic where no lexical resources (i.e., phoneme set and phoneme pronunciation lexicon) are available. More specifically, in this work we exploit the flexibility of KL-HMM in terms of choice of posterior feature space representation and transfer learning to study three different posterior features, namely,

1. multilingual posterior features: In this case, the aim is to study how acoustic and lexical resources available in other languages could be used to improve Scottish Gaelic ASR. More specifically, multilingual posterior features estimated by an MLP trained on auxiliary multilingual corpus are used as feature observations. The states of KL-HMM capture the relation between graphemes and multilingual phonemes.
2. grapheme posterior features: Grapheme-to-phoneme relationship in Scottish Gaelic is many-to-one, and is fairly regular (see Section 3). So it may be possible to train an MLP directly on Scottish Gaelic corpus with standard acoustic features (which typically depict phoneme characteristics) as input t classify graphemes. And, use the grapheme posterior feature estimates from the output of the MLP as feature observation.
3. hierarchical grapheme posterior features: In English, the relation between context-independent graphemes and phonemes is irregular. However, in our previous studies, analysis of the context-dependent grapheme models revealed that by modeling grapheme context the relationship becomes more regular and one-to-one [5]. Similarly, it may be possible to learn phoneme-to-grapheme relationship by modeling phoneme context. Such an approach coupled with acoustics could help in estimating better grapheme posterior features, while leveraging from auxiliary data. More precisely, this could be achieved using hierarchical MLP approach [12], where the first MLP is trained on auxiliary corpus to estimate multilingual phoneme posterior features, and the second MLP then uses the multilingual posterior features with longer temporal context as input to estimate grapheme posterior features.

3. SCOTTISH GAELIC

Scottish Gaelic is one of three primary Goidelic languages. Classified within the Indo-European language family, it is contained within the group of Celtic languages, and as such is only distantly related to any of the well-resourced major European languages. Scottish Gaelic is derived from and is closely related to Irish Gaelic; it is considered an endangered language, spoken by only around 60,000 speakers, mainly from the remote islands of Scotland. In this section we first briefly describe the alphabet, orthography, grapheme-to-phoneme relationship of Scottish Gaelic.

3.1. Language Characteristics

The Scottish Gaelic alphabet has 18 graphemes (A, B, C, D, E, F, G, H, I, L, M, N, O, P, R, S, T, U) and long vowels are marked with grave accents (À, È, Ì, Ò, Ù). The number of phonemes in Scottish Gaelic are approximately 51 (9 vowels, 10 dipthongs and 32 consonants) [15]. However, number of phonemes can vary depending on the dialect. The language lacks proper speech and linguistic resources (phoneme set and pronunciation lexicon).

3.2. Orthography

The number of graphemes in Gaelic words are usually significantly greater than the number of phonemes in the word, for two primary reasons: Firstly, in Gaelic, consonants are either broad (velarized) or slender (palatalized). Broad consonants are surrounded by broad vowels A, O or U on both sides and slender consonants are surrounded by slender vowels I or E on both sides. This has the consequence that many vowels are present in orthography only to denote the broad or slender nature of consonant next to it. Secondly, consonants of Gaelic words may be changed because of a process called lenition. In the orthography, grapheme [H] is added next to the consonant to mark this change, which typically results in aspiration of the consonant.

Broadly, however, with the exception of some very common function words, the grapheme-to-phoneme relationship of Gaelic is regular, and many-to-one, making the task of pronunciation prediction straightforward, at least in principle.

3.3. Resources for ASR

The corpus consists of six hours of talk radio from the BBC's *Radio nan Gàidheal*, collected by the University of Edinburgh in 2010. The broadcasts are from the morning news and discussion programme, *Aithris na Maidne* recorded in clean studio conditions and sampled at 48kHz (any telephone speech from callers to the programme was removed). Speech is transcribed by fluent Gaelic speakers at utterance level. The speech data in the corpus can be categorized into three broad genres: read news, reports from correspondents and interviews. Due to the minority status of Gaelic within the UK, the corpus also has a high proportion of English words (853). English words present in the corpus are manually labelled. The corpus does not define a phoneme set, phoneme pronunciation dictionary or language model for ASR.

4. EXPERIMENTAL SETUP AND RESULTS

We use the Scottish Gaelic speech corpus collected by CSTR, University of Edinburgh for all the experiments.

