Pronunciation Lexicon Development for Under-Resourced Languages Using Automatically Derived Subword Units: A Case Study on Scottish Gaelic

Marzieh Razavi[1,2], **Ramya Rasipuram**[1], **Mathew Magimai Doss**[1]

[1] Idiap Research Institute, CH-1920 Martigny, Switzerland
[2] École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{mrazavi, rramya, mathew}@idiap.ch

Abstract

Developing a phonetic lexicon for a language requires linguistic knowledge as well as human effort, which may not be available, particularly for under-resourced languages. To avoid the need for the linguistic knowledge, acoustic information can be used to automatically obtain the subword units and the associated pronunciations. Towards that, the present paper investigates the potential of a recently proposed hidden Markov model formalism for automatic derivation of subword units and lexicon development on a truly under-resourced and endangered language, more precisely Scottish Gaelic. Our studies show that the formalism can not only be useful in developing a lexicon that helps in building better automatic speech recognition systems, but can also be extended to find the relationship between the derived subword units and the existing knowledge about phonetic units from resource-rich languages, more precisely multilingual phones. Thus, the formalism paves a path for systematically combining acoustic and linguistic knowledge from multiple languages with the limited acoustic and linguistic knowledge of the under-resourced language in order to develop phone-like automatic subword unit based lexical resources.

1. Introduction

State-of-the-art automatic speech recognition (ASR) and text-to-speech (TTS) systems are based on phonemes or phones. This necessitates development of a pronunciation lexicon that transcribes each word as a sequence of phones. Phonetic lexicon development requires expert linguistic knowledge such as phone set of the language and knowledge about the relationship between written form, i.e., graphemes and phonemes. Such an expertise exists for majority languages (e.g., English and French) while there are many languages (e.g., Scottish Gaelic and Haitian Creole) that have little or no such expertise. This paper aims to explore methods to automatically discover "phone-like" subword units and develop lexicons for under-resourced languages given a limited amount of word-level transcribed speech data.

In the ASR community, there has been a sustained interest to automatically derive subword units and generate pronunciations using acoustic data typically for pronunciation variation modeling. With the growing interest in development of ASR systems for under-resourced languages, automatically derived subword units (ASWUs) have recently gained more attention as they can avoid the need for linguistic knowledge. In the context of unsupervised learning of the subword units, approaches based on segmentation and clustering (Lee and Glass, 2012; Garcia and Gish, 2006) and spectral based clustering (Jansen and Church, 2011) have been proposed. These approaches have been mainly limited to tasks such as keyword spotting and spoken term detection. Towards supervised learning of the subword units, in (Lee et al., 2013), a hierarchical Bayesian model approach was proposed to jointly learn the subword units and pronunciations. In (Hartmann et al., 2013), a spectral based clustering approach was used to derive subword units from a context-dependent grapheme-based system. The pronunciations were then transformed using a statistical machine translation (SMT) approach. In a more recent work, a novel hidden Markov model (HMM)-based formalism has been proposed which requires only word-level transcribed speech data for subword unit derivation and pronunciation generation (Razavi and Magimai-Doss, 2015). In this formalism, the ASWUs are derived through HMM-based clustering using transcribed speech; grapheme-to-ASWU relationship is learned through acoustic data; and finally the pronunciations are inferred using the orthographic transcriptions of the words and the learned grapheme-to-ASWU relationship. Experimental studies conducted on English showed that the derived ASWUs are "phone-like" and can yield better ASR systems than graphemes. It is worth mentioning that to date, to the best of the knowledge of the authors, the ASWU based lexicon development given transcribed speech has been investigated only on majority languages, especially English.

The present paper investigates the application of the HMM-based formalism to automatic subword unit derivation and pronunciation lexicon development on a genuinely under-resourced and endangered European language, namely, Scottish Gaelic. Specifically, we study whether the approach, which was originally investigated on a resource-rich language (English), can generalize for under-resourced languages with limited acoustic resources and no phonetic lexicon. More precisely, we investigate: (a) whether the ASWUs are "phone-like", and (b) whether the ASR system based on ASWUs would yield a better performance than graphemes.

