Island-Driven Search Using Broad Phonetic Classes

Tara N. Sainath

MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar St. Cambridge, MA 02139, U.S.A.
tsainath@mit.edu

***Abstract*—Most speech recognizers do not differentiate between reliable and unreliable portions of the speech signal during search. As a result, most of the search effort is concentrated in unreliable areas. Island-driven search addresses this problem by first identifying reliable islands and directing the search out from these islands towards unreliable gaps. In this paper, we develop a technique to detect islands from knowledge of hypothesized broad phonetic classes (BPCs). Using this island/gap knowledge, we explore a method to prune the search space to limit computational effort in unreliable areas. In addition, we also investigate scoring less detailed BPC models in gap regions and more detailed phonetic models in islands. Experiments on both small and large scale vocabulary tasks indicate that our island-driven search strategy results in an improvement in recognition accuracy and computation time.**

I. Introduction

Many speech scientists believe that human speech processing is done by first identifying regions of reliability in the speech signal and then filling in unreliable regions using a combination of contextual and stored phonological information [1]. However, most speech decoding paradigms are typically performed left-to-right without utilizing knowledge of reliable regions. In addition, the search computational effort is mainly concentrated in unreliable regions when in reality most of the information in the signal can be extracted from the reliable areas. In the case of noisy speech, if phrases are unintelligible, this may even lead the search astray and make it impossible to recover the correct answer. This is also a problem in large vocabulary speech systems, where many pruning algorithms do not utilize knowledge of reliable regions, and thus may prune away too many hypotheses in unreliable regions and keep too many hypotheses in reliable areas.

Island-driven search is an alternative method to better deal with noisy and large vocabulary systems. This strategy works by first hypothesizing islands from reliable regions in the signal, and then working outwards from these islands to recognize unreliable areas. Island-driven search has been explored in areas such as parsing and handwriting recognition, though has been relatively unexplored in automatic speech recognition (ASR) due to numerous challenges.

First, the choice of island regions is a very difficult and unsolved problem [1]. For example, [2] explores an island-driven search for ASR. In this paper, a first-pass recognition is performed and islands are identified from stable words in the N-best list of hypotheses. Words in the island regions are held constant, while words in the gap regions are re-sorted using the N-best list. However, we argue that, if a motivation behind island-driven search is to identify reliable regions to influence effective pruning, identifying these regions from an N-best list generated from a pruned search space may not be an appropriate choice. Thus, our first goal is to develop a methodology to identify reliable island regions.

Second, the nature of speech recognition poses some constraints on the preferred strategy for island-driven search. While island searches have been explored both unidirectionally and bidirectionally, unidirectional search is more attractive in ASR due to the computational benefits. Unidirectional island-driven techniques typically make use of a heuristic strategy to decrease the number of nodes expanded during search. Therefore, our second goal is to explore the use of island/gap regions in a unidirectional framework to decrease the computational effort in unreliable areas.

Third, the potential computational complexities of island-driven search have limited its use in large vocabulary tasks. For example, the BBN HWIM system [1] utilizes island information for parsing. While this type of approach has shown promise for small grammars, computational complexities of the island parser have limited its use in large scale tasks. Thus, our third goal is to investigate an island-driven technique which can be applied to small and large scale tasks.

In this paper, we look to develop a method of island-driven search which can be incorporated into an ASR framework. First, we explore utilizing broad phonetic classes (BPCs), which have been shown to represent spectrally distinct portions of the speech signal [3], to identify reliable island regions from a speech utterance. Second, we utilize island/gap knowledge in designing a pruning strategy to better guide the search. Third, to limit unnecessary computational search effort in gap regions, we look at scoring less detailed BPC models in gaps and more detailed acoustic models in island regions.

We explore the proposed island-driven techniques on small and large vocabulary noisy speech tasks. Our experiments utilize the SUMMIT segment-based recognizer [4] developed at MIT. On the small vocabulary task, we find that our island-based pruning method offers improvements in both performance and computation time, while further usage of island information to score BPC models in gaps offers additional improvements. Extending these proposed methods to a large vocabulary task, we find that recognition performance does not degrade using island-driven techniques and the methods still provide faster computation time.

