



NRL/FR/5555--07-10,145

# Variable Data Rate Voice Encoder for Narrowband and Wideband Speech

THOMAS M. MORAN  
DAVID A. HEIDE  
YVETTE T. LEE

*Transmission Technology Branch  
Information Technology Division*

GEORGE S. KANG  
*ITT Industries (AES)  
Herndon, VA 20170*

March 2, 2007

# REPORT DOCUMENTATION PAGE

*Form Approved*  
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 02-03-2007		<b>2. REPORT TYPE</b> Formal		<b>3. DATES COVERED (From - To)</b> October 1, 2004 to December 1, 2006	
<b>4. TITLE AND SUBTITLE</b>  Variable Data Rate Voice Encoder for Narrowband and Wideband Speech				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b> 33904N, 61553N	
<b>6. AUTHOR(S)</b>  Thomas M. Moran, David A. Heide, Yvette T. Lee, and George S. Kang*				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b> IT-235-009	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Naval Research Laboratory Washington, DC 20375-5320				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  NRL/FR/5555--07-10,145	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  Naval Research Laboratory Washington, DC 20375-5320				<b>10. SPONSOR / MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR / MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Approved for public release; distribution unlimited.					
<b>13. SUPPLEMENTARY NOTES</b>  *ITT Industries (AES), Herndon, VA 20170					
<b>14. ABSTRACT</b> Past designs for many military communications systems were based upon specific radio links with fixed and limited channel capacities. Accordingly, many different voice compression algorithms, operating at various fixed rates, were implemented. While still being used today, these incompatible systems are an obstacle to interoperable communications. Emerging net-centric communications promise to provide connectivity to all military users but voice interoperability will still require compatible voice encoding as well as encryption for secure communications. This report details a Variable Data Rate (VDR) voice encoder that is designed to provide interoperable secure voice communications for net-centric users. While being backwards compatible with the Federal standard voice encoder (MELP) at 2400 bits per second (bps), it operates at a range of data rates up to 26,000 bps. Because the rate setting can be changed dynamically, the VDR encoder can provide efficient use of network bandwidth yet be interoperable at any and all rates simultaneously, and, with the proper encryption, even when secure.					
<b>15. SUBJECT TERMS</b> Variable data rate vocoder      MELP vocoder      Wideband speech      Speech modeling Residual excited LPC					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			Thomas Moran
Unclassified	Unclassified	Unclassified	Unlimited	30	<b>19b. TELEPHONE NUMBER (include area code)</b> 202-404-8842



## CONTENTS

1. INTRODUCTION .....	1
1.1 Why Does the DoD Need a VDR Voice Processor for Secure Voice?.....	1
1.2 Characteristics of the VDR Voice Processor .....	1
1.3 Our Ultimate Goal.....	2
2. BACKGROUND .....	3
2.1 DoD Voice Communication Environments are Multirate.....	3
2.2 Previous Approaches to Multirate Voice Processing.....	4
2.3 History of Our VDR R&D Efforts .....	6
3. TECHNICAL APPROACH.....	6
3.1 Efficient Speech Coding Generates VDR Speech Data .....	6
3.2 VDR Generates Universally Interoperable Multirate Voice Data.....	8
3.3 Wideband VDR vs Narrowband VDR.....	12
3.4 Narrowband VDR .....	13
3.5 Wideband VDR.....	20
4. CONCLUSIONS.....	26
5. ACKNOWLEDGMENTS .....	26
REFERENCES .....	27



# VARIABLE DATA RATE VOICE ENCODER FOR NARROWBAND AND WIDEBAND SPEECH

## 1. INTRODUCTION

### 1.1 Why Does the DoD Need a VDR Voice Processor for Secure Voice?

The primary reason to use a variable data rate (VDR) speech encoder is to provide interoperability for the widest number of Department of Defense (DoD) secure voice users. Only a VDR encoder, such as the one described in this report, can interoperate both securely and efficiently in place of the many different voice encoders now being used across the DoD.

The need to update DoD secure voice is well documented in the publication “C<sup>4</sup>I for the Warrior” [1]. In this document from the early 1990s, the Joint Staff recognized the need for secure voice interoperability in the DoD. Since that time, we at the Naval Research Laboratory (NRL) have developed technology to help resolve these compatibility issues. NRL’s VDR speech encoder operates at the various data rates necessary to satisfy the DoD’s voice communication requirements and, most importantly, all of the various rates of the VDR encoder are directly interoperable. Furthermore, VDR was developed to be especially efficient over Internet Protocol (IP) networks by having the capability to dynamically adjust the encoding rate to the network traffic conditions.

### 1.2 Characteristics of the VDR Voice Processor

As part of the introduction, VDR characteristics are simply stated below, leaving further elaborations to later sections.

- The VDR voice processor is a multirate voice processor in which a single voice algorithm generates multiple data rates from 2.4 kilobits per second (kbps) to an average rate of about 23 kbps for 0 to 4 kHz input speech.
- The 2.4 kbps rate is the Federal standard algorithm for narrowband speech; the Multiple Excitation Linear Predictive (MELP) voice encoder.
- Inclusion of a few more kbps of data from the 4 to 8 kHz audio frequency band makes it possible to generate spacious and crisp FM-like wideband speech.
- The VDR bitstream has an embedded structure (in which higher-rate voice data frames contain successively lower-rate voice data frames as subsets). Deleting a certain portion of the superset makes it possible to reduce the data rate, even in the encrypted mode. Because of this embedded data structure, any of the VDR data rates are interoperable and may be switched, as often as 44 times per second, even when speech is present. Importantly, it does not create undesirable sounds such as clicks or warbles during rate changes. This is because the speech waveforms at all VDR data rates are synchronous.

- This is not a collection of separate voice encoders operating at different data rates. The VDR encoder is a single voice processing principle designed to be matched with a single encryption principle.
- VDR exploits the variable nature of the speech waveform; for example,
  - Vowels need higher data rates because the structure of complicated pitch harmonics of a vowel waveform must be well preserved, otherwise speech will sound warbled.
  - Consonants can be encoded at lower data rates because the random waveform of a consonant does not require an exact representation.
  - Speech gaps within a word, between words, and between phrases need even fewer bits to encode because speech gaps are primarily environmental noise.

Although the VDR is a multirate device, the VDR processor is not a device that hosts a multitude of voice algorithms. Voice terminals that use multiple compression algorithms do not perform well when switching algorithms in mid-conversation. When doing so, the speech waveform sometimes gets cropped because different voice algorithms can have different internal delays. This hurts speech quality and is annoying to the users.

Note also that the VDR processor does not achieve efficient coding by eliminating speech gaps. Such an approach for reducing the speech data rate is called Time Assigned Speech Interpolation (TASI). TASI was extensively used for reducing the number of trunking channels for long-distance voice communication. The idea of eliminating speech gaps that contain ambient sounds is a bad idea for military communication because speech gaps often contain critical information for gauging the battlefield conditions at the transmitter site. Therefore, VDR encodes speech gaps at appropriately low data rates that still provide audible information.

### **1.3 Our Ultimate Goal**

Our ultimate goal is to provide the core technology for universal secure voice. This core will be the VDR voice processor combined with VDR encryption. Associated with the core will be the protocols for rate control and interfacing the secure voice terminal with the underlying network. The intention is to provide the key components of a secure voice architecture that can be implemented in phases.

Most of Navy (and DoD) voice communication will require two types of terminals. One model is a desktop version that will function as a Universal Voice Terminal (UVT) (Fig. 1). We envision the UVT to have connectivity worldwide. It will function over all DoD networks and be the hub for the handheld terminal.

The other model of VDR terminal is a handheld wireless device, the Personal Secure Terminal (PST), intended to be issued to every foot soldier. It will be a short range radio that provides secure group communications but will also interoperate with the UVT to reach the command center.

