Interspeech 2018
2-6 September 2018, Hyderabad

Effect of TTS Generated Audio on OOV Detection and Word Error Rate in ASR for Low-resource Languages

Savitha Murthy[1], *Dinkar Sitaram*[1], *Sunayana Sitaram*[2]

[1]PES University, Bangalore, India
[2]Microsoft Research, Bangalore, India

savithamurthy@pes.edu, dinkars@pes.edu, sunayana.sitaram@microsoft.com

Abstract

Out-of-Vocabulary (OOV) detection and recovery is an important aspect of reducing Word Error Rate (WER) in Automatic Speech Recognition (ASR). In this paper, we evaluate the effect on WER for a low-resource language ASR system using OOV detection and recovery. We use a small seed corpus of continuous speech and improve the vocabulary by incorporating the detected OOV words. We use a syllable-model to detect and learn OOV words and, augment the word-model with these words leading to improved recognition. Our research investigates the effect on OOV detection and recovery after adding missing syllable sounds in the syllable model using a Text-to-Speech (TTS) system. Our experiments are conducted using 5 hours of continuous speech Kannada corpus. We use an already available Festival TTS for Hindi to generate Kannada speech. Our initial experiments report an improvement in OOV detection due to addition of missing syllable sounds using a cross-lingual TTS system.

Index Terms: speech recognition, Indian Language, Kannada, Out of Vocabulary, OOV, Low resource, TTS

1. Introduction

OOV words are those words that are encountered during decoding of speech and are not present in the Lexicon of an ASR or Spoken Term Detection (STD) System. There has been research on OOV detection and recovery since early nineties. Earlier method used filler models to mitigate the presence of OOV words where OOV words were represented with separate grammar and acoustic model[1][2][3][4]. Recently, Iwami et al.[5] also have used similar method for OOV term detection in STD task, where they define a separate syllable lattice for OOV terms to improve recognition. The method of hybrid search space for OOV detection has been used where, the Language Model (LM) for the OOV words was defined using sub-words (morphemes, syllables, phonemes, characters or other sub-words) along with the word model LM. This had the advantage of the filler model as well as the sub-word representation. Hybrid search space is of two types - Hierarchical and Flat. In Hierarchical hybrid LM, OOV words are represented as separate sub-word entries in the Lexicon and a separate sub-word LM is trained[6][7][8]. When an OOV word is encountered during decoding, the sub-word LM probabilities are used instead of word LM. Another approach for OOV detection is the use of Flat hybrid LM [9][10]. Here there is a single LM where the sub-word representation for OOV words are specified interspersed with the In-Vocabulary(IV) words. Alternatively, mixed LM instead of hybrid has been employed by Reveil et al.[11] where, they use separate word LM and sub-word LM for OOV detection and recovery. Along with sub-words, semantic context of the words have been used to improve OOV detection and recovery. Sheikh et al.[12] propose a neural bag-of-words for retrieval of task specific (proper names in this case) words that an ASR would come across. Semantic word classes[13] and Part-of-Speech (POS) tagging[14] along with sub-word units have been used for better OOV detection and recovery. Apart from LM and semantic features, there has been research on use of acoustic features for OOV detection. Pham et al.[15] have devised a technique of re-scoring the hypothesis for OOV detection based on acoustic similarity scores obtained using Dynamic Time Warping (DTW). They combine two different scores to re-rank the detection first, by concatenating sub-word samples and second, by segmenting the hypothesis into sub-words. Wang et al.[16] investigate the use of local acoustic probabilities to influence the decoding process. They lay emphasis on acoustic similarities over LM scores and show an improvement in OOV detection.