The rest of this paper is organized as follows. Our method for detecting islands is described in Section II. Utilization of

islands/gaps for search space pruning and scoring BPC models in gaps are presented in Sections III and IV respectively. Section V outlines the experiments performed, while Sections VI and VII discuss the results on the small and large scale tasks. Finally, Section VIII summarizes the paper.

II. Identifying Islands

We investigate a method to learn islands by using information about BPCs which have been identified with high confidence from the input speech signal. Our representation of BPCs for island detection include vowels/semi-vowels, nasals, weak fricatives, strong fricatives, stops, closures and silence, as our past research with these BPCs have illustrated that they are relatively acoustically distinct (i.e., [3], [5]). To determine confidence scores for hypothesized BPCs, we explore a BPC-level acoustic confidence scoring technique, presented in [6].

A. Confidence Features

First, we derive a series of features for each hypothesized BPC based on frame-level acoustic scores generated from a BPC recognizer described in [5]. At each frame, a maximum *a posteriori* probability and normalized log-likelihood score are computed for the hypothesized BPC. Using these frame-level acoustic confidence scores, we can derive BPC-level features, f, for each hypothesized BPC by taking various averages across the frame-level scores ([6]).

After BPC-level features are extracted from each hypothesized BPC, a Fisher Linear Discriminant Analysis (FLDA) projection is applied to reduce the set of BPC-level features f into a single dimension confidence score. The goal of the FLDA is to learn a projection vector w to reduce dimensionality of f while achieving maximal separation between two classes. Typically, these two classes are correctly and incorrectly hypothesized sub-word units (i.e., [6]). However, the goal of our work is to identify reliable island regions, not correctly hypothesized BPCs. More intuitively, a silence or stop closure could be hypothesized correctly but generally provides little reliability information on the actual word spoken relative to a voiced sound, such as a vowel. Therefore, a 2-class unsupervised k-means clustering algorithm is applied to the feature vectors f to learn a set of two classes, denoted as $class_0$ and $class_1$, which we have found in [7] to correspond to "reliable" and "unreliable" classes.

The trends in $class_0$ and $class_1$ are illustrated in Figure 1, which analyzes the concentration of BPCs belonging to the two classes. The figure shows that most of the reliable BPCs, i.e., nasals, vowels and semi-vowels, belong to $class_0$. However, typical unreliable classes such as closures, silence, and weak-fricatives, have a higher concentration in $class_1$. After a set of two classes is learned, the FLDA is then used to learn a linear projection w. The projection vector is then applied to a newly hypothesized BPC feature vector to produce a single acoustic confidence score, namely $F_{score} = w^T f$.

B. Detecting Island Regions

After confidence scores are defined for each hypothesized BPC, an appropriate confidence threshold to accept the BPC as a reliable island region must be determined. Ideally, we would like island regions to include reliable BPCs, that is vowels, semivowels and nasals. Furthermore, we would like transitions between islands and gaps to occur at true boundaries between reliable/unreliable BPCs in the utterance, but would like to minimize the transitions that occur in the middle of sequences of reliable or unreliable BPCs. Thus, we define our goal of detecting reliable BPCs as those hypothesized BPCs that provide a high probability of detecting the true reliable/unreliable transitions with a low false alarm probability.

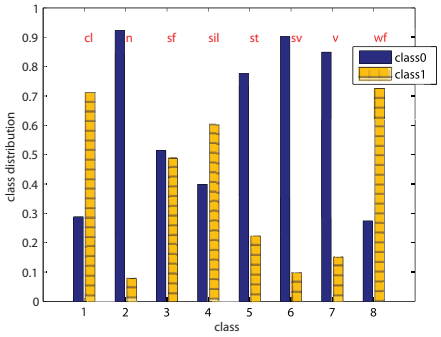

Fig. 1. Distribution of BPCs belonging to $class_0$ and $class_1$

To find an appropriate confidence threshold, we calculate a Receiver Operating Characteristic curve, a common tool used to find a suitable tradeoff between a high detection and low false alarm probability as the confidence threshold setting is varied. After an appropriate setting is determined to define island regions, we then use this information in our island-driven search methods. In Section III we discuss a method to prune the search space while in Section IV we explore a technique to reduce computation time during model scoring.