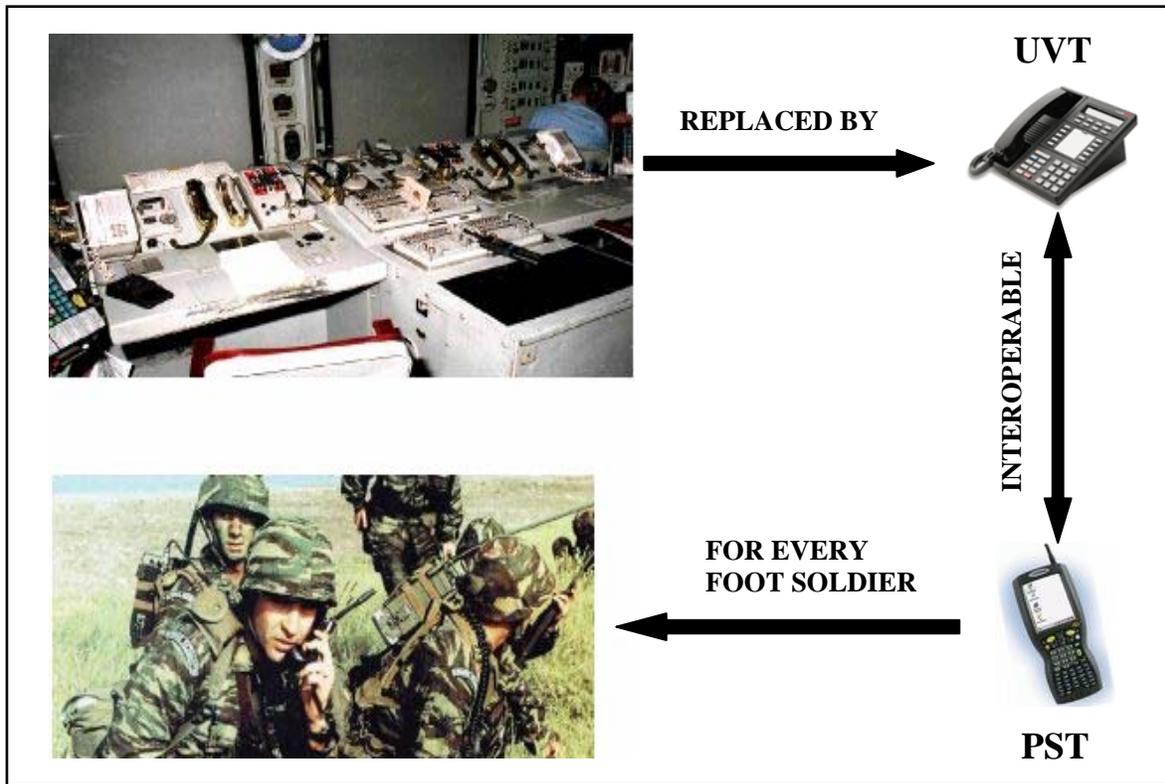


Fig. 1 — Combat Information Center (CIC) with a confusing array of secure voice terminals. VDR can integrate all these incompatible voice terminals into a single interoperable secure voice phone that we call the Universal Voice Terminal (UVT). In addition, we plan to develop a pocket-size version of VDR for every foot soldier that we call the Personal Secure Terminal (PST). PST gives connectivity among all soldiers and also interoperates with a UVT to reach the command center or other plant version of the VDR. Note that currently all soldiers are given weapons, but not a phone. In the future, every foot soldier should have a pocket-size phone that enables them to communicate with fellow soldiers and the commander. There should be no incident similar to that of Jessica Lynch who fell into enemy hands because of the inability to make contact with friendly forces.

## 2. BACKGROUND

### 2.1 DoD Voice Communication Environments are Multirate

In Fig. 2, typical tactical communication environments are capsulated into four categories in terms of usable data rates. As noted, DoD voice communication data rates range from as low as 2.4 kbps to as high as 32 kbps or higher. Figure 2 explains why so many different data rates are needed for voice communication by the Navy and DoD.

Figure 2(a) shows voice communication over narrowband links where all that may be available is a 2.4 kbps link. Figure 2(b) shows an extremely noisy platform where the 2.4 kbps voice terminal is not usable. There are ample test data within DoD to indicate as such. Note that the Joint Tactical Information Distribution Systems (JTIDS) uses the 16 kbps voice data rate in their F-14 platforms. Figure 2(c) shows President Bush on Air Force One, a much quieter environment compared with the F-14. When high ranking officers engage in high-level conversations, they deserve the very best digital voice system, where the data rate could be on the order of 64 kbps at a constant data rate or VDR at 20 to 30 kbps. Figure 2(d) shows a ship where the operating communication environment is a universe in itself. It must have the capability to transmit voice encoded at any data rate over all possible channels including HF,

UHF, VHF, and SHF, and all different satellite channels, such as, MILSTAR and FLEETSATCOM. This complicated naval communication architecture will be simplified if VDR is used.



Fig. 2 — Four examples of platforms where naval voice communications take place. For the reasons mentioned above, each operational environment needs a different data rate.

## 2.2 Previous Approaches to Multirate Voice Processing

Previous approaches to satisfying the multirate communication environments have been to develop many different voice terminals, each operating at a specific data rate (see Table 1). They can interoperate only through a tandem arrangement (speech is regenerated by the first voice terminal, and the regenerated speech is redigitized if tandemed analogly, and finally re-encoded by the second voice terminal). In these processes, speech quality will be degraded; in some cases, severely. Also the speech data must be decrypted and encrypted again. Therefore, it is impossible to achieve end-to-end encryption, which is a DoD secure voice goal.

### 2.2.1 Currently Deployed DoD's Voice Terminals

Table 1 lists some of the most common DoD voice terminals from the current inventory along with the voice algorithm used. A voice terminal is more than a voice encoder. It includes an encryptor, a modem, and sometimes an RF transceiver.

Table 1 — Some of DoD's Currently Operational Voice Terminals

<b>DoD Voice Terminal</b>	<b>Voice Algorithm</b>
ANDVT TACTERM AN/USC-43	2.4 kbps LPC
ANDVT MINTERM KY-99A	2.4 kbps LPC
ANDVT AIRTERM KY-100	2.4 kbps LPC
STU-III	2.4 kbps LPC, 4.8 kbps CELP
STE	2.4 kbps LPC, 2.4 kbps MELP, 4.8 kbps CELP, 6.4 kbps G.729, 32 kbps ADPCM
VINSON KY-58	16 kbps CVSD
DSVT KY-68	16 kbps CVSD
SINGARS	16 kbps CVSD

### 2.2.2 Other Voice Processing Algorithms

For many years the services worked with the National Security Agency (NSA) to develop secure voice algorithms. The NSA had extensive programs to improve, test, and evaluate voice encoding algorithms (such as LPC, MELP, APC, CELP, CVSD, and ADPCM) in military and other secure voice applications. In the 1970s, the NSA investigated more than a dozen voice algorithms. For each investigation, NSA published an exemplary report that characterized that particular algorithm in a wide range of DoD applications.

Commercial telecommunications are largely based on voice encoding algorithms standardized through the International Telecommunication Union (ITU). Commercial standardization makes it easier to implement compatible communications devices. However, unlike the military standard algorithms listed above, it is up to each user to test the performance of these voice algorithms to see which is suited for their applications. They are not usually optimized to the harsher military communications environments in terms of intelligibility and quality under acoustic noise and transmission errors. Table 2 lists some of the most common ITU algorithms applicable to DoD uses. None of them are directly interoperable.

