COMPARISON OF ENERGY-BASED ENDPOINT DETECTORS FOR SPEECH SIGNAL PROCESSING

by

A. Ganapathiraju, L. Webster, J. Trimble, K. Bush, P. Kornman

Department of Electrical and Computer Engineering
Mississippi State University
{ganapath,webster,trimble,bush,kornman}@isip.msstate.edu

Accurate endpoint detection is a necessary capability for construction of speech databases from field recordings. In this paper we describe the implementation of two endpoint detection algorithms which use signal features based on energy and rate of zero crossings. We have made extensive use of object-oriented concepts and data-driven programming to make our code re-usable for a variety of applications, including speech recognition. A uniform user-interface for both algorithms has been developed using a novel virtual classmethodology.We also present a comparison of the two algorithms using an objective evaluation paradigm we have developed. A small locally prepared database has been used for the purpose of evaluation.

A copy of this presentation is available at:

http://www.isip.msstate.edu/publications/1996/secon'96/endpt.fm

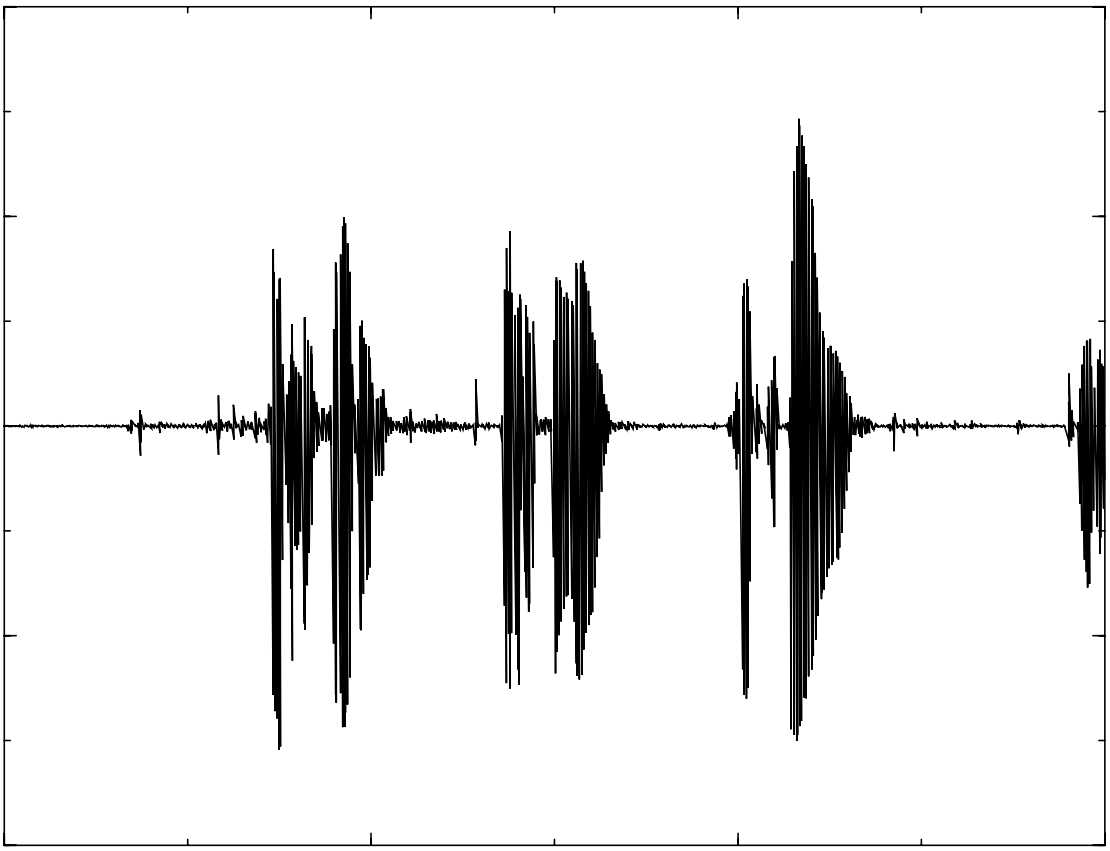
MOTIVATION FOR AN ENDPOINT DETECTOR

❒ In early stages of speech recognition research, used for the alignment process in DTW isolated word recognizers (ISR)

❒ Still used extensively today as data trimming tools in the preparation of speech databases.