III. Island-Driven Segmentation

Segment-based recognizers [4] can often be computationally expensive, as the size of the search space and number of segmentations can grow as speech is subjected to noisier environments [3]. Therefore, we explore a method known as "segmentation by recognition" to prune the segment graph. Segmentation by recognition has previously been explored (i.e., [8]) without island/gap knowledge as a means of producing a smaller segment graph with more meaningful segments. In this method, a set of acoustic landmarks, representing potential transitions between phonemes, are first placed at regions of spectral change in the speech signal. The landmarks are then connected together to create a segment network. Then, a forward *phonetic* Viterbi search is performed over this segment graph to produce a phonetic lattice, after which a backwards A^* search is carried out on this lattice to produce an *N*-best list of phonemes. This *N*-best list is then converted into a new pruned *N*-best segment graph. A second-pass *word* recognition is then performed over this new segment graph.

Segmentation by recognition offers a few attractions. First, the pruned segment graph is produced from phonetic recognition and therefore the segments are much better aligned to the

phonemes hypothesized during word recognition. Second, the segment graph is much smaller, thus reducing the chances of throwing away potentially good paths. In this work, we explore segmentation by recognition using island/gap knowledge.

More specifically, we first use the BPCs to define a set of island/gap regions as presented in Section 2. Island/gap knowledge is then used to chunk an utterance into smaller sections at islands of reliability, allowing us to vary the number of segments in island *vs.* gap regions. In each island region, a forward phonetic Viterbi search is done to produce a phonetic lattice. A backwards A^* search over this lattice then generates a smaller list of *N*-best segments, after which a new pruned segment graph is created in the island regions. Here N, the number of allowed paths, is chosen to optimize recognition performance on a held out development set.

Next, the pruned segment graphs in the island regions are used to influence segment pruning in the gap regions. More specifically, another forward Viterbi/backward A^* is performed across each gap-island-gap region. Here the pruned island segment graph from the island pruning is inserted in the island regions. Again, N is chosen to optimize performance on the development set. We chose N in the gap regions to be smaller than the N chosen in the island regions to allow for fewer segments in less confident gap regions and more detailed segments in reliable island regions.

Finally, the *N*-best segments from the island and gap regions are combined to form a pruned segment graph. Then, given the new segmentation by recognition graph, a second-pass full word recognition is done over this pruned search space. We will refer to this segment-pruning technique described above as an island-driven segmentation, as fewer segments are permitted in areas of reliability and denser segmentation is allowed during regions of less confidence.

IV. Island Information for Model Evaluation

In this section, we explore the utilization of island/gap regions to further differentiate between the search effort in islands *vs.* gaps, by scoring less detailed phonetic models in gap regions and more detailed models in island regions. For example, the Aurora-2 corpus [9] contains 28 phones, and therefore effectively scores 157 diphone acoustic models (after clustering) for each possible segment. If less detailed BPC models are scored for each segment, this can reduce the number of acoustic models to approximately 49, roughly one-third. In order to implement this joint BPC/phonetic recognizer, we make changes to both the Finite State Transducer (FST) search space and acoustic model scoring phase, discussed below.

A. Finite State Transducer Formulation

The SUMMIT recognizer utilizes an FST framework [10] to represent the search space. In order to allow for BPC models in the search space, we represent the FST network R as being composed of the following components:

$$R = C \circ B \circ P \circ L \circ G \tag{1}$$

C typically represents the mapping from context-dependent (CD) phonetic labels to context-independent (CI) phonetic labels. Our CD labels include both phonetic and BPC labels, so C now represents the mapping from CD joint BPC/phonetic labels to CI BPC/phonetic labels. We next compose C with B, which represents a mapping from joint CI BPC/phonetic labels to CI phonetic labels. The rest of the composition is standard, with P representing the phonological rules, L the word lexicon and G the grammar. Thus, the full composition R maps input context-dependent BPC/phonetic labels directly to word strings. Therefore each word in the lexicon is represented as a combination of BPC and phoneme sub-word units.