Table 2 — A Sample of the Most Common Voice Processing Algorithms

<b>Standard Number</b>	<b>Voice Processing Algorithm</b>
G.711	64 kbps PCM
G.722	64, 56, 48 kbps Wideband ADPCM
G.727	40,32,24, 16 kbps ADPCM
G.728	16 kbps Low Delay CELP
G.729	8 kbps CS-ACELP
G.729D	6.4 kbps CS-ACELP

## 2.3 History of Our VDR R&D Efforts

In 2001, Kang documented our initial R&D efforts on VDR in an NRL report [2]. Since then, our insight into VDR has grown substantially through the following VDR-related activities:

- The VDR algorithm has been implemented and demonstrated in real-time at our laboratory.
- The process of estimating the network traffic density for controlling the VDR encoder rate (which we call the Network Arbitrator) has been implemented in house. The Network Arbitrator is essential for the VDR to operate over real-world IP networks.
- We have received test and evaluation data and feedback from the staff of the SPAWAR engineering facility at St. Juliens Creek in Chesapeake, VA. They are naval communication experts specializing in installation, maintenance, and support of naval secure voice terminals.

## 3. TECHNICAL APPROACH

### 3.1 Efficient Speech Coding Generates VDR Speech Data

In VoIP applications, users share fixed network resources. Often these network resources are limited, which also limits the number of users able to communicate simultaneously. Maximizing the number of users for a given network condition requires efficient speech encoding. Since speech is an inherently variable signal, VDR encoding naturally provides the necessary efficiency for a range of quality levels. In the implementation of the VDR voice encoder, we exploit three main aspects. These are 1) the nature of the speech waveform, 2) human auditory perception characteristics, and 3) the operational network constraints.

#### 3.1.1 *Exploitation of the Nature of the Speech Waveform*

The speech waveform is a variable information source. In other words, the encoding of consonants (/s/, /sh/, /t/, /p/, etc.) requires lower data rates than the encoding of vowels, and the encoding of speech gaps between words, phrases, or within a word requires even lower data rates (Fig. 3). As such, speech is a variable-data-rate information source. The optimum speech data rate is automatically determined on a frame-by-frame basis (every 22.5 ms).

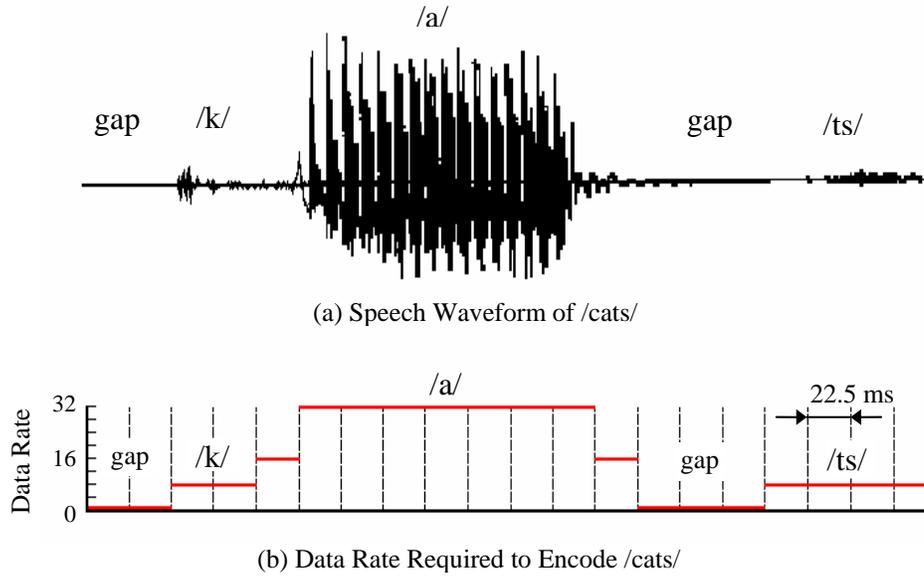


Fig. 3 — Variable-Data-Rate nature of the speech waveform. As noted, the data rate to encode the word /cats/ varies from less than 1 kbps for gaps to 32 kbps for vowel /a/. Achieving high-quality speech transmission does not require high data rates (viz., 64 kbps) all the time; high data rates are only required briefly for the vowels or other complex speech waveforms. The VDR encoder exploits these inherent VDR characteristics of the speech waveform by optimizing the data rate every 22.5 ms.

### 3.1.2 Exploitation of Human Auditory Perception Characteristics

Human ears and brains resolve lower frequencies more accurately than higher frequencies. Thus, the fidelity of low frequency encoding is critical to achieving acceptable speech quality. Hence, the VDR encodes lower frequencies of the speech content more accurately (using more bits) than higher frequencies. Based on a well-known experiment on audio perception [3] and our own experiment based on speech-like sounds (pitch-modulated sounds with three to four resonant frequencies), we use the frequency resolution that approximates those experiments; i.e., we decrease resolution by approximately one dB per octave (Fig. 4).

### 3.1.3 Operational Network Constraints

It is more important to communicate at lower data rates (with reduced speech quality) than to entirely disrupt communication by preemption when being affected by overloaded network conditions. It is an issue of survivability of communication. VDR has an option to select seven different operating modes with seven different average data rates. The network traffic density significantly influences the preferred operating mode. NRL developed a processor called the Network Arbitrator that measures the traffic density, which in turn selects the preferred operating mode.

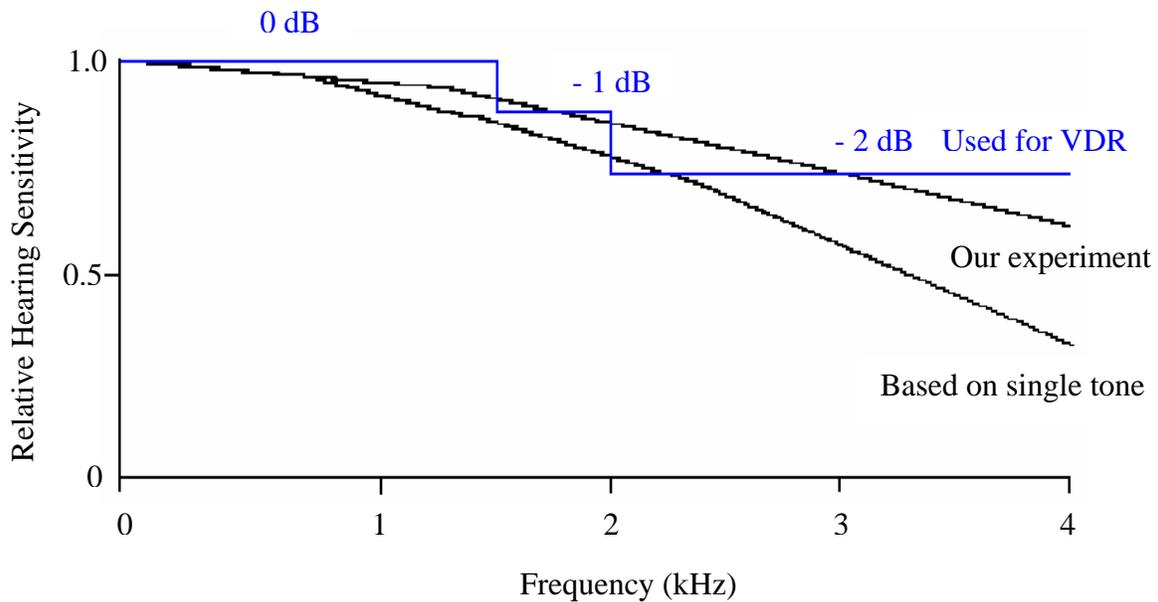


Fig 4 — Relative Hearing Sensitivity to Frequency Difference. Our experiment was based on a speech-like tone with three resonant frequencies, which are repetitive at a pitch frequency (80 Hz). VDR uses the frequency-dependent resolution that approximates the two curves, as shown in this figure. The idea is that if the human ears and brain cannot resolve higher frequencies very accurately, those frequency components only need to be represented at a comparably low level of resolution. Reducing frequency resolution in this way can lower any particular data rate as much as 5 kbps.