The remainder of the paper is organized as follows. Section 2. describes in detail the HMM-based formulation for ASWU-based lexicon development. Section 3. provides information about the Scottish Gaelic language and the experimental setup. Section 4. presents a method to relate ASWUs to multilingual phone units to ascertain if the

This work was supported by Hasler foundation through the grant AddG2SU.

ASWUs are indeed phone-like. Section 5. presents an ASR study to show the potential of ASWUs in speech technology development. Finally Section 6. concludes the paper.

2. Approach

The recently proposed HMM-based formalism for subword unit derivation and pronunciation generation consists of three phases: 1) automatic derivation of subword units, 2) learning the probabilistic relationship between graphemes and ASWUs through acoustic information, and 3) pronunciation inference given the learned grapheme-to-ASWU (G2ASWU) relationship. This section briefly explains each phase of the HMM-based formalism.

2.1. Automatic Subword Unit Derivation

In this formalism, the subword units are derived from the clustered context-dependent (CD) units in a grapheme based system using maximum-likelihood criterion. More precisely, the ASWUs $\{a_d\}_{d=1}^{D}$ are the tied states of a grapheme based HMM/Gaussian mixture model (GMM) system obtained through decision tree clustering (Figure 1, part *(A)*). The underlying idea is that the clustering of CD grapheme HMM states yields units that can be linked to both graphemes and the standard spectral feature observations, specifically cepstral features that tend to capture information related to phones. It was demonstrated in the previous study on English that the derived ASWUs tend to be "phone-like" (Razavi and Magimai-Doss, 2015).

2.2. Learning the G2ASWU Relationship

In order to generate pronunciations based on ASWUs, the first step is to learn the relationship between graphemes and ASWUs. This is done through use of acoustic information in two stages (as shown in Figure 1, part *(B)*). In the first stage, the relation between the acoustic observations $\mathbf{x}_t$ (e.g., cepstral features) and ASWUs $\{a_d\}_{d=1}^{D}$ is modeled though an artificial neural network (ANN). Then in the second stage, the relation between the graphemes and ASWUs is learned in the grapheme-based Kullback-Leibler divergence based HMM (KL-HMM) framework in which (Aradilla et al., 2008):

1. The posterior probabilities of ASWUs $\mathbf{z}_t = [P(a_1|\mathbf{x}_t), \ldots, P(a_d|\mathbf{x}_t), \ldots, P(a_D|\mathbf{x}_t)]^{\mathrm{T}}$ estimated from the ANN are used as feature observations.
2. The HMM states $\{l_i\}_{i=1}^{I}$ represent CD grapheme states. Each HMM state is parameterized by a categorical distribution $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,d}, \ldots, y_{i,D}]^{\mathrm{T}}$ with $y_{i,d} = P(a_d|l_i)$ which models the relationship between the ASWUs $\{a_d\}_{d=1}^{D}$ and the CD grapheme state l_i.
3. The local score S defined at each state is based on the KL-divergence between the ASWU feature $\mathbf{z}_t$ and categorical distribution $\mathbf{y}_i$:
$$S(\mathbf{z}_t, \mathbf{y}_i) = \sum_{d=1}^{D} z_{t,d} \log(\frac{z_{t,d}}{y_{i,d}}) \quad (1)$$
4. The parameters (categorical distributions) are then estimated through Viterbi Expectation-Maximization by minimizing a cost function based on KL-divergence.

The parameters $\{\mathbf{y}_i\}_{i=1}^{I}$ in this setup then capture the probabilistic relationship between graphemes and ASWUs.

2.3. Pronunciation Inference

In the inference phase (as illustrated in Figure 1, part *(C)*), the learned G2ASWU relationships $\{\mathbf{y}_i\}_{i=1}^{I}$ in the KL-HMM together with the orthography of the word are exploited to generate pronunciations. More precisely, given the orthographic transcription of the word, the grapheme-based KL-HMM acts as a generative model and emits a sequence of ASWU posterior probabilities. The sequence of ASWU posterior probabilities is then decoded using an ergodic HMM, i.e., the ASWUs are connected in an ergodic fashion in HMM.[1]

3. Experimental Setup

This section describes the Scottish Gaelic language, the database and the experimental setups for subword unit derivation and pronunciation generation.

