Submission Cover Sheet for 1999 IEEE SoutheastCon

Submission Type (Please Check One):

☐ *Full Length Paper - (Due 11/1/98)* ☑ *Concise Paper - (Due 12/1/98)*

Title of manuscript:
FAST SEARCH ALGORITHMS FOR CONTINUOUS SPEECH RECOGNITION

Authors and affiliations (List authors in the order to appear on the manuscript. List affiliations (companies, universities, etc.) and addresses exactly as they should appear. If there are more than 4 authors, please attach a second sheet.

1. Name: **Jie Zhao**
Affiliation: **Mississippi State University**
Address: **PO Box 9571**
Mississippi State, MS 39762

2. Name: **Jonathan Hamaker**
Affiliation: **Mississippi State University**
Address: **PO Box 9571**
Mississippi State, MS 39762

3. Name: **Neeraj Deshmukh**
Affiliation: **Mississippi State University**
Address: **PO Box 9571**
Mississippi State, MS 39762

4. Name: **Aravind Ganapathiraju**
Affiliation: **Mississippi State University**
Address: **PO Box 9571**
Mississippi State, MS 39762

Corresponding Author:
Name: **Jie Zhao**
Address: **PO Box 9571**
Mississippi State, MS 39762

Telephone: **601-325-8335**
FAX: **601-325-3149**
E-mail: **zhao@isip.msstate.edu**

Key Words: **Speech Recognition, Search, Fast-match**

For Technical Committee Use Only -- Do Not Write Below

Date Received: ____ Manuscript # Assigned ____

Author Notification Date: ____ Date Camera Ready Received: ____

Decision: ____ Session: ____

Notes: ____

Submission Cover Sheet for 1999 IEEE SoutheastCon

Submission Type (Please Check One):

☐ *Full Length Paper - (Due 11/1/98)* ☑ *Concise Paper - (Due 12/1/98)*

Title of manuscript:
FAST SEARCH ALGORITHMS FOR CONTINUOUS SPEECH RECOGNITION

Authors and affiliations (List authors in the order to appear on the manuscript. List affiliations (companies, universities, etc.) and addresses exactly as they should appear. If there are more than 4 authors, please attach a second sheet.

5. Name **Dr. Joseph Picone**
Affiliation **Mississippi State University**
Address **PO Box 9571**
Mississippi State, MS 39762

6. Name
Affiliation
Address

Corresponding Author:
Name: **Jie Zhao**
Address: **PO Box 9571**
Mississippi State, MS 39762

Telephone: **601-325-8335**
FAX: **601-325-3149**
E-mail: **zhao@isip.msstate.edu**

Key Words: **Speech Recognition, Search, Fast-match**

For Technical Committee Use Only -- Do Not Write Below			
Date Received:		Manuscript # Assigned	
Author Notification Date:		Date Camera Ready Received:	
Decision:		Session:	

Notes:

FAST SEARCH ALGORITHMS FOR CONTINUOUS SPEECH RECOGNITION

J. Zhao, J. Hamaker, N. Deshmukh, A. Ganapathiraju, J. Picone
Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
Mississippi State, Mississippi 39762, USA
Ph (601) 325-8335 - Fax (601) 325-3149
{zhao, hamaker, deshmukh, ganapath, picone}@isip.msstate.edu

ABSTRACT

The most important component of a state-of-the-art speech recognition system is the decoder, or search engine. Given this importance, it is no surprise that many algorithms have been devised which attempt to increase the efficiency of the search process while maintaining the quality of the recognition hypotheses. In this work we present one such algorithm which uses a two-pass fast-match search to effectively prune away unlikely parts of the search space. This system is compared to a state-of-the-art Viterbi decoder with beam pruning in evaluations on the OGI Alphadigits Corpus.

FAST SEARCH ALGORITHMS FOR CONTINUOUS SPEECH RECOGNITION

SUMMARY

In recent years we have seen speech recognition technology advance to the forefront of commercial applications. Typically these systems are restricted to a particular domain such as automatic dictation or command-and-control applications. With these restrictions, developers are able to create highly efficient systems which run in real-time with very low error rates. However, the primary goal of speech research is to produce systems that allow users to interact naturally without restrictions to either content or style of speech. Unfortunately, the resources required for a conversational speech recognition system to be commercially viable are far beyond the hardware currently available to consumers.

The majority of this resource consumption is owed to the search process inherent in finding the string of words spoken. The decoder searches through every possible word path to find the most likely string of words according to statistical models of speech. The Viterbi search algorithm is used at the core of most state-of-the-art decoders, but the search space for this algorithm is too large even in speech recognition tasks of moderate complexity. Thus, there is a need for algorithms that can intelligently limit the search space while not affecting the word error rate (WER).

Most state-of-the-art speech recognition systems use pruning techniques to reduce the search space. In these, a threshold is set at each level in the search where only paths whose score falls above that threshold are extended to the next level, pruning away all others. In this work we explore a fast-match technique which uses a two-pass decoding strategy. The first pass quickly finds an approximate solution by applying a simple heuristic at each level of the search. The second pass uses the knowledge gained from the first pass to perform a more detailed search. The art to this type of algorithm is determining the heuristic which can find a high quality partial solution using very limited resources.

As a first try we implemented a heuristic which extends only the N-best paths at each point through the search. At the end of this first pass, a best overall hypothesis is found and the partial path score at every level of the hypothesis is then used as the threshold for the second pass of the search. We compared this algorithm against Viterbi with beam pruning on a small subset of the OGI Alphadigits Corpus and found that the fast-match search produced a much tighter beam with little effect on WER. The fast-match scheme gave a second pass which was significantly faster than Viterbi with beam pruning. However, the overhead required to find the thresholds in the first pass was substantial.

This work is currently being extended to include heuristics which provide a more efficient overall search process. An example which has given encouraging preliminary results is the use of simple monophone acoustic models for the first pass of the search and context-dependent triphone models for the second pass. We are also verifying our initial Alphadigit results on the far more complex SWITCHBOARD task.