The concerns for OOV words are different based on the tasks under consideration. Reduction in OOV words is an important aspect of open vocabulary speech recognition without emphasis of detection or recovery. Hierarchical and Flat hybrid models with words and sub-words have been used to improve the performance of open vocabulary ASR[17][18][19]. For STD or Keyword Search (KWS) tasks, the emphasis is on OOV term detection and not recovery. Sub-word models have been used for OOV detection in KWS task[20][13][5][8][15]. Approaches other than use of sub-words have also been employed. The research of Asami et al.[21] involves finding recurrent OOV words in spoken document based on the hypothesis that repeating segments of OOV exhibit the same consistency. Another approach of handling OOV words (without use of sub-words) in KWS task is the proxy based approach. Here the OOV words are assigned IV proxy words and recognition of these proxy words is deemed as recognizing the OOV, resulting in better handling of OOV words[20][22][23]. Research regarding OOV words for ASR systems investigates OOV detection as well as recovery for better decoding. Sub-word models for OOV detection and recovery in ASR are used in [7][8][9][10][11][13][14][15].

In this paper, we address the problem of OOV detection and recovery for Kannada, a major language of India and the official language of Karnataka state. Kannada is a low-resource language with respect to building an ASR. ASR system for low-resource languages, having limited Lexicon and speech data, will undoubtedly come across OOV words. Additionally, Kannada is a highly inflected and agglutinative language (Proto-Dravidian), and each root word can have up to 400 forms depending upon case, number, gender, and so on [24]. In order to achieve good ASR accuracy, we would require a corpus with all possible words in the language. A lot of time and effort is required to build such a comprehensive speech corpus. Our ap-

10.21437/Interspeech.2018-1555

proach involves starting with a small seed corpus, recognizing OOV words, and incorporating them into ASR system thus increasing the size of the corpus. We infer from our study that OOV detection in ASR is achieved by sub-sequence representations with focus on improving WER for languages with sufficient resources. The focus of our work is to build an ASR system for a low-resource language, that can detect and learn OOV words in test data and enhance its vocabulary automatically. We make use of a syllable model ASR for OOV detection and recovery. We propose the use of a TTS System to generate audio for syllables in the test audio that are missing in the training data. We study the effect of incorporating TTS generated audio (for the missing syllables) on OOV detection and WER. We believe this approach would increase the ability of an ASR system to recognize more words and also reduce WER.

2. Method

Study by Qin et al.[25]and Qu et al.[26] suggests that sub-word models perform better than phonemes for low resource conditions. We employ a method where the hypothesis of a syllable (as sub-word) model, after post-processing is used along with the Lexicon of the word model to detect OOV words. Kannada is alphasyllabary with its orthography (known as aksharas) representing the syllables sounds[27] which form basic pronunciation units. This makes the representation of syllables as words in the Lexicon more straightforward without requiring any Grapheme to Phoneme mapping system.

2.1. Baseline

We use a word-model ASR built with a seed corpus of Kannada containing 5 hours of continuous (read) speech as the baseline. A context dependent Acoustic Model (AM) is trained using 80% of the audio and corresponding transcripts. Since the data for training is very less($\sim$ 4 hours of recording), we used a Gaussian Mixture Model (GMM)-Hidden Markov Model (HMM) model and not Deep Neural Network (DNN) with 8 Gaussians per state and 200 tied states using Baum-Welch algorithm with 39 MFCC features. We generate a tri-gram LM for word-model from the original transcripts containing words. The Lexicon for the word model contains pronunciations of the words in training set and are represented as phonemes. The phone-set comprises of 48 phonemes, including SIL for silence. We use CMU PocketSphinx ASR toolkit[1] and Language Modeling Toolkit.

2.2. OOV Detection and Recovery

We train a separate syllable based ASR for OOV detection and recovery. For syllable-model LM, we first syllabify the word transcripts using syllabification rules for Kannada. The LM for syllable-model is generated using these syllabified transcripts. The Lexicon for the syllable model contains syllables listed as words. The pronunciations for syllables are automatically derived from the words using orthographic rules and are defined in terms of the same phoneme set used for word model. Combinations of word boundary markers (for beginning of a word and end of a word) along with the syllables are specified as different words in the Lexicon resulting in a total of 856 syllables that appear in the training set.