4.1. Speech Data

The corpus consists of speech from 46 speakers. This includes 4818 utterances and 5083 unique words. The corpus did not have train, and test set division for the purpose of ASR. Therefore in this work we divided the database in to train, development and test sets in a speaker independent way. The training set consists of 22 speakers, 2389 utterances amounting to 3 hours of speech, the development set consists of 12 speakers, 1112 utterances amounting to 1 hour of speech and the test set consists of 12 speakers, 1317 utterances amounting to 1 hour of speech. The test data consists of 2246 unique words which includes 772 words not seen during training.

4.2. Pronunciation Lexicon

The database does not contain phoneme pronunciation lexicon. Also, the language lacks standard and well developed pronunciation dictionaries that can be used to develop grapheme-to-phoneme conversion systems [15]. In this work, the grapheme pronunciation lexicon was created for the words in the database. During the development of grapheme lexicon

- vowel graphemes (A, E, I, O, U) and long vowel graphemes or grave accents (À, È, Ì, Ò, Ù) were treated as separate graphemes.
- lenited consonants (BH, CH, DH, FH, GH, MH, PH, SH and TH) were treated as separate graphemes.
- consonant graphemes can be broad or slender. However, if the broad/slender assignment is ambiguous (i.e., they can be preceded by a broad vowel and followed by a slender vowel), the consonants are left as they are.
- word initial and final graphemes were treated as separate graphemes

Table 1 presents the list of graphemes in the dictionary. The graphemes J, K, Q, V, W, X, Y and Z, though not present in Gaelic words are present in the grapheme set because of the English words in the corpus. Each of the graphemes listed in the table can have three different variations, grapheme at the begin of word, end of word and middle of word. For example, the grapheme pronunciation of Gaelic word "CIAMAR" is [bs_C] [I] [A] [b_M] [A] [b_Rl]. Where 'bX' represents grapheme [X] is word begin grapheme, 'Xl' represents grapheme [X] is word final grapheme, 'b_X' represents [X] is a broad consonant and 's_X' represents [X] is a slender consonant. However, for English word "AIR" pronunciation is [bA] [I] [Rl], i.e., there are no broad and slender consonants. This resulted in total 248 context-independent graphemes.

Type	Graphemes
Vowels	A, E, I, O, U
Long Vowels	À, È, Ì, Ò, Ù
Broad consonants	b_B, b_BH, b_C, b_CH, b_D, b_DH, b_F, b_FH, b_G, b_GH, b_H, b_L, b_M, b_MH, b_N, b_P, b_PH, b_R, b_RR, b_S, b_SH, b_T, b_TH
Slender consonants	s_B, s_BH, s_C, s_CH, s_D, s_DH, s_F, s_FH, s_G, s_GH, s_H s_L, s_M, s_MH, s_N, s_P, s_PH, s_R, s_RR, s_S, s_SH, s_T, s_TH
Consonants	A, B, BH, C, CH, D, DH, E, F, FH, G, GH, H, I, J, K, L, M, MH, N, O, P, Q, R, S, T, TH, U, V, W, X, Y, Z

Table 1. Graphemes in Gaelic dictionary. 'b_X' represents [X] is a broad consonant and 's_X' represents [X] is a slender consonant

4.3. Systems

As done in previous works on under-resourced ASR, we build a HMM/GMM system with cepstral features to ascertain how well the standard ASR approach performs. We then build KL-HMM systems using the posterior features motivated earlier in Section 2.

All the systems use the grapheme lexicon presented earlier in Section 4.2 and model either context-independent (*mono*) or context-dependent (*tri*) grapheme subword units. Each grapheme subword unit was modeled by a 3 state left-to-right HMM. For context-dependent systems, word internal context models were trained. In the case of KL-HMM systems, the unseen contexts were backed-off to a seen context. As mentioned earlier in Section 3, the corpus does not include a language model. Therefore, we trained a bigram language model using sentences from the test set[1]. The different systems investigated in this work are

[1]In our future work we intend to build a language model.

