A Variable Rate Voice Coder using LPC-10E

J.P. Macker, R.B. Adamson

Communications Systems Branch, Naval Research Laboratory, Washington, D.C. 20375

Abstract

This paper describes the current status of an ongoing research and development effort whose objective is a variable rate voice coder with a high degree of speech intelligibility and natural voice quality with an average throughput rate near 1200 b/s. The voice coder described here is based on the DoD FS 1015 LPC-10 2400 b/s vocoder. A method of silence detection and the use of variable size data frame formats to distinguish voiced and unvoiced frames are discussed. This voice coding technology is a critical component in supporting efficient integrated services over bandwidth-constricted radio networks, satellite systems, and handheld portable communications devices.

I. Introduction

There is increasing interest in low rate speech vocoding for both military and commercial communications systems. Both the government and industry are actively pursuing the development and fielding of integrated services networks. The use of interactive multimedia applications is also on a dramatic rise and places a severe strain on bandwidth-constricted communication networks, particularly mobile radio communications.

The use of variable rate vocoding technology (<2400 b/s) will be useful in future mobile communications devices and within existing datagram packet networks (e.g. INTERNET) to support multimedia communications while reducing data throughput requirements.

II. Vocoder Operation

A block diagram of the LPC-10 vocoder is shown in Figure 1. Linear prediction analysis is performed at a 22.5 ms frame rate by an open loop tenth-order covariance method. There are two important estimated frame-by-frame parameters, energy and voicing, which will be used to perform variable rate encoding decisions. A frame-by-frame rms energy measurement is already implemented by the LPC analysis routine and its output will be used to perform silence detection processing. In addition, unvoiced LPC frames contain redundant information which can be removed for network voice applications resulting in a shorter frame length. The rationale and method will be discussed in the following paragraphs.

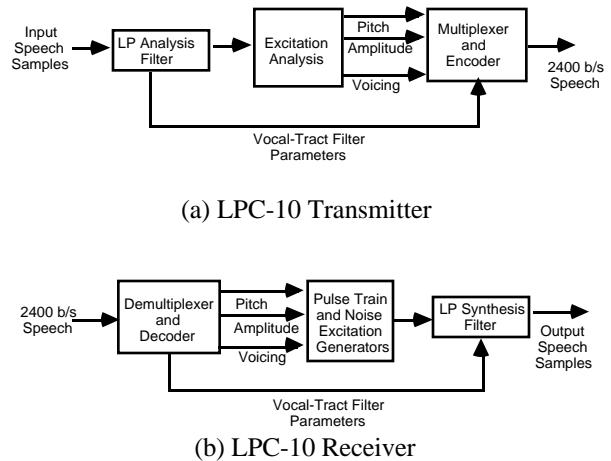(a) LPC-10 Transmitter

(b) LPC-10 Receiver

Fig.1 - LPC-10 Voice Processor

The FS 1015 LPC-10 standard employs a Hamming error correction coding method to improve performance in high bit error rate environments. In layered communication systems such as integrated networks, there is generally less need for traditional bit error correction at the application layer. The application layer will need to deal with handling lost frames or packets and assume some responsibility for data transport efficiency. Within a network environment, the transport, network, and link layers generally provide delivery of error-free datagram data, although order and arrival times are not always guaranteed. Therefore, the Hamming error correction coding does nothing to help in this case. We use this rationale to create a variable frame structure and allow silence frames to be dropped.

The FS1015 LPC-10 Hamming code parity information is embedded within the unvoiced frames resulting from the LPC analysis routines. Unvoiced sounds with significant energy are a critical part of human speech intelligibility. Therefore, such frames will be encoded and transmitted, but will require a fewer number of bit to quantize without coding. If the energy value for an unvoiced frame is below a threshold value it is considered a silence frame and need not be transmitted. The resulting vocoder transmission source will produce vocoder frames during "voice spurt" periods in which the frame rms energy is above a predefined threshold. Unvoiced frames considered critical to speech intelligibility will be retained, but will be shortened in length. To make this work, the Hamming decoder within LPC-10 has been deactivated and a silence and

1. REPORT DATE **JUL 1994**	2. REPORT TYPE	3. DATES COVERED **00-00-1994 to 00-00-1994**
4. TITLE AND SUBTITLE **A Variable Rate Voice Coder using LPC-10E**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Communications Systems Branch,4555 Overlook Avenue, SW,Washington,DC,20375**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**2**	

variable frame size packing and unpacking routine has been added. The initial prototype vocoder analysis process produces three basic voice frame types as defined in Table 1.

Frame Type	rms Value	Length (bits)
Voiced	>minEnergy	54
Unvoiced Speech	>minEnergy	34
Silence	<minEnergy	0

Table 1: LPC-10 Variable Rate Frame Types

III. Application

The authors' initial use of this experimental design is in the NRL network voice terminal application. This application has been designed to operate on standard computer workstations and provide both a "telephone-like" user interface and a datagram network interface for voice communications. A number of vocoder algorithms and operational modes are available upon configuration and number of fixed rate vocoders have been incorporated to run under this environment including, 2400 (LPC-10E) and 1200 b/s. A variable rate source design paradigm provides a good match to a datagram network service interface and will provide a more efficient, yet intelligible, interactive network voice application.

Initial experiments with this variable rate vocoding scheme have demonstrated good intelligibility, as well as efficient use of the available transmission bandwidth. Over 1 sec interval measurements, typical voice conversations have indicated an average transmission rate requirement of 1000-1200 b/s.

IV. Future Work

There is desire in some areas to apply resource reservation mechanisms to improve the performance of real-time network applications (e.g., voice, video). A variable rate source presents a unique problem in terms of bandwidth resource reservation. If resources are reserved at the expected peak rate, then resources will be used inefficiently on the average. On the other hand, if an average throughput rate is reserved then the possibility of congestion and service denial results during peak activity periods. Reserving "virtual bandwidth" based on service priority level is one potential method for avoiding bandwidth management

difficulties. Assigning voice services an appropriate high priority level, means that delay and bandwidth requirements will be met on the average. Resource reservation issues will be explored by experiments with variable rate voice applications in a variety of communication networks and loading situations.

More experiments are also planned regarding the appropriate framing of talkspurts and the application of these technique to other LPC-based vocoders. In order to preserve human speech transition fidelity, the transmission of voice frames directly before and after a talkspurt period may be desired. Even though, the rms energy appears low there may be critical partial transition information within these frames.

IV. Conclusion

A variable rate vocoder has been demonstrated achieving average throughput rates close to 1000 b/s. The technique involves some simple modification to the existing standard FS 1015 LPC-10 algorithm. Performance with low network frame loss is effectively the same as achieved with unmodified LPC-10. Such very low rate voice technology is expected to be a critical component in future military and commercial communications systems.