## 3.2 VDR Generates Universally Interoperable Multirate Voice Data

### 3.2.1 Multiple Voice Data Rate from a Single Voice Processor

Kang's 2001 report [2] describes one voice processing principle that is used in four operating modes. This LPC-based speech analysis/synthesis system is capable of generating multirate speech by altering the resolution of the residual samples.

Following that work in 2001, we have now added three modes. Table 3 defines all seven modes.

Table 3 — VDR Operating Modes

Mode #	Description	Average Data Rate for Clean Conversational Speech
Mode 1	MELP Standard	2.4 kbps Fixed
Mode 2	Hybrid of Mode 1 (MELP) and Mode 3 VDR	7 kbps
Mode 3	VDR with spectral replication above 1.5 kHz	12 kbps
Mode 4	VDR with spectral replication above 2 kHz	15 kbps
Mode 5	VDR with spectral replication above 3 kHz	19 kbps
Mode 6	VDR with no spectral replication	23 kbps
Mode 7	Mode 6 with upper-band (4-8 kHz) added	26 kbps

Note that Mode 1 is exactly the standardized MELP algorithm selected for use in the DoD as the preferred 2.4 kbps algorithm. The MELP algorithm is interoperable with legacy 2.4 kbps terminals (ANDVT and STU-III) through the use of a transcoding technique developed by the authors [4]. To conserve data we use several of the parameters in the MELP bitstream to generate common parameters used in the VDR algorithm. Mode 2 is actually a hybrid of the Mode 1 MELP mode and the Mode 3 VDR mode. Mode 7 adds a wideband (0 to 8 kHz) capability to Mode 6 of the VDR algorithm. All of these modes are discussed in more detail later in this report.

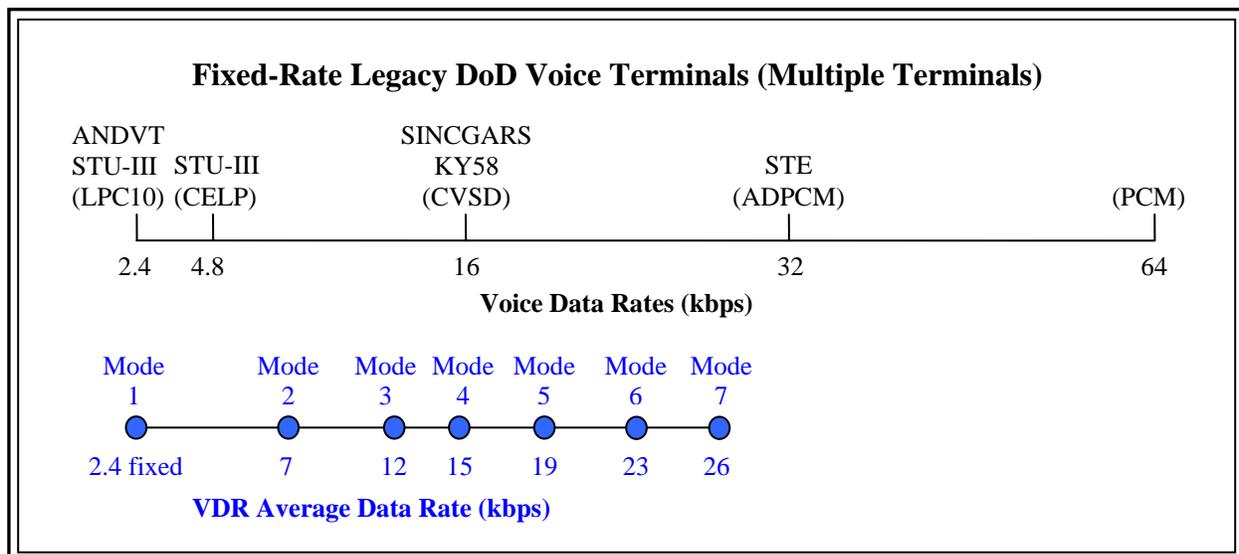


Fig. 5 — The VDR encoder matches the lowest data rates used by legacy DoD voice terminals and ranges up to maximum rate of 32 kbps. According to our listening tests, VDR at an average data rate of 23 kbps compares favorably with the fixed-rate 64 kbps Pulse Code Modulator (PCM). At the highest average setting of 26 kbps, the input speech bandwidth is 8 kHz vs 4 kHz for all the other rate settings.

### 3.2.2 Embedded Data Structure Makes Universal Interoperation Possible

The VDR bitstream has an embedded structure (i.e., a frame of high-rate voice data contains subframes of lower-rate data, which makes it possible to interoperate between any two different VDR rates (Fig. 6).

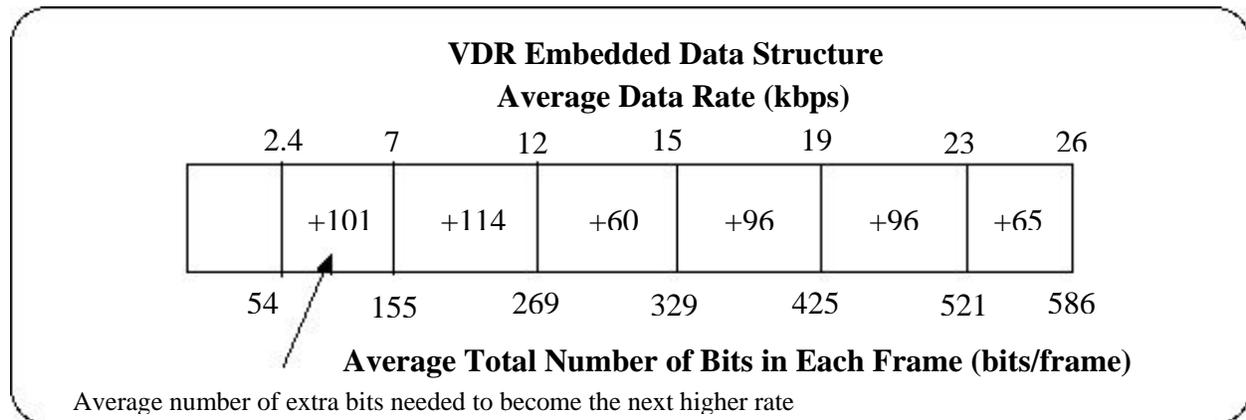


Fig. 6 — Embedded data structure of VDR in each frame. A lower-rate voice data frame plus extra bits (to improve speech) becomes a higher-rate data frame. Note that the numbers of bits in the VDR modes are, of course, *variable*, and the above bitstream just shows the embedded data structure with an average number of bits for each mode. Later we will show how the upper-band speech data can be added to any part of this bitstream to give upper-band capability to any of the VDR modes, not just the highest mode.

### 3.2.3 Speech Waveforms at All VDR Data Rates are in Sync

All the speech waveforms generated from the VDR data are synchronized (Fig. 7). Therefore, a VDR data rate can be switched to a lower VDR data rate on fly (even while talking). With the network arbitrator, the VDR data rate can be lowered, or raised, without user intervention. Because all VDR speech waveforms are synchronized, undesirable clicking noise will not be generated at the data-rate transitions.