3.1. Scottish Gaelic

Scottish Gaelic belongs to the class of Celtic languages. It is considered as an endangered language spoken by only 60,000 people. There are about 51 phonemes in the language (Wolters, 1997). However, the number of phonemes can change depending on the dialect. The language lacks a proper phonetic lexicon and the available transcribed speech data are limited.

Scottish Gaelic alphabet has 18 letters, consisting of five vowels and thirteen consonants. The long vowels are represented with grave accents (À, È, Ì, Ò, Ù). There are twelve basic consonant types in Scottish Gaelic (B, C, D, F, G , I , L, M, N, P, R, S, T):

- Each consonant is either fortis or lenis (i.e., they are produced with greater or lesser energy). The lenited consonants are presented in the orthography with a grapheme [H] next to them.
- Each consonant is either broad (velarized) or slender (palatalized). Broad consonants are surrounded by broad vowels (A, O or U), while slender consonants are surrounded by slender vowels (E or I).

Due to the effect of lenited and broad/slender letters on the pronunciation, typically the number of graphemes in a word is relatively larger than the number of phonemes. The grapheme-to-phoneme relationship in Scottish Gaelic is therefore many-to-one in most cases.

3.2. Database

The Scottish Gaelic corpus was collected by the University of Edinburgh in 2010 and contains recordings from broadcast news and discussion programs[2]. In this paper, the database is partitioned into training, development and test sets according to the structure provided in (Rasipuram and Magimai.-Doss, 2015). The training set contains 2389 utterances with 3 hours of speech and 22 speakers. The development set has 1112 utterances with 1 hour of speech and 12 speakers. The test set consists of 1317 utterances from 12 speakers amounting to 1 hour of speech. There are a total of 2246 unique words in the test set of which 772 are not seen during training.

[1] Note that the pronunciation generation process was originally proposed in (Rasipuram and Magimai-Doss, 2012).

[2] http://forum.idea.ed.ac.uk/tag/scots-gaelic

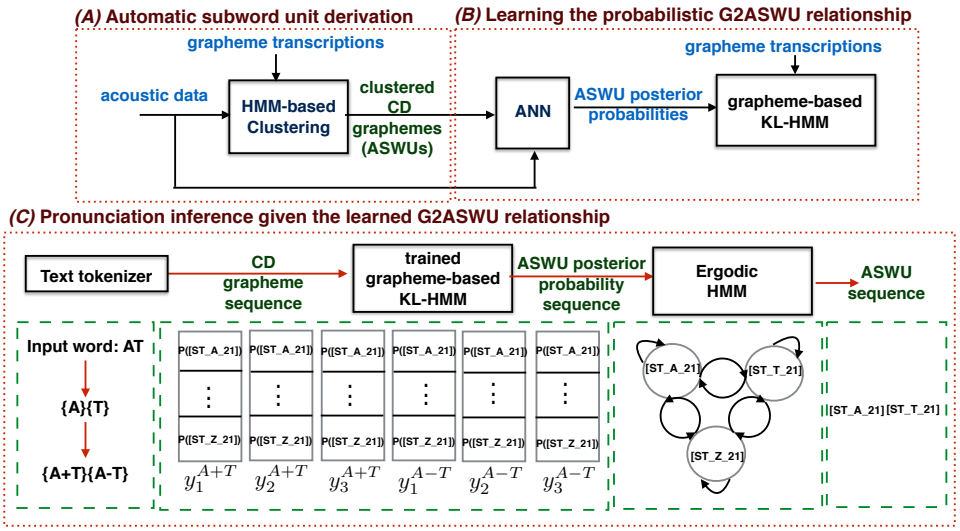

Figure 1: Block diagram of the HMM formalism for subword unit derivation and pronunciation generation. The subword units are represented in the form of HTK clustered states as $[ST\_G\_N]$, with G denoting a grapheme and N denoting a number.

