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:
IMPLEMENTATION AND ANALYSIS OF SPEECH RECOGNITION FRONT-ENDS

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: **Vishwanath Mantha**
Affiliation: **Mississippi State University**
Address: **PO Box 9571**
Mississippi State, MS 39762

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

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

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

Corresponding Author:
Name: **Vishwanath Mantha**
Address: **PO Box 9571**
Mississippi State, MS 39762

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

Key Words: **digital signal processing, speech recognition, front-end**

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:
IMPLEMENTATION AND ANALYSIS OF SPEECH RECOGNITION FRONT-ENDS

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 **Aravind Ganapathiraju**
Affiliation **Mississippi State University**
Address **PO Box 9571**
Mississippi State, MS 39762

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

Corresponding Author:
Name: **Vishwanath Mantha**
Address: **PO Box 9571**
Mississippi State, MS 39762

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

Key Words: **digitial signal processing, speech recognition, front-end**

For Technical Committee Use Only -- Do Not Write Below

Date Received: _____ Manuscript # Assigned _____

Author Notification Date: _____ Date Camera Ready Received: _____

Decision: _____ Session: _____

Notes: _____

IMPLEMENTATION AND ANALYSIS OF SPEECH RECOGNITION FRONT-ENDS

V. Mantha, R. Duncan, Y. Wu, J. Zhao, A. Ganapathiraju, J. Picone
Institute for Signal and Information Processing
Mississippi State University
Mississippi State, Mississippi 39762, USA
Ph (601) 325-3149 - Fax (601) 325-3149
{mantha, duncan, wu, zhao, ganapath, picone}@isip.msstate.edu

ABSTRACT

We have developed a comprehensive front-end module integrating several signal modeling algorithms common to state-of-the-art speech recognition systems. The algorithms presented in this work include mel frequency cepstra, perceptual linear prediction, filter bank amplitudes, and delta features. The framework for the front-end system was carefully designed to ensure simple integration into speech processing software. The modular design of the software along with an intuitive GUI provide a powerful tutorial by allowing a student of speech processing to easily interchange algorithms and vary every aspect of each model parameter. The software is written in a tutorial fashion, with a direct correlation between algorithmic lines of code and equations in the technical paper. The effectiveness of the different front-end algorithms is also being evaluated on a common set of speech data.

IMPLEMENTATION AND ANALYSIS OF SPEECH RECOGNITION FRONT-ENDS

SUMMARY

Current speech recognition technology requires that the speech signal first be processed into observation vectors representing events in the probability space. This process, known as signal modeling, is the function of the front-end module. These observation vectors are typically closely related to important aspects of the human hearing process (e.g. the frequency warping that takes place in the auditory canal). Using these acoustic observation vectors and some language constraints, a network search algorithm finds the most probable sequence of events and hypothesizes the textual content of the audio signal.

The field of speech recognition research is currently hindered by a lack of public domain software. To facilitate smaller entry costs for research groups in this field, the Institute for Signal and Information Processing (ISIP) is developing a freely available state-of-the-art speech recognition tool-kit. This paper describes the development of a comprehensive front-end module which is an integral part of the ISIP speech recognition system but remains flexible enough to be used in a wide variety of signal modeling applications.

The first step in speech signal modeling is to use digital filter banks, the Fourier transform, or linear prediction to estimate the spectrum of the signal. Feature coefficients are derived from the spectrum by calculating the filter bank amplitudes and the mel-scaled cepstrum.The perceptual linear prediction (PLP) algorithm more closely relates these spectral features to auditory psychophysical concepts such as critical-band spectral estimation, the equal loudness curve, and the intensity-loudness power law. From these first-order features differential coefficients are also used to model temporal variation in the signal.

All software for this front-end module was developed in C++ using the public-domain GNU compiler. Our software is comprehensive, allowing the user complete control over all aspects of the signal modeling process. This includes algorithm selection, frame and window duration, and internal parameters. A Tcl-Tk based graphical user interface (GUI) is also available to facilitate user interaction with the numerous parameters. The GUI allows the user to vary different modeling parameters and study the effect on the output observations. It also assists in the comparison of different algorithms on the same data.

The final aspect of this project is to evaluate the effectiveness of each algorithm. Our evaluation performs frame-level classification experiments on a reduced phone set of the OGI Alphadigits Corpus. Each phone is modeled by extracting observation vectors corresponding to multiple instances of the phone across the corpus. The location of these frames within the speech file is determined by looking at time aligned state level transcriptions automatically obtained through forced alignments. We use Support Vector Machines (SVMs) and decision trees to assist in classification. By comparing the output of this classifier and the reference information, we can evaluate the performance of each front-end algorithm.