B. Acoustic Model

The acoustic model calculates the probability of an observation o_t given sub-word unit u_n as $P(o_t|u_n)$. In island regions, the sub-word unit u_n is a context-dependent phonetic model Phn and the acoustic model is scored as $P(o_t|Phn)$ for each Phn. In the gap region, the sub-word unit is a context-dependent BPC model, BPC. We calculate $P(o_t|BPC)$ by taking the average of all the phonetic model scores which make up the BPC. The expression for the BPC acoustic model score is given more explicitly by Equation 2. Here M is the number of Phn models which belong to a specific BPC. Details on the justification of this approach for scoring BPC models can be found in [7].

$$P(o_t|BPC) = \frac{1}{M}\left(\sum_{Phn \in BPC} P(o_t|Phn)\right) \tag{2}$$

V. Experiments

Island-driven search experiments are first conducted on the small vocabulary Aurora-2 corpus [9]. This task consists of clean TI-digit utterances with artificially added noise at levels of -5db to 20db. We utilize this corpus because of its simple nature, which allows us to explore the behavior of the proposed island-driven search techniques in noisy conditions. Results are reported on Test Set A, which contains noise types similar to those in the training data, namely subway, babble, car, and exhibition hall noise. For word recognition experiments, global multi-style diphone acoustic models are used. Acoustic models are trained specific to each segmentation investigated, namely the baseline spectral change segmentation in SUMMIT [4], a BPC segmentation method presented in [3] which has been shown to be robust in noisy conditions, and the proposed island-driven segmentation techniques.

Experiments are then conducted on the CSAIL-info corpus, which contains information about people, rooms, and events in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. The large vocabulary nature of the task, coupled with the various non-stationary noises which contaminate the speech utterances, motivate us to explore island techniques on this task. Results are reported on the development and test sets. For word recognition experiments, diphone acoustic models are trained using only the spectral change method on data collected from the telephone-based weather system [10].

A variety of experiments are conducted on both corpora to analyze the behavior of the proposed island-driven strategy. First, we explore the robustness of the technique discussed in Section 2 to identify islands and gaps. Second, we analyze the word error rate (WER) of the island-based segment pruning and joint BPC/phonetic model scoring methods. Third, the computational benefits of the island methods are investigated.

VI. Results on Aurora

A. Island Quality Investigation

First, we investigate the robustness of the technique to hypothesize islands and gaps proposed in Section II. Ideally, a robust island will have a high concentration of vowels, semi-vowels and nasals, which correspond to more reliable, robust parts of the speech signal. Figure 2 illustrates for each phoneme in the digit "zero", the distribution of islands and gaps within that phoneme. The distribution is normalized across each phoneme, so, for example a distribution of 0.3 in the island region for /z/ in "zero" means that 30% of the time /z/ is present in an island region and 70% of the time it is contained in a gap region. The plot indicates that most of the vowels and semi-vowels in the word, containing the information-bearing parts of the signal, are concentrated in the island regions. However, most of the non-harmonic classes belong to the gap regions. This trend was observed for all eleven digits in the Aurora-2 task.

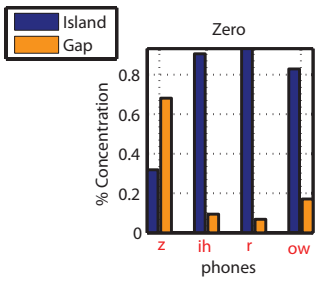

Fig. 2. Phoneme Concentration of Islands/Gaps in the digit "zero"

B. Performance of Island-Driven Techniques

1) Island-Based Segment Pruning: Second, to explore the behavior of the island-based segment pruning method, Table I compares the WER of this approach to the spectral change and BPC segmentation methods. The results are averaged across all noise conditions in Test Set A. Table I indicates that the island segmentation method has the lowest error rate, and a Matched Pairs Sentence Segment Word Error (MPSSWE) significance test indicates that the island segmentation is statistically significant from the other two approaches. These results verify that recognition results can be improved by using the island/gap regions to reduce the segmentation graph and keeping the most promising segments.

2) Joint BPC/Phonetic Model Scoring: Third, we explore the benefit of scoring BPC models in less reliable gap regions. The first question explored is how many BPC models should be scored in gap areas? Figure 3 shows the WER on the development set for the joint BPC/phonetic method as the number of BPC models is varied. Here, the additional BPC chosen at each point on the graph is picked to give the maximum decrease in WER. We also analyze the WER when phonetic models are scored in both regions, as indicated by the flat line in the figure.