The database does not provide any phonetic lexicon. The graphemic lexicon can be simply obtained from the orthography of the words. As the corpus also contains borrowed English words, the graphemes J, K, Q, V, W, X, Y and Z are also present in the lexicon. Therefore the lexicon consists of 32 graphemes including silence.

3.3. ASWU-based Pronunciation Generation

As explained in Section 2., the HMM-based formalism for subword unit derivation and pronunciation generation has three phases. This section explains the setup for each phase as follows:

(A) Automatic Subword Unit Derivation: In order to automatically derive the subword units, cross-word CD grapheme-based HMM/GMM systems were trained using HTK toolkit (Young et al., 2000). Each grapheme was modeled with a single HMM state. The decision tree based clustering was done with singleton questions using maximum likelihood criterion to derive the subword units. Different number of ASWUs were obtained by adjusting the log-likelihood increase during decision-tree based state tying. Note that the total number of graphemes in Scottish Gaelic (considering the broad, slender and lenition effects) are around 83 units (Rasipuram et al., 2013) which could be a good indicator for selecting the minimum number of ASWUs. Taking this observation into account, we initially selected the number of ASWUs as 85, 91 and 97.

(B) Learning the G2ASWU Relationship: As the first stage in learning the G2ASWU relationship, a five-layer ANN, more specifically multilayer Perceptron (MLP) was trained to classify the ASWUs. We used 39-dimensional PLP cepstral features with four preceding and four following frame context as MLP input. The optimal number of hidden units were obtained based on the frame accuracy on the development set. In most cases, each hidden layer had 1000 hidden units. The MLP was trained with output non-linearity of softmax and minimum cross-entropy error criterion, using Quicknet software (Johnson et al., 2004).

As the second stage, using the posterior probabilities of ASWUs as feature observations in the grapheme-based KL-HMM system, context-dependent (single preceding and single following) grapheme subword models were trained. Each grapheme subword unit was modeled with three HMM states. The parameters of the KL-HMM were estimated by minimizing a cost function based on the KL-divergence local score defined in Equation (1). For tying KL-HMM states, we applied KL-divergence based decision tree state tying method proposed in (Imseng et al., 2012). As a result of applying the state tying method, unseen grapheme contexts can be easily handled.

(C) Pronunciation Inference: In the pronunciation inference phase, each ASWU in the ergodic HMM was modeled with three left-to-right HMM states. During the pronunciation inference, some of the ASWUs that had less probable G2ASWU relationships were pruned out and a subset of derived ASWUs appeared in the lexicon. This can be seen from Table 1 which shows the properties of the ASWU-based lexicons together with the MLPs used. The MLPs are denoted as *MLP-*N where N denotes the number of ASWUs, and the lexicons are represented as *Lex-ASWU-*M with M denoting the actual number of subword units used. It can be seen that some of ASWUs are eliminated in the lexicons. For instance in the *Lex-ASWU-76*, from the 85 ASWUs obtained through clustering, only 76 are used.

Lexicon	# of units	MLP
Lex-ASWU-76	*76*	*MLP-85*
Lex-ASWU-82	*82*	*MLP-91*
Lex-ASWU-86	*86*	*MLP-97*

Table 1: Summary of the ASWU-based lexicons and the MLPs used.

We decided on the optimal number of ASWUs based on the performance at the speech recognition level. More precisely, we selected the ASWU-based lexicon which led to the best performing ASR system on the cross-validation set. In our experiments *Lex-ASWU-82* led to the best ASR

performance and is used in the rest of the paper. More information about the ASR setup is provided in Section 5..