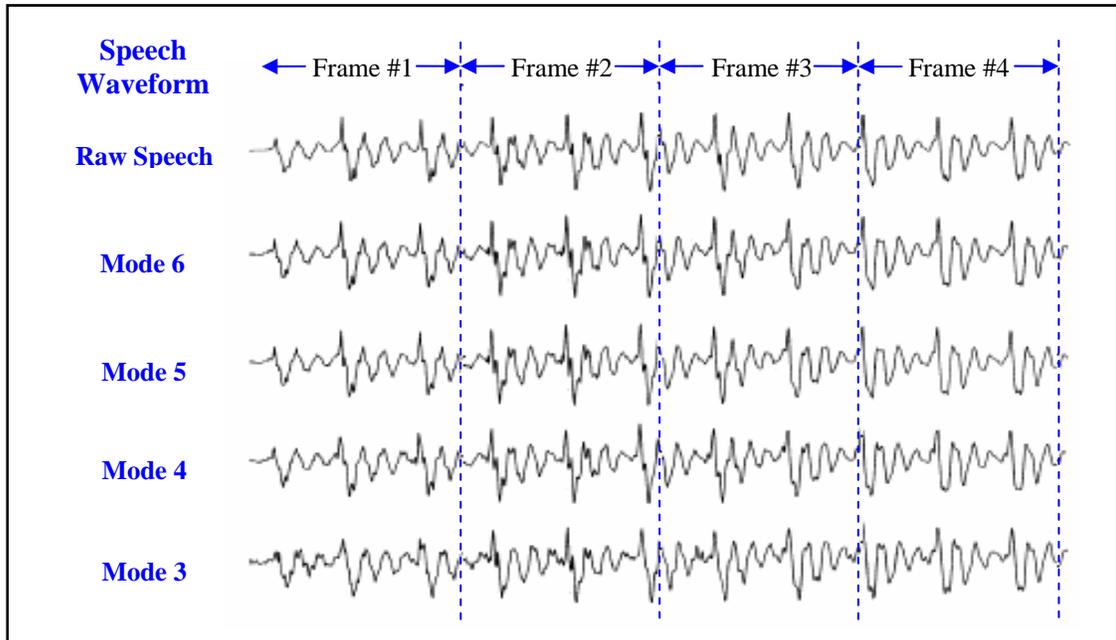


Fig. 7 – The speech waveform generated at all VDR data rates is in sync. Therefore, switching of data rates does not generate clicking noises that would otherwise be caused by the waveform discontinuities at the transition time instant.

### 3.2.4 Two-Dimensional Dynamic Data-Rate Optimization

VDR has seven operating modes, from which one may be chosen based on network traffic conditions. As indicated by Fig. 8, for the average-data-rate range selected (by the network arbitrator), there are seven possible instantaneous data rates (i.e., data rates at each frame) from which one optimum data is automatically selected at each frame (22.5 ms) based on the complexity of the speech waveform. Later we will discuss the seventh mode, where a wideband (0-8 kHz) speech capability is added.

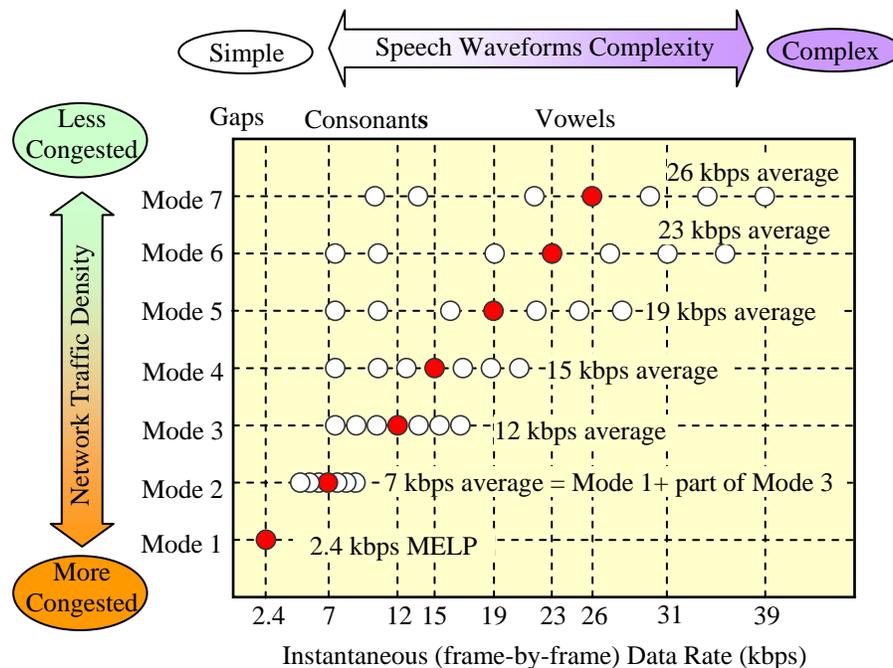


Fig. 8 — Two-dimensional optimization of data rates based on network traffic conditions and the complexity of the speech waveform. The red dots give the average rate for each mode. Mode 2 is a difficult mode. To make it sound better than 2.4 kbps speech and work better with extremely noisy input speech, Mode 2 is a superposition of two speech inputs: The audio band below about 700 Hz is encoded using a portion of the VDR residual encoder. The band in the 700-4000 Hz range is encoded using 2.4 kbps MELP. The presence of a portion of VDR speech as a supplement to the 2.4 kbps MELP speech provides a much improved tolerance to noise.

### 3.3 Wideband VDR vs Narrowband VDR

The term *Wideband VDR* refers to the version of VDR in which the input speech frequency has a bandwidth of 0 to 8 kHz. The earlier VDR [2], which we now call *Narrowband VDR*, has a bandwidth of 0 to 4 kHz.

In the early days of telecommunications, transmission channels did not support wideband analog speech; for example,

- Switched analog telephone networks typically have a 0 to 3 kHz bandwidth.
- Signal bandwidth for AM radio broadcasts is typically 3 kHz.
- Signal bandwidth for HF channels (shortwave circuits) is around 2 kHz.

Truncating the bandwidth of an analog speech signal still provides usable speech intelligibility because the uncompressed analog speech has many redundancies. In most digital voice communication the speech signal bandwidth is still limited to 0 to 4 kHz, but speech redundancies are removed (by compressing). Our tests of digitally encoded speech consistently indicates that female speech intelligibility is lower than that of male speech when the speech bandwidth is limited to 0 to 4 kHz, especially in noise [5 (Section 2)].

Human speech is wideband, 0 to 8 kHz, and often higher (Fig. 9). Wideband speech (0 to 8 kHz) is more intelligible than the standard 0 to 4 kHz narrowband, particularly for female voices, and has more tolerance to acoustic noise interference.