TABLE I
WER for Segmentation Methods on Aurora-2 Test Set A

Segmentation Method	WER
Baseline Spectral Change Segmentation	31.9
BPC Segmentation Baseline	22.8
Island-Based Segmentation	**22.3**

Point A in the figure corresponds to the location where the WER of the joint BPC/phonetic approach equals that of the phonetic approach. This point corresponds to the following 8 BPCs: silence, vowel, semi-vowel, nasal, closure, stop, weak fricative, strong fricative. If the number of BPCs is increased, and in particular both strong and weak fricatives are split into voiced and unvoiced classes, the WER continues to decrease. This best set of BPC models is depicted by *Point B* in Figure 3. There is no extra benefit to increasing the number of BPCs past 10, as illustrated by the increase in WER.

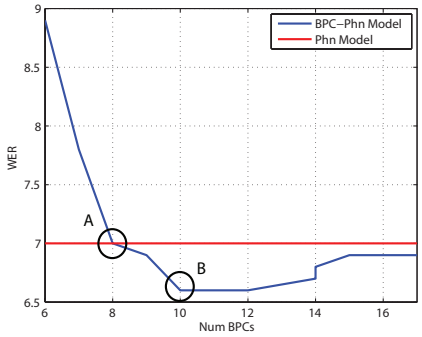

Fig. 3. WER vs. Number of BPC Models when joint BPC/phonetic models are scored vs. scoring only phonetic models

Using these 10 BPC models to score the gap regions, Table II compares the WER when only phonetic models are scored *vs.* scoring BPC/phonetic models. Notice that there is an improvement when BPC models are scored in gap regions, showing that performing a less detailed evaluation in unreliable regions does not lead to a degradation in performance.

TABLE II
WER for Island-Based Methods on Aurora-2 Test Set A

Scoring Method	WER
Island-Based Seg, Phonetic Models	22.3
Island-Based Seg, BPC/Phonetic Models	**22.1**

3) Error Analysis: To better understand the improvement in error rate offered by the island-driven techniques, Table III breaks down the WER for the BPC segmentation, island segmentation method scoring phonetic models and the island segmentation method scoring joint BPC/phonetic models, and lists the corresponding substitution, deletion and insertion rates. Notice that the main advantage to the island based approach is the large decrease in insertion rate.

TABLE III
BREAKDOWN OF ERROR RATES ON AURORA-2 TEST SET A

Method	WER	Subs	Del	Ins
BPC Segmentation	22.8	9.9	6.8	6.1
Island Seg, Phn Models	22.3	10.8	7.6	3.9
Island Seg, BPC/Phn Models	**22.1**	**11.1**	**8.0**	**3.0**

A closer investigation of these insertion errors is illustrated in the top panel of Figure 4, which displays the number of insertion errors for the three methods, when errors occur purely in islands, gaps, or span over a combined island&gap region. In addition, the bottom panel of Figure 4 illustrates the relative reduction in insertion errors over the BPC segmentation method. Notice that most of the insertions occur in gap only and island&gap regions where the signal is less reliable compared to pure island areas. In addition, the biggest reduction in insertions occur in gap only regions, showing one of the strengths of island-driven search. Having a detailed segmentation and phonetic model scoring in unreliable regions can throw the search astray without taking into account future reliable areas, resulting in large insertion errors.

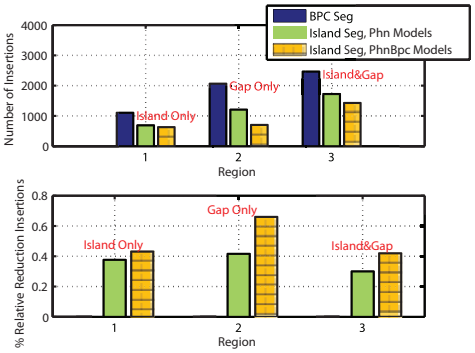

Fig. 4. Insertion Errors in Various Regions

C. Computational Efficiencies

In this section, we explore the computational efficiencies of the island-based approach. First, we compare the Viterbi path extensions for the BPC segmentation and island segmentation approaches, calculated by counting the number of paths extended by the Viterbi search through the length of the utterance. Figure 5 shows a histogram of the Viterbi extensions on all utterances in Test Set A for the two approaches. Notice that the island segmentation extends fewer paths and has an average path extension of about 9.5 (in ln scale), compared to the BPC segmentation which extends roughly 10.4 paths.