- Large databases have become imperative for training the large vocabulary continuous speech recognizers.
- Telephone data collection is the easiest and fastest way to collect data for large databases.
- A major problem associated with this approach is that the characteristics of speech and noise are not very different in such environments.
- To make available as public domain software an accurate and simple tool for endpoint detection

THE ENDPOINT DETECTION PROBLEM



- Low energy fricatives
- End of word plosives
- Intraword energy cutoff etc.

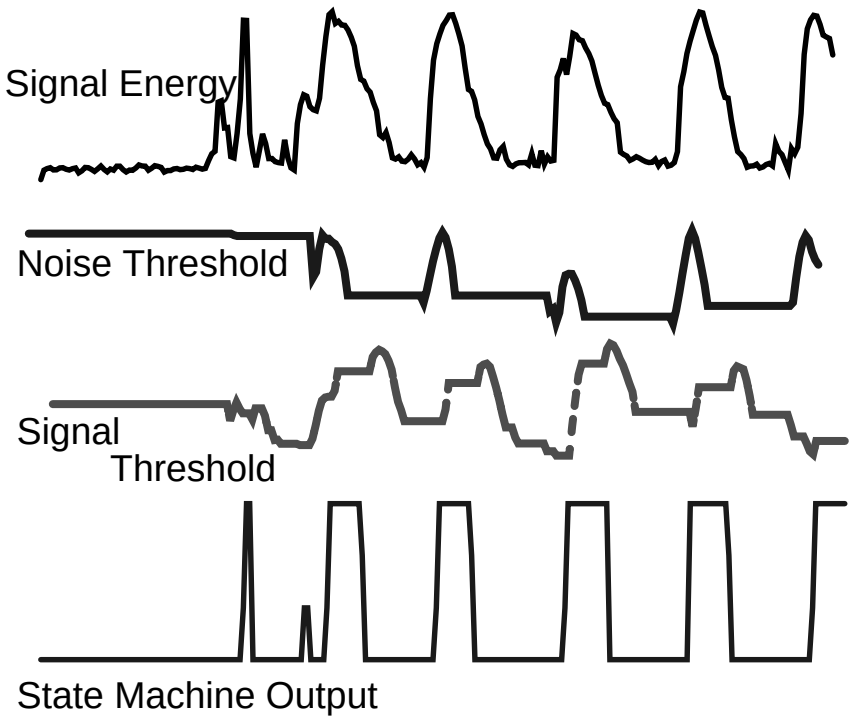
ALGORITHMS

- Available algorithms for endpoint detection are based on:
  - Energy
  - Zero-Crossings
  - Periodicity Measure
  - Hidden Markov Models
  - Spectral Slope
- Most commonly used algorithms are energy and zero-crossing rate
  - Simplicity of algorithm implementation
  - Number of parameters needed to be tweaked
  - User can easily adapt the system to any application or environment