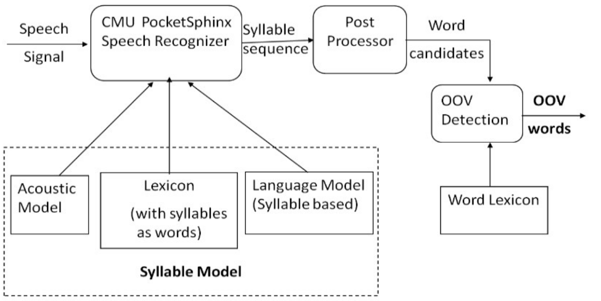

Figure 1: *Syllable based decoding with post-processing for OOV detection.*

We post-process the syllable hypothesis (1-best) from the syllable ASR to form word candidates. The word candidates are looked up in the Lexicon of the word model to determine potential OOV words. These words are then validated and pruned against Kannada wiki corpus[2] resulting in a set of valid OOV words that are not present in the training set. We add the valid OOV words to the Lexicon of the word-model. The pronunciations of recovered Kannada words are derived from orthographic rules for Kannada. Sentences containing the OOV words are extracted from the Kannada wiki corpus. The LM generated for these sentences is merged with the existing word-based LM to form an enhanced word LM with the learned OOV words. Figure 1 depicts our approach of OOV detection using syllable-model.

2.3. Inclusion of New Syllables using TTS into the Syllable ASR

The total syllables in Kannada, considering vowel(V), (consonant-vowel)CV, CCV and CCCV as syllable forms [28] amounts to more than half a million. Eliminating some of the combinations we can estimate the number of syllables used in practice approximately to 15000 or more. Combining word boundaries with the syllables would result in around 45000 syllables or more. The Lexicon defined based on the training data contains 856 syllables considering word boundary context, which is a minute fraction of the total possibilities of syllable sounds that can appear in Kannada speech. The ability of an ASR system to recognize *OOV syllables* not in the Lexicon would definitely aid in learning more OOV words. Our objective is to incorporate OOV syllables into the syllable-model using a TTS.

For our experiments, we identify the syllables present in the test data that are not present in the training set. The transcripts for these new syllables and words containing these syllables are prepared. Since Indian Languages are syllable based with similar sounds we can cross lingually use a TTS trained in one Indian Language to generate speech for another. We use the already trained and available Festival TTS[3] for Hindi to generate audio for Kannada speech. We generate the audio waveforms for the new syllables and the words based on the prepared transcripts. We augment the Lexicon of the syllable-model ASR with the new syllables and words. We generate a new LM for the transcripts containing new syllables and words and merge

[1] Shmyrev, Nickolay. "CMUSphinx Open Source Speech Recognition." CMUSphinx Open Source Speech Recognition. Accessed July 13, 2017. http://cmusphinx.sourceforge.net/

[2] "Kannada Wikipedia Dump, January 8th 2015". Accessed July 13, 2017 ttps://archive.org/details/knwiki-20150108.

[3] "Festival Speech Synthesis System", Accesses March 13, 2018. http://www.cstr.ed.ac.uk/projects/festival/

it with the existing LM, resulting in an augmented LM for the syllable-model ASR. We retrain the AM after including the TTS generated audio for the new syllables and words. The augmented syllable-model ASR is used for OOV detection and recovery.

3. Experiments

3.1. Data

For our experiments, we recorded a multi-speaker corpus using the Kannada transcripts from IIIT-Hyderabad[4] . The transcripts contain 1000 sentences in Kannada with a vocabulary of 1741 words. The training data was recorded for the first 700 sentences by 34 speakers resulting in 4 hours of training audio. The test data was recorded for the remaining 300 sentences from 12 speakers resulting in 1 hour of test audio. We cleaned the transcripts to remove punctuations and then transliterated them to English alphabets using Baraha software. We then converted the Baraha English scripts to English alphabet notations used are from the common label set provided by IIT-Madras[5]. Below is an example of the conversion to IIT-M label set.