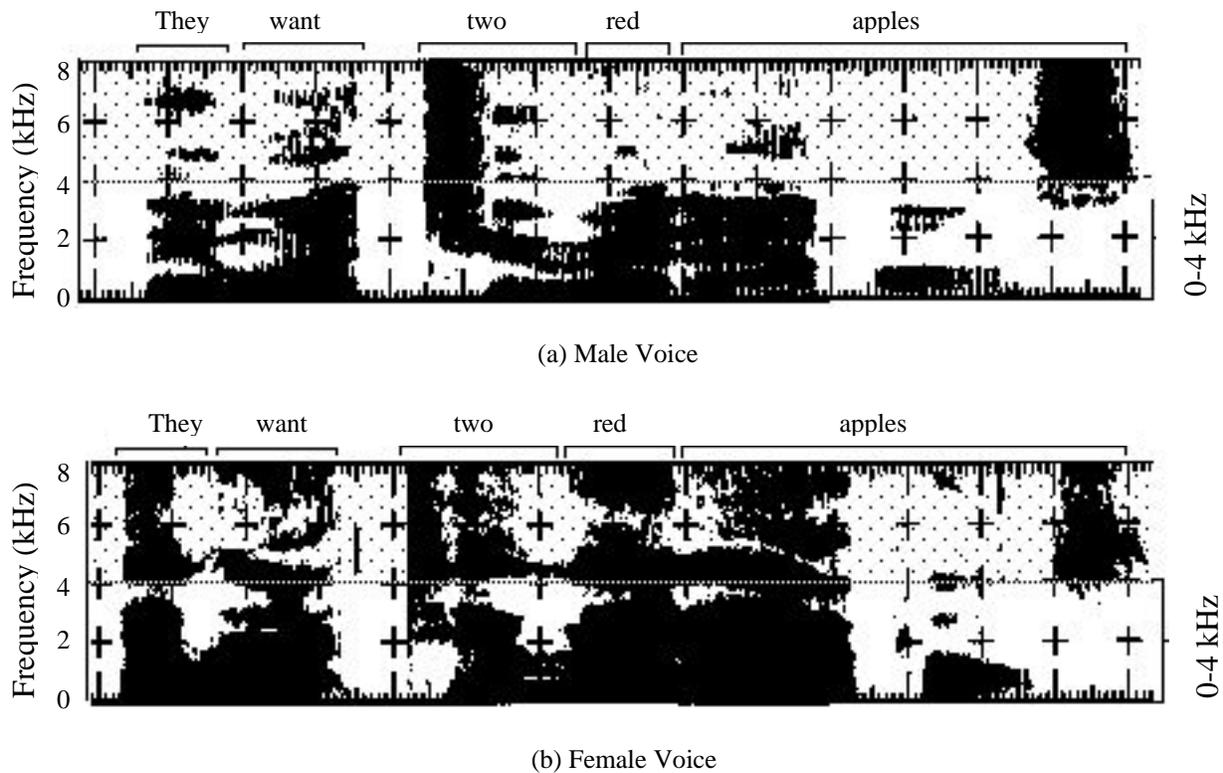


Fig. 9 — Speech spectrograms of typical male and female speech. This figure shows why the intelligibility of female speech is always lower than that of male voice if frequencies above 4 kHz are removed. Female speech has much more energy above 4 kHz.

In general, the spectrum of female speech has a considerable amount of speech energy above 4 kHz; much more than the male speech. Therefore, narrowband male speech scores better than narrowband female speech in formalized speech intelligibility tests, such as the Diagnostic Rhyme Test (DRT). When we first developed VDR in 2001 [2], the input speech was limited to a bandwidth of 0 to 4 kHz which is the standard bandwidth for most telephony. The addition of the upper-band (4 to 8 kHz) speech data now makes this Narrowband VDR into Wideband VDR.

### 3.4 Narrowband VDR

The narrowband VDR is documented in the earlier NRL report [2]. In this section, highlights of the narrowband VDR are summarized to facilitate discussions of the wideband VDR to follow.

#### 3.4.1 Block Diagram

A block diagram of the narrowband VDR is shown in Fig. 10. Among speech analysis/synthesis systems, the LPC-based analysis/synthesis system was chosen for the following two reasons: (1) the VDR system is capable of directly interoperating with DoD's latest standard 2.4 kbps vocoder MELP and, indirectly, with the legacy standard 2.4 kbps LPC-10 vocoder used in the widely deployed ANDVT, and (2) the LPC analysis/synthesis system allows for the linear scaling of the data rate because it is a *unity-*

*gain system.* The output speech improves as the resolution of the error signal (the prediction residual) becomes finer (i.e., encoded at a higher data rate). At the finest level of resolution, the system generates an output signal that equals the input. In other words, this one system is capable of generating speech at widely varying rates with correspondingly varying levels of speech quality.

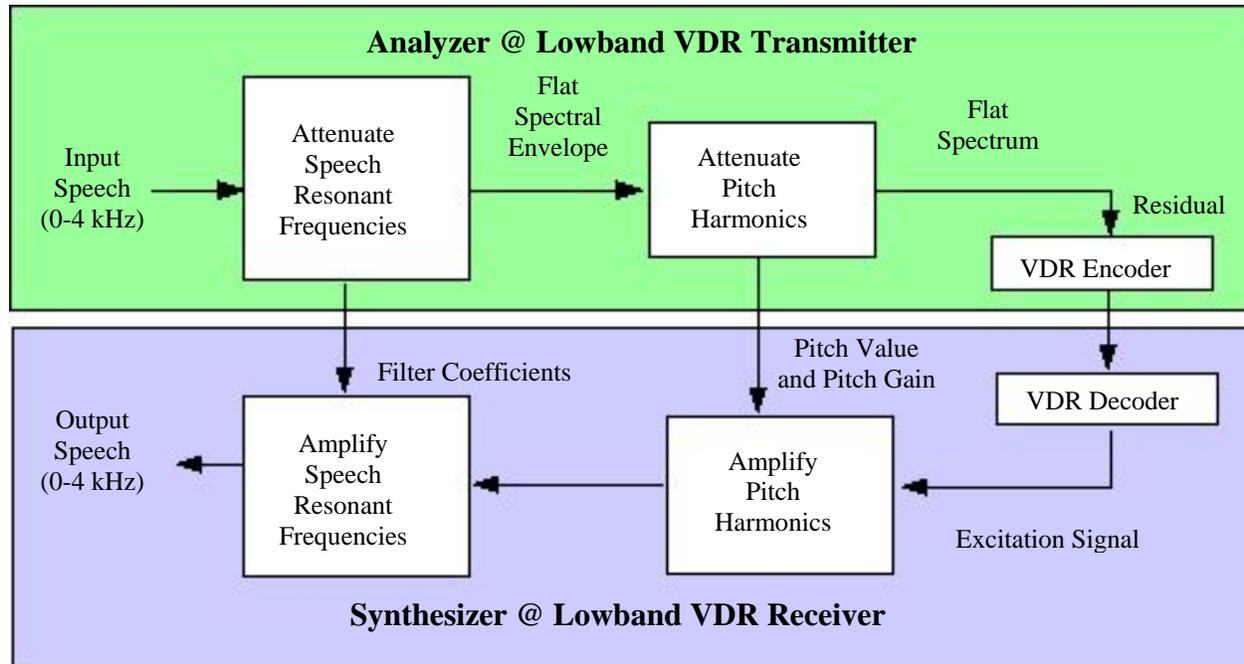


Fig. 10 — The block diagram of narrowband VDR based on the LPC analysis/synthesis system. The output speech quality is solely dependent on the resolution of (the number of bits used to encode) the residual.

The LPC analysis/synthesis system decomposes the speech waveform into slowly time-varying components and fast time-varying components. The slowly time-varying components include filter coefficients, the pitch value, and speech loudness. They are updated only once per frame (22.5 ms). The fast time-varying components are the prediction residual samples. They are updated sample by sample, 8,000 times per second (or every 125  $\mu$ s).

The LPC analysis/synthesis system is a two-stage spectral whitening (flattening) process; the first stage attenuates speech resonant frequencies, and the second stage attenuates pitch harmonics. Note that, even if the slowly time-varying components are quantized, as long as the prediction residual samples are computed from the quantized slowly time-varying components, the output speech quality is solely dependent on the resolution of the prediction residual. Thus, the data rate of the VDR system and the output speech quality can be controlled by the number of bits used to encode the prediction residual.

To ensure compatibility with the new MELP 2.4 kbps standard vocoder, the exact 54 bit MELP bitstream is used as the base kernel of the VDR bitstream. Because MELP and VDR are both based on LPC we are able to use common parameters from MELP to save bits in the VDR portion of the bitstream. The common parameters used are the LPC parameters (in the form of Line Spectral Pairs) and the pitch.

### 3.4.2 Advantages for Encoding Residual Samples in the Frequency Domain

The prediction residual may be encoded in the time domain or in the frequency domain. Encoding of the residual in the frequency domain (our approach), however, has many advantages:

**Ease of Incorporating Perception Characteristics in Coding:** More efficient encoding of the prediction residual can be achieved by exploiting human auditory perception of sound frequencies during the quantization process. These characteristics are easier to accommodate in the frequency domain, as shown in Fig. 4.

**Amplitude-Dependent Phase Coding:** Encoding in the frequency domain makes it possible to perform amplitude-dependent phase resolution. In this process, the phase resolution is encoded more coarsely when the amplitude spectrum is low (and so less detectable by the listener). In this way, we are able to save as much as 5 kbps.