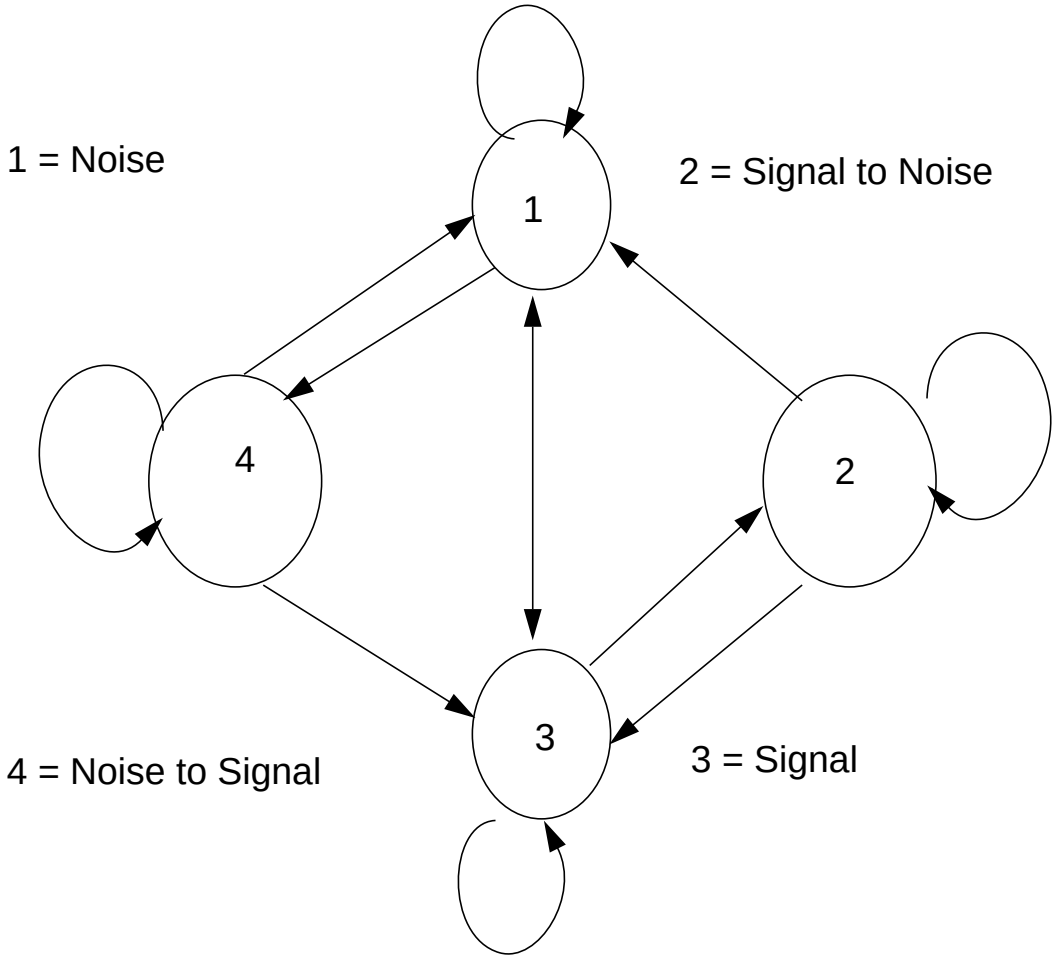
ALGORITHM DESCRIPTION

❒ **Energy Algorithm**:

- Preprocessing operations include preemphasis, windowing, normalization performed on input data
- State machine created . Parameters updated dynamically
- Make a decision on endpoints based on the state machine output



STATE MACHINE GENERATION



❒ **Energy and Zero-crossing Algorithm**:

- Zero-crossing rate used only for refining the endpoints

- Zero-crossing rate activity gives a more reliable estimate in cases where energy level is too low for endpoint detection