1. *HMM/GMM:* The HMM/GMM system was trained with 39 dimensional perceptual linear prediction (PLP) cepstral coefficients ($c_0 - c_{12} + \Delta + \Delta\Delta$). Decision tree state trying method was used to cluster context-dependent grapheme models. Question set used for state tying is constructed by grouping a grapheme, its word begin, word final, broad and slender variants. Clustering resulted in 1934 tied states. State emission distributions were modeled with a mixture of 8 Gaussians.
2. *KL-HMM-MULTI:* We use an *of-the-shelf* MLP trained on SpeechDat(II) to estimate multilingual phoneme posterior features (MLP-MULTI) [13, 16]. The multilingual MLP was trained by pooling acoustic and lexical resources from five different languages of SpeechDat(II) corpus, namely British English, Italian, Spanish, Swiss French and Swiss German to classify 117 phonemes. Approximately, 12 hours of speech data from each language (totally amounting to 63 hours) was used to train the MLP. The input to the MLP was 39 dimensional PLP cepstral feature with 4 frames preceding context and 4 frames following context. For more details on MLP-MULTI, the reader is referred to [16]. SpeechDat(II) is a telephone speech corpus, hence, the Gaelic speech was down sampled to 8kHz before extracting PLP cepstral features. Gaelic PLP features were forward passed through MLP-MULTI to obtain multilingual posterior features which were then modeled by KL-HMM system.
3. *KL-HMM-GRAPH:* Grapheme posterior features were estimated using an MLP trained on the Gaelic speech corpus (MLP-GAELIC). The input to the MLP was 39 dimensional PLP features with four preceding and four following frame context. The frame level labeling (targets) for MLP training were obtained by performing force alignment of the training and development data using HMM/GMM system. Context-independent graphemes with atleast 100 frames were chosen as MLP targets. This resulted in 207 graphemes and hence the MLP is trained to classify 207 context-independent graphemes. The KL-HMM system was then trained with 207 dimensional grapheme posterior feature.
4. *KL-HMM-HIER:* As shown in Figure 1, the hierarchical MLP (MLP-HIER) used for grapheme posterior feature estimation consists of two MLPs. More specifically, grapheme posterior features were estimated using an MLP trained on multilingual posterior features (rather than traditional PLP features) with eight preceding and eight following frame context. The choice of the temporal context was based on previous studies [12]. The targets to train second MLP were obtained from HMM/GMM system (same as the targets used to train MLP-GRAPH). The KL-HMM system was trained using the grapheme posterior features estimated by MLP-HIER.

For all the KL-HMM systems, categorical state distributions were estimated by optimizing all the three local scores and the local score which resulted in minimum KL-divergence on the training set was selected, which in this case was RKL.

4.4. Results

Table 2 presents the word error rates in percentage on the test set of Gaelic corpus for different systems modeling contexts *mono* and *tri*. The results show that, a) all KL-HMM systems yield better performance than HMM/GMM system for both *mono* and *tri* contexts, b) System *KL-HMM-MULTI* which uses auxiliary acoustic and lexical resources yields better performance than System *HMM/GMM*.

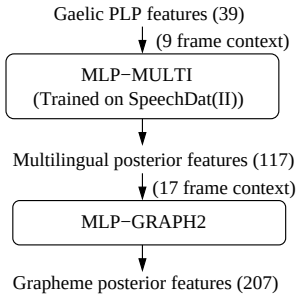

Fig. 1. Hierarchical MLP classifier to estimate grapheme posterior features. Numbers in the brackets indicate dimension of the feature vectors

This suggests that multilingual phoneme posterior features can be effectively ported across languages, c) It is interesting to note that System *KL-HMM-GRAPH* which uses grapheme posterior features (estimated by MLP trained with acoustic feature as input) yields performance comparable to System *KL-HMM-MULTI*, and d) System *KL-HMM-HIER* which uses both auxiliary multilingual corpus and Gaelic corpus to estimate the grapheme posterior features significantly outperforms all the other systems. Thus, supporting the idea that the estimation of grapheme posterior features could be effectively improved by modeling phoneme context information in conjunction with acoustics. This is not only interesting for under resourced languages but also for resource rich and majority languages.

Table 2. Word error rate in percentage on the test of Gaelic corpus

System	Context	
	mono	*tri*
HMM/GMM	42.7	35.2
KL-HMM-MULTI	41.8	27.2
KL-HMM-GRAPH	38.7	28.9
KL-HMM-HIER	29.8	22.6

5. CONCLUSIONS

In this paper, we studied the grapheme-based ASR using KL-HMM for under-resourced language, Scottish Gaelic. We investigated two different posterior features, namely multilingual posterior features and grapheme posterior features. The ASR studies conducted showed that irrespective of the type of the posterior feature used the KL-HMM approach outperforms the traditional HMM/GMM approach. Furthermore, the KL-HMM system using grapheme posterior features estimated by an hierarchical MLP (trained on both auxiliary data and Gaelic data) yields the best system. Some future directions that are worth investigating are,

- improving grapheme posterior feature estimates using deep neural networks [17], combining multiple feature streams [18].
- alternate posterior feature space representations. For example, the MLP could be trained to estimate context-dependent graphemes, or articulatory posterior features which are typically considered to be language independent [19].
- generation of lexical resources. The parameters of the System *KL-HMM-MULTI* capture the relation between graphemes and multilingual phonemes. The captured relationship could be exploited to generate phoneme lexical resources for Scottish Gaelic [14].