Kannada script:	ಶಿಕ್ಷಣ ಪಡೆದ ಬಳಿಕ
Romanized notation:	“śikṣaṇa paḍeda baḷika”
Baraha Transliteration:	“shikShaNa paDeda baLika”
Common Label : set:	“SHIKSXANXA PADXEDA’ BALXIKA”

We then defined the Lexicon from the prepared common label notation transcripts . The Lexicon for the word-model contains words and their pronunciation defined in terms of phoneme as shown below:
Word-model transcript: “SHIKSXANXA PADXEDA BALXIKA”
Lexicon entry for the word-model:

SHIKSXANXA	SH I K SX A NX A
PADXEDA	P A DX E D A
BALXIKA	B A LX I K A

We prepared separate Lexicon for the syllable-model. Entries in syllable-model Lexicon contain syllables defined as words and their pronunciations as phonemes. We used syllabification rules for Kannada to convert word transcripts into syllable form. Word boundaries (ˆ , $) along with syllables[10] were used for better recognition. Below is as example:
Syllable based transcript: “ˆSH_I K_SX_A NX_A$ ˆP_A DX_E D_A$ ˆB_A LX_I K_A$”

The syllables constituting the words “SHIKSXANXA PADXEDA” defined in the Lexicon using phones as given below:

ˆSH_I	SH I
K_SX_A	K SX A
NX_A$	NX A
ˆP_A	P A
DX_E	DX E
D_A$	D A

3.2. Evaluation

We use Festival TTS for Hindi to generate audio for the missing Kannada syllables and words containing the syllables. We use diphone model and Hindi voice from an adult male to generate Kannada audio. Training the initial system with multi-speaker corpus enabled us to incorporate the Hindi TTS male voice into the Acoustic Model.

For comparison purpose, we prepare three different syllable-models (Model A, B and C) and evaluated their effect on OOV detection. We describe the three syllable-models in the following subsection:

3.2.1. Model A:

For Model A, we train a syllable model from the initial multi-speaker corpus. These syllabified transcripts and audio include only those syllables that occurred in the original training data and no new syllables are added. We use this syllable-model to detect OOV words and augment the word-model with the detected OOV words and sentences.

3.2.2. Model B:

For Model B, we identify 41 new syllables present in the test set but not in the training set. We prepare 82 transcripts for the new syllables. Two transcripts are defined for each syllable - the first transcript contains only the syllable representation (repeated 10 times) and, the second transcript consists of the Kannada words (defined to accommodate different contexts for the syllable, repeated 4-5 times) containing the syllable. We add these syllables and words to the syllable-model Lexicon. We train a new LM for the 82 transcripts and merge them with the existing LM to obtain an enhanced LM for the syllable-model. The AM is same as Model A. We use this syllable-model with augmented Lexicon and LM (but not AM) for OOV detection and enhance the word-model with the detected OOV words and sentences.

3.2.3. Model C:

For Model C, we use Festival Hindi TTS to generated audio for the 82 Kannada transcripts (described in Model B). We use these audio recordings to retrain the AM along with the other multi-speaker recordings. We also augmented the Lexicon and LM with the new syllables and words. We used this syllable-model containing augmented Lexicon, LM and AM for OOV detection and enhanced the word-model with the detected OOV words and sentences.

4. Results and Discussion

We use the WER from the word-model ASR trained on multi-speaker recordings (without OOV detection and recovery) as baseline. We report WERs for the word-model after employing OOV detection and recovery using the three different syllable models. We also compare the performance of the different syllable models.

Fig.2 depicts WERs for the baseline and different word models after OOV inclusion using different syllable models. The least WER of 38.02% is obtained using the syllable-model ASR incorporating retrained AM with TTS audio and, augmented Lexicon and LM with the new syllables and words (Model C). We also report the number of OOV words and new syllables recognized by the enhanced word-model ASR after OOV recovery in Table 1.