4. Relating ASWUs to Phonetic Units

One of the fundamental questions that arises is that whether the ASWUs and the pronunciations obtained are linguistically meaningful. In the previous study (Razavi and Magimai-Doss, 2015), this was addressed by computing the KL-divergence between a Gaussian distribution modeling a mono-phone unit and the Gaussian distribution modeling an ASWU in the HMM/GMM setup. In the case of Scottish Gaelic, however, there is no phonetic lexicon available. So such an approach can not be pursued. In this section, we show that such linguistic interpretations can be achieved by relating the ASWUs to multilingual phones obtained from auxiliary resource-rich languages exploiting the KL-HMM framework. The underlying assumption here is that speech sound units are shared across languages, as the human speech production mechanism is common across languages.

4.1. Learning ASWU-to-Multilingual Phone Relationships

In Section 2., it was explained how the relationship between the graphemes and ASWUs can be learned through the parameters of the KL-HMM using acoustic data. Now, to learn the relationship between the ASWUs and multilingual phones, we can again exploit the same KL-HMM framework. Specifically, instead of using an MLP classifying ASWUs, a multilingual MLP trained on auxiliary acoustic and lexical resources is used to estimate the posterior probabilities of the multilingual phones $\mathbf{z}_t$; and instead of training a grapheme-based KL-HMM, an ASWU-based KL-HMM is trained. In other words the KL-HMM states l_i represent context-independent (CI) ASWUs. In this setup, the parameters of the KL-HMM, $\{\mathbf{y}_i\}_{i=1}^{I}$, capture the probabilistic relationship between the CI ASWUs and the multilingual phones.

Towards that, we used an off-the-shelf multilingual MLP trained on 63 hours of speech from five languages in SpeechDat(II) corpus to classify multilingual phones of size 117 (Rasipuram and Magimai.-Doss, 2015). We refer to the multilingual MLP as *MLP-MULTI-117*. We trained the CI-ASWU-based KL-HMM system by using the posterior probabilities of multilingual phones $\mathbf{z}_t$ estimated on the Scottish Gaelic data as feature observations. Given the categorical distribution $\mathbf{y}_i$ for each of CI ASWU l_i, the relationship between each ASWU and the multilingual phones is ascertained by choosing the phone with the highest probability.

4.2. Interpretation of the ASWUs

Table 2 provides the ASWU-to-multilingual phone mappings for some of the ASWUs. The mapped phones are shown in the SAMPA[3] format along with the probability of the phone within the brackets. Furthermore, we have provided example Gaelic words which contain the ASWUs within their pronunciations. In each example the graphemes which have been mapped to the ASWU in the pronunciation are highlighted.

[3]http://www.phon.ucl.ac.uk/home/sampa/

ASWU	mapped phone	example word	ASWU	mapped phone	example word
ST_C_22	/C/ [0.7]	SMAOINI**CH**	ST_T_21	/h/ [0.6]	**TH**OG
ST_C_23	/k/ [0.9]	**C**ADAL	ST_T_24	/t/ [0.7]	MO**T**A
ST_S_21	/S/ [0.8]	RI**S**	ST_G_22	/g/ [0.5]	**G**AD
ST_S_23	/s/ [0.8]	THU**S**A	ST_G_23	/k/ [0.5]	LA**G**
ST_F_21	/f/ [0.7]	**PH**ÀIRT	ST_R_22	/r/ [0.4]	MA**R**
ST_B_21	/b/ [0.5]	**B**RIS	ST_L_21	/l/ [0.8]	SAOI**L**
ST_B_22	/v/ [0.4]	A-**BH**OS	ST_L_23	/l/ [0.5]	SGEU**L**
ST_À_21	/a/ [0.5]	MH**À**L	ST_Ò_21	/o/ [0.3]	SP**Ò**RS
ST_A_212	/@/ [0.4]	AG**A**D	ST_O_23	/o/ [0.3]	ST**O**C
ST_E_21	/@/ [0.4]	S**E**	ST_I_23	/I/ [0.7]	TR**I**C
ST_E_23	/l/ [0.3]	WHAL**E**S	ST_I_28	/i/ [0.2]	TR**Ì**

Table 2: Some of the ASWUs together with their mapped phonemes in SAMPA format and some example words.