\

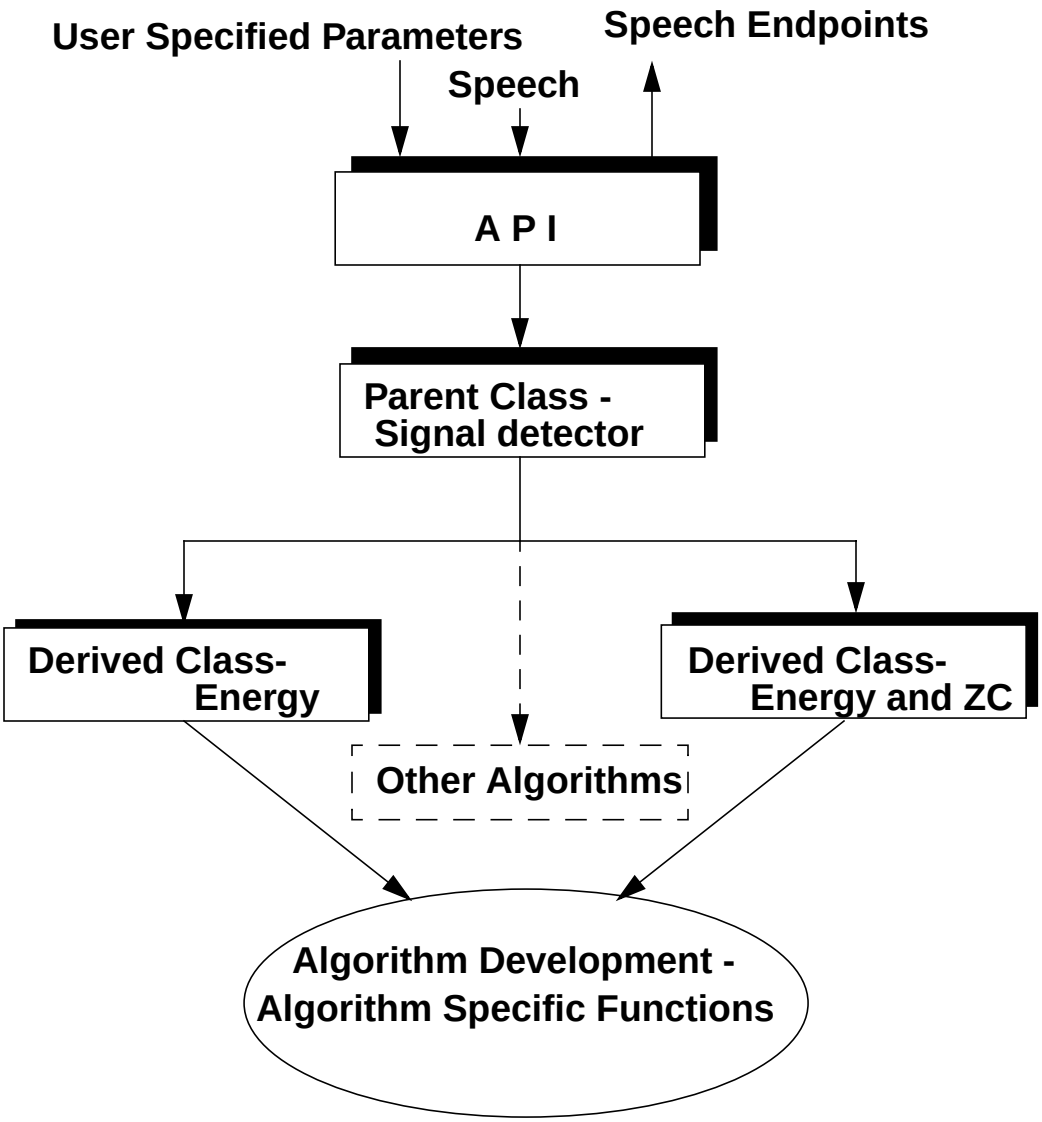
- Initial estimate of endpoints got from the energy algorithm

- Backtrack for 15 frames and if zero-crossing activity is present in more than 10 frames endpoints are shifted to first occurrence of activity

SOFTWARE STRUCTURE

- Structured implementation of the algorithms required for further improvements and additional research
- Data-driven programming concepts used for making code flexible to suit the requirements of the user
- Implementation specifics hidden from the user while giving him/her the control over the algorithm choice and the choice of the parameters
- Modular code structure to allow for any modifications in existing functions
- Keeping in mind the above facts C++ was chosen because of its applicability to the problem at hand