The Kannada script and corresponding romanized notations

[4] Accessed July 13, 2017. http://festvox.org/databases/iiit_voices

[5] Source: “IIT-Madras-Indic TTS Common label Set”. Accessed July 13, 2017 https://www.iitm.ac.in/donlab/tts

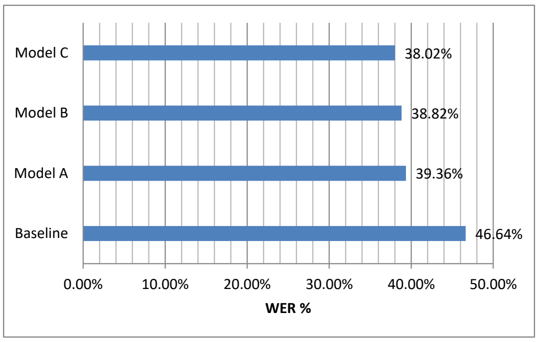

Figure 2: *Wordmodel WER at using different stages of enhanced syllable-model ASR for OOV detection.*

Table 1: *OOV words and new syllables recognized using different syllable models after OOV recovery.*

Syllable Model	OOV words recognized	New syllables recognized
Model A	77	-
Model B	82	17
Model C	86	21

along with IIT-M label set notation for the word examples used in the discussion is listed in 2. Our results show that use of a separate syllable based ASR helps detect and recover OOV. For example, the word "NIWRXTTARAADARU" is not recognized in the initial word model since it does not appear in the Lexicon and is an OOV. This word can be broken down into syllables as NI, WRX, TTA, RAA, DA and RU. All these syllables are present in the training set for the syllable model and hence are recognized correctly. These syllables when concatenated (considering the word boundaries) form the word NIWRXTTARAADARU which is a valid Kannada word (meaning retired in English). Thus syllables can be used to include new words that are not present in the Lexicon.

Table 2: *Kannada word examples with different notations.*

Native script	Romanized notation	IIT-M label set notation
ಆಯುರ್ವೇದ	āyurvēda	AAYURWEEDA
ನಿವೃತ್ತರಾದರು	nivṛttarādaru	NIWRXTTARAADARU
ಬೃಂದಾವನ	bṛṃdāvana	BRXQDAAWANA
ವೃಂದಾವನ	vṛṃdāvana	WRXQDAAWANA
ಭಾರತದ	bhāratada	BHAARATADA

However, due to the limited size of the corpus not all the syllables are present in the training set. For example words AAYURWEEDA and BRXQDAAWANA are not recognized because the syllable RWEE (CCV) and syllable BRX are not present in the training set. Augmenting the syllable-model ASR to include these syllables in the Lexicon and Language Model (Model B) facilitates detection and inclusion of the word AAYURWEEDA but not BRXQDAAWANA. This is because all the sounds in the word AAYURWEEDA are present in the initial acoustic model. The word BRXQDAAWANA is decoded as WRXQDAWANA in the syllable model. This is because the syllable BRX which is not present in the training data is acoustically closest to the syllable WRX that is present in the training data. Incorporating TTS generated audio for the new syllables (including BRX) results in correct detection and recovery of the word BRXQDAAWANA. However, only a 0.8% improvement using Model C over Model B for OOV detection and recovery can be attributed to the fact that the frequency of the new syllables in the test data is very low (1-2 times).

Table 3: *WER of different syllable models and corresponding word model after OOV inclusion.*

Model Model	Syllable Model WER	Word Model WER (with OOV inclusion)
Model A	53.23%	39.46%
Model B	49.26%	38.82%
Model C	49.23%	38.02%

Table 3 depicts the comparison of WER of different syllable models (after post-processing to form word-candidates) and WER of word-model using corresponding syllable models for OOV detection and recovery. Syllable based recognition as a separate framework has the advantage of incorporating OOV terms into and at the same time preserving the word level constraints of the word model. The WER for the word model after OOV inclusion is lower than the WER for the corresponding syllable model used. This reinforces the claim that using a separate syllable ASR does not affect the IV words in the word model.