It can be observed from Table 2 that the ASWUs are indeed related to the phonetic units. For example, the ASWU [ST_S_21] is mapped to the sound /S/ (as found in the pronunciation of the English word *ACTION*: /{/ /k/ /S/ /n/) and is used in the pronunciation of the Scottish Gaelic word *RIS* which has the slender consonant *S*. On the other hand, the ASWU [ST_S_23] is mapped to the sound /s/ (as used in the pronunciation of the English word *EAST* : /i:/ /s/ /t/) and is found in the pronunciation of the Gaelic word *THUSA* which contains the broad consonant *S*. Similarly the consonant ASWUs [ST_F_21] and [ST_R_22] are related to sound units /f/ and /r/ . For the vowel ASWUs such as [ST_I_28] and [ST_E_21], the ASWUs are related to the phonetic units with a relatively lower confidence (i.e., lower probability). This is typical as the vowel grapheme units are mapped to more than one phone, while the grapheme consonants have mostly one-to-one relationship to the phones.

4.3. Interpretation of Generated Pronunciations

Table 3 presents a few words, due to space limitation, together with the ASWU-based pronunciations. To have a better sense of the generated pronunciations, each ASWU has been mapped to a multilingual phone according to the information in Table 2. We have also provided the 'perceived' pronunciations for each word through informal hearing of the Gaelic words from an online community-driven dictionary for Gaelic in which for most of the words an audio file pronouncing the word is available[4].

It is worth mentioning that in Scottish Gaelic, broad consonants *MH* and *PH* are pronounced as the English sounds $/v/$ and $/f/$ respectively; and the broad consonant *TH* is pronounced as the English /h/ sound[5]. It can be seen that the ASWU-based pronunciations to a certain extent capture the linguistic rules related to pronunciations. For instance, in the word *PHOS* the broad consonant PH is mapped to the /f/ sound. Similarly, in the word *MHÀL*, the broad consonant *MH* corresponds to $[ST\_B\_22]$ which is mapped to the $/v/$ sound. In fact, it can be observed that the mapped pronunciations corroborate well with the perceived pronunciations in several cases.

[4]http://www.learngaelic.net/dictionary/index.jsp

[5]https://en.wikipedia.org/wiki/Scottish_Gaelic_orthography

For some of the borrowed English words (e.g., *YOU* and *KATY*), on the other hand, the generated pronunciations using ASWUs seem to be influenced dominantly by Gaelic pronunciations. One explanation for such behavior is that the English words could be accented within the corpus. This behavior could also be attributed to the limited amount of English words available.

Word	*Lex-ASWU-82*	mapped pron.	perceived pron.
MHÀL	[ST_B_22] [ST_À_21] [S_L_23]	/v/ /a/ /l/	/v/ /a/ /l/
THOG	[ST_T_21] [ST_O_23] [ST_G_23]	/h/ /o/ /k/	/h/ /O/ /g/
PHÒS	[ST_F_21] [ST_Ò_21] [ST_S_23]	/f/ /o/ /s/	/f/ /o/ /s/
VOTE	[ST_B_22] [ST_O_23] [ST_T_24] [ST_E_21]	/v/ /o/ /t/ /@/	/v/ /@U/ /t/
YOU	[ST_I_28] [ST_O_23]	/i/ /o/	/j/ /u:/
KATY	[ST_G_23] [ST_A_212] [ST_T_24] [ST_I_28]	/k/ /@/ /t/ /i/	/k/ /eI/ /t/ /i/

Table 3: Example words from *Lex-ASWU-82*, together with their ASWU-based pronunciations, their mapped pronunciations based on the sequence of multilingual phone units and their perceived pronunciations.

5. ASR Studies

In the previous section, it was observed that the ASWUs are phone-like and the pronunciations are linguistically meaningful. The next question that arises is that whether these phone-like ASWUs and the developed pronunciation lexicon can bring any advantages for speech technology systems on the under-resourced language. For that purpose, we compared the ASWU-based ASR system with the grapheme-based ASR system which is an alternative approach when phonetic lexicon is not available. More precisely, we built two systems:

1. *HMM-GMM-GRAPH*: A cross-word context-dependent grapheme-based HMM/GMM system with 39 dimensional PLP cepstral features extracted using HTK toolkit. Each subword unit was modeled with three HMM states. For tying the HMM states, singleton questions were used. Each HMM state was modeled by a mixture of 8 Gaussians.
2. *HMM-GMM-ASWU*: A cross-word context-dependent HMM/GMM system using *Lex-ASWU-82* as the lexicon in the same setup as *HMM-GMM-GRAPH* system.