**Replication of the Residual Spectrum is Possible:** This process allows for the most effective and efficient residual coding and is the single most important topic of VDR. Therefore, it will be discussed in a separate section later.

To quantize residual samples in the frequency domain, however, requires more computations than in the time domain. First of all it is necessary to overlap the analysis frames to reduce the noise created by waveform discontinuities. We overlap the 180 residual time samples in each frame with 12 samples of the previous frame. Secondly, a fast Fourier transform is required to obtain the frequency components of the residual. The Winograd transform is used (because the number of samples is not a binary number) on a total of 192 samples to generate 96 real and 96 imaginary components. The transform process gives us 24 spectral components in each of the four 1000 Hz frequency bands. The DC component and the first spectral component (at  $f = 41.67$  Hz) are not transmitted because they do not result in audible sounds.

To speed up the spectral encoding/decoding process, we use look-up tables. In these tables, the real and imaginary parts of each input spectral component are quantized to represent an address which is used by the synthesizer to directly read the corresponding spectral code from the look-up table. We have 7 different coding tables (9-bit, 8-bit, 7-bit, 6-bit, 5-bit, 4-bit, and 3-bit tables). The 9-bit table has 512 spectral codes. If the decoded real and imaginary values are plotted in a unit-circle of the  $z$ -plane, they form a constellation made of 512 points.

### 3.4.3 Parameter that Indicates the Preferred Instantaneous Data Rate

The instantaneous data rate is the data rate for each individual frame. It is confined to a range of values determined by the operating mode selected. After the mode of the narrowband VDR is chosen based on the traffic condition of the network, the instantaneous data rate is determined based on the complexity of the speech waveform. As stated earlier, encoding a vowel requires a high data rate, whereas encoding a consonant requires a lower data rate. We found that the peak magnitude of the 96 residual spectral vectors is a reliable indicator for the instantaneous data rate of that particular frame. The reasons why this parameter works well are:

- When the residual spectrum is computed for each speech frame, the largest magnitude spectral component is used to normalize all the components in that frame. (This is because the normalized spectrum is simpler to quantize using a unit circle representation). This makes the total number of bits required to encode the residual proportional to the peak spectral amplitude. There is no reason to allocate more bits to encode each residual spectrum component than the number of bits to encode the peak spectrum component.

- If the speech waveform is more complicated (for example, the speech waveform has many resonant frequencies, or a vowel is modulated by noise (as in /z/, /j/) the LPC prediction process produces more errors. Since the residual signal is the computed error of the LPC prediction signal, the amplitude spectra of the residual will be larger when more prediction errors are produced. Therefore, the peak amplitude spectrum is a good indicator of the signal complexity, and thus the number of bits required to encode the residual spectral components in order to capture that complexity.

#### 3.4.4 Spectral Replication of Residual Spectrum as a Means to Reduce Average Data Rate

The narrowband VDR quantizes the residual spectrum using four separate frequency bands. Only at the highest rate setting, all of the quantized residual frequency components are transmitted. At lower rate settings the higher frequency components are stripped off at the transmitter. At the receiver, these higher frequencies are reproduced from the lower frequencies using spectral replication. The spectral replication process allows the rate to change without noticeably affecting speech quality. This is the key technique, patented [6] by the Navy that makes VDR encoding possible. Unfortunately, the patent has expired.

The overall residual spectrum quantizer for narrowband VDR operates in Mode 6 (the highest average data rate of the Narrowband VDR, see Table 4). The number of bits for each residual spectral component is assigned based upon the speech waveform complexity. There are seven settings of bit size ranging from three to nine bits. The speech waveform complexity level is indicated by the peak value of the residual spectral component observed in each frame. As noted, we encoded the entire residual spectrum from near DC to 4 kHz.

In Mode 5, narrowband VDR transmits the residual spectrum from near DC to 3 kHz. The spectral components from 3 to 4 kHz are not transmitted. Instead, they are replicated at the receiver from the lower frequency spectral components. (see Table 5 and Fig. 11).

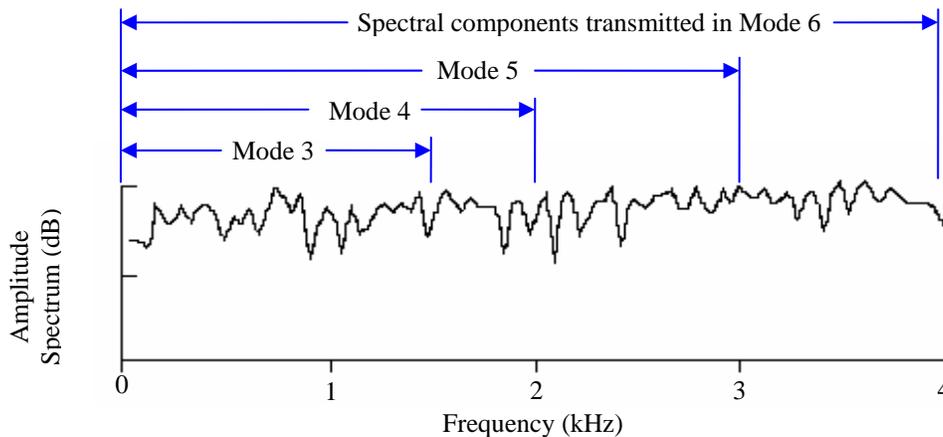


Fig. 11 – Residual spectrum and the portion used for a given operating mode, as indicated. Due to the relative flatness of the envelope, the lowband residual (0 to 1 kHz) may be upconverted (moved up in frequency) and used as the excitation signal for the higher frequency bands, i.e., the 3 to 4 kHz band. This spectral replication process makes VDR possible. In general, to implement VDR voice processing over a wide data-rate range by only adjusting residual quantization steps would result in terrible speech quality.

If the network arbitrator selects Mode 4, then spectral components from 2 to 4 kHz are not transmitted and they are replicated from those of 0 to 2 kHz (see Table 6). If the network arbitrator selects Mode 3, then spectral components from 1.5 to 4 kHz are not transmitted and they are replicated from those of 0 to 1.5 kHz (see Table 7).

In other words, low-frequency residual spectra may be upconverted as high-frequency residual spectra. The amplitude spectral error is small because the residual amplitude spectrum is relatively flat. The phase spectral error is inconsequential to the human auditory system because human ears cannot perceive the phase information. Therefore, spectral replication is an efficient way of reducing voice data rate, while minimizing any degradation of speech quality. One caveat in replicating spectral components is that they should be consecutive for 1 kHz or more, otherwise speech quality will be poor. Of course, the replicated high-frequency excitation signal is not the same as the original high-frequency excitation signal (used in Mode 6) but human ears cannot discern the difference too readily because the difference is primarily in the high-frequency phase spectrum.

If the network arbitrator selects Mode 2, then spectral replication cannot be used without a significant loss of speech quality. At this rate setting, only the spectral components below 700 Hz are transmitted. The band from 0.7 kHz to 4 kHz is then derived not from spectral replication but from that region of the 2.4 kbps MELP signal (see Table 8). This hybrid system of combining the lower residual spectral components of VDR and the band above 700 Hz of the MELP signal gives much more tolerance to high noise conditions than 2.4 kbps MELP alone.

Finally, under periods of very high network congestion the network arbitrator will select the base 2.4 kbps MELP standard mode, Mode 1. This selection will allow as many users as possible to keep communicating in a crisis without being preempted from the network.

Table 4 — Mode 6 Quantization Table for Narrowband VDR

Peak Amplitude of Pitch-Filtered Residual (# of Bits)		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)
		0.1-1.5 kHz (34)	1.5-2.0 kHz (12)	2-3 kHz (24)	3-4 kHz (24)		
Complex Waveform	9	9x34=306	8x12=96	7x24=168	7x24=168	803	36
	8	8x34=272	7x12=84	6x24=144	6x24=144	709	31
	7	7x34=238	6x12=72	5x24=120	5x24=120	615	27
	6	6x34=204	5x12=60	4x24=96	4x24=96	521	23
	5	5x34=170	4x12=48	3x24=72	3x24=72	427	19
Simple Waveform	4	4x34=136	3x12=36	0 (note 1)	0 (note 1)	237	11
	3	3x34=102	0 (note 1)	0 (note 1)	0 (note 1)	167	7

Note 1: The 0 bit means random noise having a unit variance is used for excitation.