5. Conclusions

Our experiments show that the WER for a low-resource language ASR system can be reduced by starting with a small seed corpus and learning OOV words. Our approach of using a separate syllalbe model for OOV detection and recovery does not affect the recognition of In-Vocabulary words and at the same time improves recognition of the word-model. Also, our study of different types of syllable-model for OOV detection depicts that enhancing the Acoustic Model with TTS generated audio helps learn new syllables leading to better OOV detection and improved WER. We believe this approach of using TTS audio has not been tried before. Our approach is applicable to other Indian Languages or any language with a close relation between orthography and pronunciations.

We believe the WER we obtained are due to low amount of data. The WER we have obtained are still comparable with those reported as baseline using GMM for Interspeech Challenge. In future, we plan to evaluate our approach on other low resource languages with more data than what we have experimented on. We also plan to make use of DNNs for better acoustic modeling. There has been research on use of web data for augmenting LMs[29][30][31] and AMs[32]. We plan to study the use of Kannada wiki corpus for generating LM and TTS audio for better learning.

6. References

[1] A. Asadi, R. Schwartz, and J. Makhoul, "Automatic modeling for adding new words to a large-vocabulary continuous speech recognition system," *Acoustics, Speech, and Signal Processing*, vol. 1, pp. 305–308, 1991.

[2] A. Asadi, "Automatic Detection and Modeling of New Words in a Large Vocabulary Continuous Speech Recognition System," pp. 263–265, 1991.

[3] G. Jusek, A., Fink, G. A., Kummert, F., Rautenstrauch, H., & Sagerer, "Detection of unknown words and its evaluation," in *Fourth European Conference on Speech Communication and Technology*, 1995.

[4] T. Kemp and A. Jusek, "Modelling unknown words in spotaneous speech," *International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 1, pp. 530–533.

[5] K. Iwami, Y. Fujii, K. Yamamoto, and S. Nakagawa, "Out-of-vocabulary term detection by N-gram array with distance from continuous syllable recognition results," *2010 IEEE Workshop on Spoken Language Technology, SLT 2010 - Proceedings*, pp. 212–217, 2010.

[6] J. Bazzi, I., & Glass, "Heterogeneous lexical units for automatic speech recognition: preliminary investigations," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference*, 2000.

[7] I. Bazzi, "Modelling Out-of-Vocabulary Words for Robust Speech Recognition," *Dissertation Abstracts International, B: Sciences and Engineering*, vol. 63, no. 11, 2003.

[8] Y. Y. Wang, Xuyang, Zhang, Pengyuan, NA Xingyu, PAN Jielin, "Handling OOV Words in Mandarin Spoken Term Detection with an Hierarchical n -Gram Language," vol. 26, no. 6, 2017.

[9] M. Yazgan, A., & Saraclar, "Hybrid language models for out of vocabulary word detection in large vocabulary conversational speech recognition," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference*, 2004, pp. Vol. 1, pp. I–745.

[10] L. Qin and A. Rudnicky, "OOV Word Detection using Hybrid Models with Mixed Types of Fragments," pp. 2–5, 2012.

[11] B. Réveil, K. Demuynck, and J. P. Martens, "An improved two-stage mixed language model approach for handling out-of-vocabulary words in large vocabulary continuous speech recognition," *Computer Speech and Language*, vol. 28, no. 1, pp. 141–162, 2014.

[12] I. Sheikh, I. Illina, D. Fohr, and G. Linarès, "Improved neural bag-of-words model to retrieve out-of-vocabularywords in speech recognition," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-Sept, pp. 675–679, 2016.

[13] A. Horndasch, A. Batliner, C. Kaufhold, and E. Nöth, "Combining semantic word classes and sub-word unit speech recognition for robust OOV detection," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-Sept, pp. 1335–1339, 2016.

[14] L. Qin and A. Rudnicky, "Building a vocabulary self-learning speech recognition system," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, no. September, pp. 2862–2866, 2014.