Table 4 presents the HMM/GMM performance in terms of word accuracy (WA). It can be observed that the *HMM-GMM-ASWU* system performs significantly better than the *HMM-GMM-GRAPH* system. As the number of tied states in both HMM/GMM systems are roughly the same, the two systems have similar complexity, and thus the improvements in the accuracy for the *HMM-GMM-ASWU* system can be attributed to the use of ASWUs.

System	# of units	# of tied states	WA
HMM-GMM-GRAPH	32	1158	64.6
HMM-GMM-ASWU	82	1161	**66.4**

Table 4: Performance of HMM/GMM systems in terms of word accuracy, i.e., (100 - word error rate).

6. Conclusion

In this paper, we investigated the potential of ASWUs for developing linguistically meaningful pronunciation lexicons. Our studies on Scottish Gaelic showed that the HMM-based formalism for subword unit derivation and pronunciation lexicon development can be effectively scaled to under-resourced languages. Furthermore, the studies also showed how auxiliary languages resources and prior linguistic knowledge can be exploited to understand the ASWUs and the inferred pronunciations in terms of meaningful linguistic units. Our future aim, in addition to extending the investigations to other under-resourced languages, is to systematically evolve the formalism and standardize the ASWU-based lexicon development approach with the help of the linguistic community.

7. References

Aradilla, G., H. Bourlard, and M. Magimai-Doss, 2008. Using KL-based acoustic models in a large vocabulary recognition task. In *Proceedings of Interspeech*.

Garcia, A. and H. Gish, 2006. Keyword spotting of arbitrary words using minimal speech resources. In *Proceedings of ICASSP*.

Hartmann, W., A. Roy, L. Lamel, and J. Gauvain, 2013. Acoustic unit discovery and pronunciation generation from a grapheme-based lexicon. In *Proceedings of ASRU*.

Imseng, D. et al., 2012. Comparing different acoustic modeling techniques for multilingual boosting. In *Proceedings of Interspeech*.

Jansen, A. and K. Church, 2011. Towards unsupervised training of speaker independent acoustic models. In *Proceedings of Interspeech*.

Johnson, D. et al., 2004. ICSI Quicknet Software Package. http://www.icsi.berkeley.edu/Speech/qn.html.

Lee, C. and J. R. Glass, 2012. A nonparametric Bayesian approach to acoustic model discovery. In *Proceedings of ACL*.

Lee, C., Y. Zhang, and J. R. Glass, 2013. Joint learning of phonetic units and word pronunciations for ASR. In *EMNLP*. ACL.

Rasipuram, R., P. Bell, and M. Magimai.-Doss, 2013. Grapheme and multilingual posterior features for under-resourced speech recognition: a study on Scottish Gaelic. In *Proceedings of ICASSP*.

Rasipuram, R. and M. Magimai-Doss, 2012. Acoustic data-driven grapheme-to-phoneme conversion using KL-HMM. In *Proceedings of ICASSP*.

Rasipuram, R. and M. Magimai.-Doss, 2015. Acoustic and lexical resource constrained asr using language-independent acoustic model and language-dependent probabilistic lexical model. *Speech Communication*, 68:23–40.

Razavi, M. and M. Magimai-Doss, 2015. An HMM-based formalism for automatic subword unit derivation and pronunciation generation. *Proceedings of ICASSP*.

Wolters, M., 1997. *A Diphone-Based Text-to-Speech System for Scottish Gaelic*. Ph.D. thesis, M.S. thesis, University of Bonn.

Young, S., D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, 2000. *The HTK Book Version 3.0*. Cambridge University Press.