6. REFERENCES

[1] S. Kanthak and H. Ney, "Context-Dependent Acoustic Modeling using Graphemes for Large Vocabulary Speech Recognition," in *Proc. of ICASSP*, 2002, pp. 845–848.

[2] M. Killer, S. Stüker, and T. Schultz, "Grapheme based Speech Recognition," in *Proc. of European Conference on Speech Communication and Technology (EUROSPEECH)*, 2003.

[3] T. Schultz and K. Kirchhoff, "Multilingual Acoustic Modeling," in *Multilingual Speech Processing*. Academic Press, 2006.

[4] J. Dines and M. Magimai-Doss, "A Study of Phoneme and Grapheme based Context-Dependent ASR Systems," in *Proc. of Machine Learning for Multimodal Interaction (MLMI)*, 2007, pp. 215–226.

[5] M. Magimai.-Doss, R. Rasipuram, G. Aradilla, and H. Bourlard, "Grapheme-based Automatic Speech Recognition using KL-HMM," in *Proc. of Interspeech*, 2011, pp. 2693–2696.

[6] S. Stüker, "Integrating Thai Grapheme Based Acoustic Models into the ML-MIX Framework - For Language Independent and Cross-Language ASR," in *Proc. of the Spoken Languages Technologies for Under-resourced Languages (SLTU)*, 2008.

[7] S. Stüker, "Modified Polyphone Decision Tree Specialization for Porting Multilingual Grapheme Based ASR Systems to New Languages," in *Proc. of ICASSP*, 2008, pp. 4249–4252.

[8] V. B. Le and L. Besacier, "Automatic Speech Recognition for Under-Resourced Languages: Application to Vietnamese Language," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 17, pp. 1471–1482, 2009.

[9] E. Barnard, J. Schalkwyk, C. Heerden, and P. J. Moreno, "Voice Search for Development," in *Proc. of Interspeech*, 2010.

[10] T. Schlippe, E. G. K. Djomgang, N. T. Vu, S. Ochs, and T. Schultz, "Hausa Large Vocabulary Continuous Speech Recognition," in *Proc.of SLTU*, 2012.

[11] G. Aradilla, H. Bourlard, and M. Magimai Doss, "Using KL-Based Acoustic Models in a Large Vocabulary Recognition Task ," in *Proc. of Interspeech*, 2008.

[12] J. P. Pinto, G. S. V. S. Sivaram, M. Magimai.-Doss, H. Hermansky, and H. Bourlard, "Analysis of MLP Based Hierarchical Phoneme Posterior Probability Estimator," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, pp. 225–241, 2011.

[13] D. Imseng, R. Rasipuram, and M. Magimai.-Doss, "Fast and Flexible Kullback-Leibler Divergence based Acoustic Modeling for Non-native Speech Recognition," in *Proc. of Automatic Speech Recognition and Understanding (ASRU)*, 2011, pp. 348–353.

[14] R. Rasipuram and M. Magimai.-Doss, "Acoustic Data-driven Grapheme-to-Phoneme Conversion using KL-HMM," in *Proc. of ICASSP*, 2012, pp. 4841–4844.

[15] M. Wolters, "A Diphone-Based Text-to-Speech System for Scottish Gaelic," M.S. thesis, University of Bonn, 1997.

[16] D. Imseng, H. Bourlard, and M. Magimai.-Doss, "Towards mixed language speech recognition systems," in *Proc. of Interspeech*, 2010, pp. 278–281.

[17] A. Mohamed, G. Dahl, and G. Hinton, "Deep Belief Networks for Phone Recognition," in *Proc. of NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*, 2009.

[18] F. Valente, "A Novel Criterion for Classifiers Combination in Multistream Speech Recognition," *IEEE Signal Processing Letters*, vol. 16, no. 7, pp. 561–564, 2009.

[19] R. Rasipuram and M. Magimai.-Doss, "Integrating Articulatory Features using Kullback-Leibler Divergence based Acoustic Model for Phoneme Recognition," in *Proc. of ICASSP*, 2011, pp. 5192–5195.