[15] V. T. Pham, H. Xu, X. Xiao, N. F. Chen, E. S. Chng, and H. Li, "Rescoring hypothesized detections of out-of-vocabulary keywords using subword samples," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-Sept, pp. 933–937, 2016.

[16] X. Wang, T. Li, P. Zhang, J. Pan, and Y. Yan, "Enhanced out of vocabulary word detection using local acoustic information," *Proceedings - 2014 10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2014*, pp. 594–597, 2014.

[17] M. A. Basha Shaik, D. Rybach, S. Hahn, R. Schlüter, and H. Ney, "Hierarchical Hybrid Language models for Open Vocabulary Continuous Speech Recognition using WFST," *Proceedings of the Workshop on Statistical and Perceptual Audition (SAPA - SCALE)*, no. October 2015, pp. 46–51, 2012.

[18] M. Bisani and H. Ney, "Open vocabulary speech recognition with flat hybrid models." *Interspeech*, pp. 3–6, 2005.

[19] T. Hirsimäki, M. Creutz, V. Siivola, M. Kurimo, S. Virpioja, and J. Pylkkönen, "Unlimited vocabulary speech recognition with morph language models applied to Finnish," *Computer Speech and Language*, vol. 20, no. 4, pp. 515–541, 2006.

[20] G. Chen, O. Yilmaz, J. Trmal, D. Povey, and S. Khudanpur, "Using proxies for OOV keywords in the keyword search task," *2013 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013 - Proceedings*, pp. 416–421, 2013.

[21] T. Asami, R. Masumura, Y. Aono, and K. Shinoda, "Recurrent Out-of-Vocabulary Word Detection Using Distribution of Features," pp. 1320–1324, 2016.

[22] M. S. Batuhan, Gundogdu, "DISTANCE METRIC LEARNING FOR POSTERIORGRAM BASED KEYWORD SEARCH," *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference*, pp. 5660–5664, 2017.

[23] Z. Lv, J. Kang, W. Q. Zhang, and J. Liu, "An LSTM-CTC based verification system for proxy-word based OOV keyword search," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 5655–5659, 2017.

[24] Krishnamurti Bhadriraju, *The Dravidian Languages - Bhadriraju Krishnamurti - Google Books*, 1st ed. Cambridge University Press, 2006.

[25] L. Qin, M. Sun, and A. Rudnicky, "OOV Detection and Recovery Using Hybrid Models with Different Fragments." *Interspeech*, no. August, pp. 1913–1916, 2011.

[26] Z. Qu, P. Haghani, E. Weinstein, and P. Moreno, "Syllable-Based Acoustic Modeling with CTC-SMBR-LSTM," pp. 173–177, 2017.

[27] S. Nag, "Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness," *Journal of Research in Reading*, vol. 30, no. 1, pp. 7–22, 2007.

[28] M. M. Prasad, M. Sukumar, and a. G. Ramakrishnan, "Divide and conquer technique in online handwritten Kannada character recognition," *Proceedings of the International Workshop on Multilingual OCR - MOCR '09*, p. 1, 2009.

[29] C. Xie, W. Guo, G. Hu, and J. Liu, "Web Data Selection Based on Word Embedding for Low-Resource Speech Recognition," pp. 1340–1344, 2016.

[30] A. Gorin, R. Lileikyte, G. Huang, L. Lamel, J. L. Gauvain, and A. Laurent, "Language model data augmentation for keyword spotting in low-resourced training conditions," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-Sept, pp. 775–779, 2016.

[31] G. Mendels, E. Cooper, V. Soto, J. Hirschberg, M. Gales, K. Knill, A. Ragni, and H. Wang, "Improving speech recognition and keyword search for low resource languages using web data," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2015-Janua, pp. 829–833, 2015.

[32] J. Nouza, R. Safarik, and P. Cerva, "ASR for south slavic languages developed in almost automated way," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-Sept, pp. 3868–3872, 2016.