In addition, to evaluate the benefit in computational effort with the joint BPC/phonetic approach, we explore the number of models requested by the search during recognition. Every time paths are extended, the search requests a set of models to extend these paths. The number of models evaluated per utterance is computed by calculating the total number of models requested through the length of an utterance. Figure 6 illustrates a histogram of the number of models evaluated (in ln scale) for all utterances in Test Set A, in both the island and gap regions. The joint BPC/phonetic method is much more

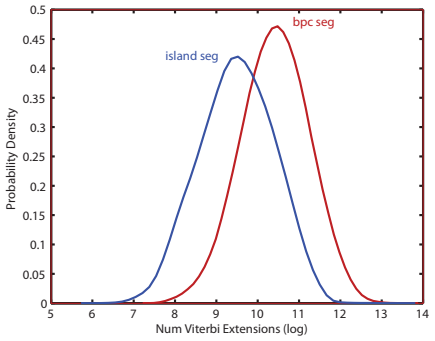

Fig. 5. Histogram of Number of Viterbi Extensions (ln scale)

efficient, particularly in the gap region, and evaluates fewer models compared to the phonetic method.

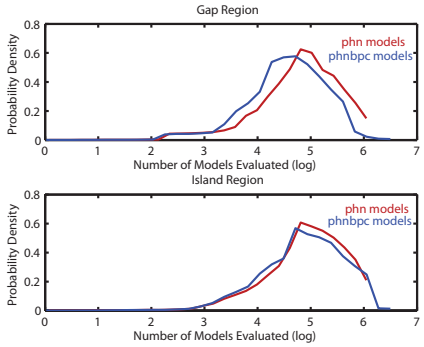

Fig. 6. Histogram of No. of Models Evaluated in Islands and Gaps

VII. RESULTS ON CSAIL-INFO

A. Island Quality Analysis

First, we explore the quality of the island detection technique. It has been suggested that stressed syllables in English carry more acoustically discriminatory information than their unstressed counterparts and therefore provide islands of reliability [1]. To analyze the behavior of stressed syllables, the vocabulary in the CSAIL-info corpus was labeled with stress markings, obtained from the IPA stress markings in the Merriam-Webster dictionary. It has also been shown that identifying stressed syllables from nucleus vowels offers more reliability than also using stress information for non-vowel segments. Thus, we explore using the BPC island-detection technique discussed in Section III such that islands are identified to maximize the detection of true stressed vowels.

First, we analyze the distribution of just stressed vowels in islands and gaps. Figure 7 shows the distribution of stressed vowels per utterance in the island and gap regions. Specifically, the figure indicates for a given % of stressed vowels per utterance (x-axis), the % of these stressed vowels found solely in island regions (y-axis). The graph illustrates that a significantly higher number of stressed vowels, in fact 84% on average, appear in island regions compared to gaps. Furthermore, because stressed vowels should ideally represent stable portions of the signal, they should also be recognized

with high probability. In [7], we observed that approximately 84% of the stressed vowels found in island regions are correct. Thus, we can conclude that most of the information-bearing parts of the signal are found in the island regions, and also that most of these stressed vowels are correctly hypothesized.

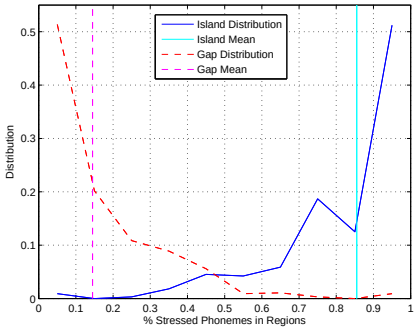

Fig. 7. Distribution of Stressed Vowels in Islands and Gaps

B. Performance of Island-Driven Techniques

1) Island-Based Segment Pruning: Table V shows the results for the three segmentation techniques. The island-based technique has slightly worse performance than the BPC segmentation method, though a MPSSWE significance test indicates that these two methods are not statistically significant. However, the island method still offers similar computational benefits, as discussed in Section VI-C, over the BPC approach.