SOFTWARE STRUCTURE SCHEMATIC



SALIENT FEATURES OF THE CODE

- A circular buffer is used to store the energy and zero-crossing information.This has two advantages:
  - Makes the code capable of performing real-time processing
  - Helps simplify the smoothing and filtering operations
- To manage the memory requirement, many smoothing operations are performed
  - Long stretches of silence can be removed
  - Long stretches of signal can be removed
  - Similarly short bursts of noise and signal can be rejected
- Object Oriented Concepts used
  - Concepts of data-driven programming make the code extremely user controlled
  - Concept of Virtual Functions hides algorithm specific functions from the user

SAMPLE PARAMETER FILE

algorithm = simple energy

number_of_channels	= 2 channels	Signal Processing related parameters
sample_size	= 2 bytes	
channel_to_be_processed	= 0 channel	
sample_frequency	**= 8000.000 Hz**	
frame_duration	= 0.020 secs	
window_duration	= 0.030 secs	
preemphasis	= 0.950 units	
nominal_signal_level	**= -35.00 dB**	Signal and Noise related parameters
signal_adaptation_delta	= 15.00 dB	
signal_adaptation_constant	= 0.50 units	
nominal_noise_level	**= -60.00 dB**	
noise_adaptation_delta	= 15.00 dB	
noise_adaptation_constant	= 0.75 units	
noise_floor	**= -70.00 dB**	
utterance_delta	= 6.000 dB	
minimum_utterance_duration	**= 0.100 secs**	
maximum_utterance_duration	**= 50.00 secs**	

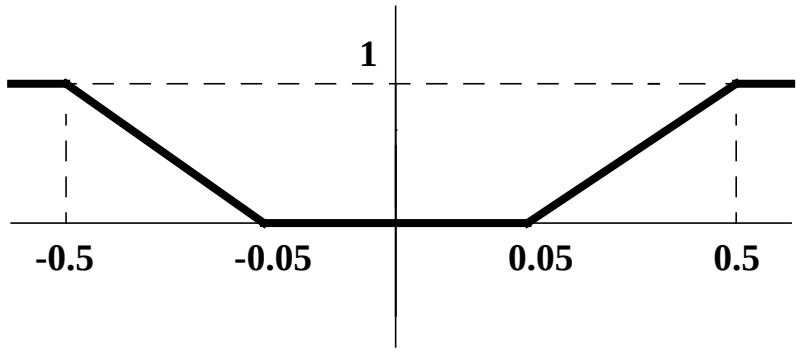
debug_level= 0 level

EXPERIMENTS

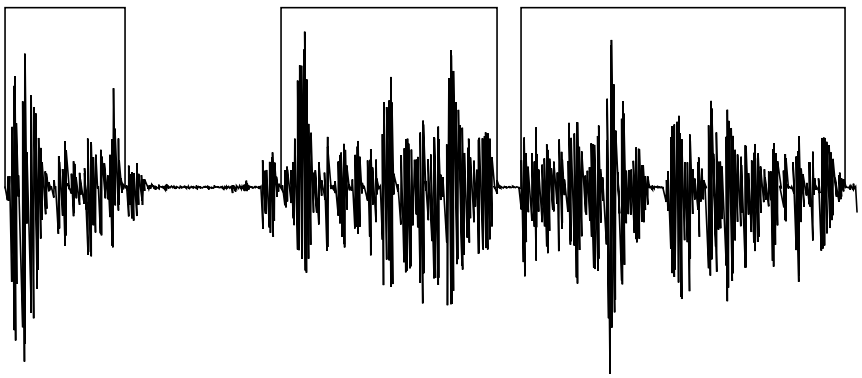
- **Database preparation**:
  - Test utterances recorded under typical office environment
  - Sampled at 16kHz and stored as 16bit integers
  - Data is composed of three male and three female speakers
  - The utterances are of different types like digits, phrases, short sentences and spontaneous speech.
  - On the average each file contains 10 utterances