TABLE IV
WER FOR SEGMENTATION TECHNIQUES ON CSAIL-INFO TASK

Method	WER (dev)	WER (test)
Spectral Change Seg	26.5	28.6
BPC Seg	**24.3**	**27.6**
Island-Based Seg - Broad Classes	24.8	28.2

One hypothesis for the slight deterioration in performance in the island-driven technique is that acoustic models are trained on a weather domain system [10] using the spectral segmentation method, which behaves more similarly to the BPC segmentation technique compared to the island-based approach. We have observed in the Aurora-2 task that retraining acoustic models specific to each segmentation method offered improvements in recognition accuracy. However, due to the limited data in the CSAIL-info training set, better performance was found using Jupiter acoustic models, rather than training acoustic models specific to each segmentation.

2) Joint BPC/Phonetic Model Scoring: Next, we explore the performance of the joint BPC/phonetic approach on the CSAIL-info task, which is shown in Table V for various BPC splits. First, notice that using noise and nasal BPCs leads to a slight improvement in performance on the development set but not the test set. However, as the number of clusters is increased past the nasal class, the error rate increases. Because of the large scale nature of the CSAIL-info task, scoring less detailed BPC models increases the confusability among words. For example, consider the words "bat" and "pat", which have the same BPC transcription. To address this issue, in the future, we would like to consider exploring a lexical access technique, where a first pass recognition is performed to determine an N-best list of BPC/phonetic hypotheses, after which a second-pass word recognition is done over this cohort of words.

TABLE V
WER FOR DIFFERENT BPCS IN GAP REGIONS ON CSAIL-INFO

BPCs	WER (dev)	WER (test)
No BPCs - Phonetic Models	24.8	**28.2**
Noise (Laughter, Cough, Babble)	**24.7**	28.9
+Nasal	24.8	
+Alveolar+Labial+Dental Closures	25.1	
+Voiced+Unvoiced Stops	25.2	
+Voiced+Unvoiced Weak Frics	25.5	

VIII. CONCLUSIONS

In this paper, we explored an island-driven search method which we incorporated into an ASR framework. More specifically, we utilized BPC information to identify a set of island and gap regions. We illustrated that this proposed method to identify islands was able to identify information-bearing parts of the signal with high probability. On the Aurora-2 noisy digits task, we demonstrated that utilizing island/gap information to prune the segmentation graph and to score fewer models in gaps resulted in improvements in both performance and computation time. Furthermore, on the CSAIL-info task, we showed that utilizing island information for segment pruning offered comparable performance to the BPC segmentation approach, though further utilization of BPC knowledge in gap regions during final search resulted in a slight degradation in performance. In the future, we would like to explore a bidirectional island-driven search strategy, as well as other techniques to detect islands from the input signal.

IX. ACKNOWLEDGEMENTS

Thank you to Victor Zue for helpful discussion in shaping this work. This work was sponsored by the Office of Secretary of Defense under Air Force Contract FA8721-05-C-0002.

REFERENCES

[1] W. A. Lea, *Trends in Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1980.
[2] R. Kumaran, J. Bilmes, and K. Kirchhoff, "Attention Shift Decoding for Conversational Speech Recognition," in *Proc. Interspeech*, 2007.
[3] T. N. Sainath and V. W. Zue, "A Comparison of Broad Phonetic and Acoustic Units for Noise Robust Segment-Based Phonetic Recognition," in *Proc. Interspeech*, 2008.
[4] J. Glass, "A Probabilistic Framework for Segment-Based Speech Recognition," *Computer Speech and Language*, vol. 17, no. 2-3, 2003.
[5] T. N. Sainath, D. Kanevsky, and B. Ramabhadran, "Broad Phonetic Class Recognition in a Hidden Markov Model Framework using Extended Baum-Welch Transformations," in *Proc. ASRU*, 2007.
[6] S. Kamppari and T. Hazen, "Word and Phone Level Acoustic Confidence Scoring," in *Proc. ICASSP*, 2000.
[7] T. N. Sainath, "Applications of Broad Class Knowledge for Noise Robust Speech Recognition," Ph.D. dissertation, MIT, 2009.
[8] S. C. Lee and J. Glass, "Real Time Probabilistic Segmentation for Segment-Based Speech Recognition," in *Proc. ICSLP*, 1998.
[9] H. G. Hirsch and D. Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Condidions," in *ISCA ITRW ASR2000 "Automatic Speech Recognition: Challenges for the Next Millennium"*, 2000.
[10] J. Glass, T. Hazen, and I. Hetherington, "Real-time Telephone-Based Speech Recognition in the JUPITER Domain," in *Proc. ICASSP*, 1999.