File #	Data Type	Sample Data
1	single numbers	"zero" "two"
2	'teens'	"eighteen" "sixteen"
3	two word numbers	"65" "44"
4	very large number	"300,188"
5	phrases	"high definition television"
6	sentence pattern	"we hold these truths to be self-evident, that all men are created equal"
7	spontaneous speech	********** (approx. 5 sentences)

- **Objective Evaluation:**
  - An evaluation paradigm developed for comparison
  - Based on perception of speech by humans
  - Performance of both algorithms compared to a hand-marked database



SCORING FUNCTION



RESULTS

- Comparison of the overall performance of the two algorithms :

Algorithm	Subs.	Deletion	Insertion	Average
Energy	0.23	2.0	1.0	1.10
Energy and Zero-crossing	0.24	2.0	1.0	1.06

Table 1: COMPARISON OF ALGORITHMS

- Comparison of performance of algorithms on different types of utterances :

UTTERANCE	ENERGY	ENERGY & ZC	MUD (secs)	MUS (secs)
Isolated Digits	0.11	0.12	0.20	0.06
Teens	3.10	3.06	0.40	0.10
Multi-syl. Digits	0.12	0.11	0.30	0.10
Long Digit Strings	1.17	1.15	1.00	0.10
Phrases	0.24	0.28	1.00	0.10
Sentences	0.20	0.32	1.00	0.10
Spont. Speech	0.73	0.71	3.00	0.10

Table 2: PERFORMANCE BY UTTERANCE TYPE

CONCLUSIONS

- The effect of zero-crossing rate is not significant for endpoint detection in average SNR environments
- Software structure designed to make the code reusable
- The code is being presently used for preparation of a large database under the JEIDA project.
- The code structure will be used in the continuous speech recognizer we are presently working on
- Though present day CSR systems do not use an endpointer as an integral part, their use as important tools in database preparation will continue
- The first version of this code and the database is available via anonymous ftp at http://www.isip.msstate.edu/software

BIBLIOGRAPHY

1. B. Wheatley, et. al.,"Robust Automatic Time Alignment of Orthographic Transcriptions with Unconstrained Speech," *Proc. ICASSP 92,* Vol. 1, pp. 533-536, 1992.
2. J. Junqua et. al.,"A Robust Algorithm for Word Boundary Detection in Presence of Noise," *IEEE Trans. on Speech and Audio Processing*, Vol. 2, No. 3, pp. 406-412, July 1994.
3. L. Lamel et. al., "An Improved Endpoint Detector for Isolated Word Recognition," *IEEE Trans. Acoustics., Speech, Signal Processing,* Vol. 29 pp. 777-785, Aug. 1981.
4. L.Rabiner and M. Sambur, "An algorithm for determining the endpoints of isolated utterances," *Bell Syst. Tech. J.*, Vol. 54, pp. 297-315, 1975.
5. Minsoo Hahn et. al.,"An Improved Speech Detection Algorithm for Isolated Korean Utterances,"*Proc. ICASSP 92*, Vol. 1, pp. 525-528,1992.
6. L.Rabiner and R.Schafer,"*Digital Processing of Speech Signals,*" Englewood Cliffs, N.J., Prentice-Hall,1978.
7. J.Deller et. al.," *Discrete Time Processing of Speech Signals,*" N.Y., Macmillan, 1993.
8. B. Reaves,"Comments on an improved endpoint detector for isolated word recognition," *Corresp. IEEE Acoust.,Speech and Signal Processing,* Vol. 39, pp. 526-527, Feb. 1991.
9. R.Tucker, "Voice Activity Detection Using a Periodicity Measure," *IEE Proceedings, Part 1, Communications, Speech, Vision,* Vol. 139, pp. 377-380, Aug. 1992.