Note 2: The total number of bits includes 65 bits for the MELP standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 5 — Mode 5 Quantization Table for Narrowband VDR

Peak Amplitude of Pitch-Filtered Residual (# of Bits)		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 3)	Instantaneous Data Rate (kbps)
		0.1-1.5 kHz (34)	1.5-2.0 kHz (12)	2-3 kHz (24)	3-4 kHz (24)		
Complex Waveform	9	9x34=306	8x12=96	7x24=168	Not transmitted (note 2)	635	28
	8	8x34=272	7x12=84	6x24=144		565	25
	7	7x34=238	6x12=72	5x24=120		495	22
	6	6x34=204	5x12=60	4x24=96		425	19
	5	5x34=170	4x12=48	3x24=72		355	16
Simple Waveform	4	4x34=136	3x12=36	0 (note 1)		237	11
	3	3x34=102	0 (note 1)	0 (note 1)		167	7

Note 1: The 0 bit means random noise having a unit variance is used for excitation.

Note 2: The un-transmitted spectral components are replicated by the transmitted spectra in the lower bands.

Note 3: The total number of bits includes 65 bits for the MELP standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 6 — Mode 4 Quantization Table for Narrowband VDR

Peak Amplitude of Pitch-Filtered Residual (# of Bits)		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 3)	Instantaneous Data Rate (kbps)
		0.1-1.5 kHz (34)	1.5-2.0 kHz (12)	2-3 kHz (24)	3-4 kHz (24)		
Complex Waveform	9	9x34=306	8x12=96	Not transmitted (note 2)	467	21	
	8	8x34=272	7x12=84		421	19	
	7	7x34=238	6x12=72		375	17	
	6	6x34=204	5x12=60		329	15	
	5	5x34=170	4x12=48		283	13	
Simple Waveform	4	4x34=136	3x12=36		237	11	
	3	3x34=102	0 (note 1)		167	7	

Note 1: The 0 bit means random noise having a unit variance is used for excitation.

Note 2: The un-transmitted spectral components are replicated by the transmitted spectra in the lower bands.

Note 3: The total number of bits includes 65 bits for the MELP standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 7 — Mode 3 Quantization Table for Narrowband VDR

Peak Amplitude of Pitch-Filtered Residual (# of Bits)		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)	
		0.1-1.5 kHz (34)	1.5-2.0 kHz (12)	2-3 kHz (24)	3-4 kHz (24)			
Complex Waveform	9	9x34=306	Not transmitted (note 1)				371	16
	8	8x34=272					337	15
	7	7x34=238					303	13
	6	6x34=204					269	12
	5	5x34=170					235	10
Simple Waveform	4	4x34=136					201	9
	3	3x34=102					167	7

Note 1: The un-transmitted spectral components are replicated by the transmitted spectra in the lower bands.

Note 2: The total number of bits includes 65 bits for the MELP standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 8 — Mode 2 Quantization Table for Narrowband VDR

Peak Amplitude of Pitch-Filtered Residual (# of Bits)		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)	
		0.1-0.7 kHz (15)	0.7-2.0 kHz (31)	2-3 kHz (24)	3-4 kHz (24)			
Complex Waveform	9	9x15=135	Not transmitted- MELP used above 0.7 kHz (note 1)				200	9
	8	8x15=120					185	8
	7	7x15=105					170	8
	6	6x15=90					155	7
	5	5x15=75					140	6
Simple Waveform	4	4x15=60					125	6
	3	3x15=45					110	5

Note 1: The band from 0.7 kHz to 4 kHz is then derived not from spectral replication but from that region of the 2.4-kbps MELP signal.

Note 2: The total number of bits includes 65 bits for the MELP standard, pitch gain, residual peak amplitude, and the operating mode selector.

### 3.4.5 Summary of Bit Allocation for Narrowband VDR

Table 9 below gives the overall bit allocation for Narrowband VDR. In addition to the spectral components given in Tables 4 through 8, there is the MELP standard, pitch gain, residual peak amplitude, and the data rate selector. Note that VDR derives the LPC coefficients (in the form of line spectral pairs) and the pitch directly from the MELP bitstream to save bits in the VDR portion of the bitstream.

Table 9 — Overall Bit Allocation for Narrowband VDR

2.4 kbps MELP standard	54
Pitch gain	3
Residual peak amplitude	5
Operating mode selector	3
Variable number of spectral components given in Tables 4 through 8	variable

### 3.5 Wideband VDR

#### 3.5.1 Perceptual Differences Between Wideband Speech and Narrowband Speech

Narrowband speech is not as good as wideband speech in terms of intelligibility and in terms of a perceptual quality. If we hear speech over FM radio, the sound quality is spacious, crisp, with sharp stop consonants. If we hear speech over AM radio, the speech sounds muddy and fuzzy and lacks in tonal definition. Table 10 summarizes perceived differences between wideband and narrowband speech.

Table 10 — Comparison Between Narrowband Speech and Wideband Speech

	WIDEBAND (0 to 8 kHz) SPEECH	NARROWBAND (0 to 4 kHz) SPEECH
SOUND QUALITY	Comparable to FM radio broadcast	Comparable to AM radio broadcast
	Generally crisp and spacious	Generally muffled and constricted
	Spectrally balanced sound	Bass heavy sound
SPEECH INTELLIGIBILITY	Good for female and male speech	Poor for female speech
	Tolerant to noisy speech	Significant degradation for noisy speech

Since narrowband speech is less intelligible than wideband speech, even in an ideal quiet environment, we at NRL developed methods to improve narrowband speech intelligibility by giving it some wideband speech characteristics. We did this for fricatives (/s/, /sh/, /ch/, etc.) by spreading some of the high-frequency speech energy into the lowband region. We used two different approaches: one, by exploiting the aliasing phenomenon [7], and another, by transferring the spectrum [5 (Section 1)]. For both methods, we increased intelligibility by as much as 4 points on the DRT (indicating a substantial improvement) for female speech encoded at 2.4 kbps.

For the wideband VDR discussed in this report, we encode and transmit the upper-band (4 to 8 kHz) speech information. Note from Fig.9 that upper-band speech energies occur intermittently in contrast to narrowband speech energies which are usually continuous; which means that, encoding wideband speech (0 to 8 kHz) does not produce twice as much data as encoding narrowband speech (0 to 4 kHz), although the bandwidth is twice as large. It is significant to note that when the network is too congested, the wideband VDR can be converted to the narrowband VDR by discarding the upper-band speech data.

#### 3.5.2 Conceptual Flow Diagram

Our approach for the wideband VDR is best explained through the simplified flow diagram shown in Fig. 12.

- Speech is divided into two frequency bands: the lower band from 0 to 4 kHz, and the upper-band from 4 to 8 kHz.
- The lower-band speech is encoded using the narrowband VDR.
- The upper-band speech is encoded using the approach discussed in the Upper-Band Encoding (Noise Excited LPC) section.

We once experimented with a wideband voice algorithm having a fixed rate of 48 kbps [5 (Section 2)]. It was created by summing upper-band speech data encoded at 16 kbps with the lower-band speech data encoded with the 32 kbps Adaptive Differential PCM (ADPCM); which is included in the current Secure Telephone Equipment (STE). The intelligibility of female speech in a quiet environment was improved by 4.1 DRT points by adding the upper-band speech information. The intelligibility of female speech in the destroyer environment (noisy ambient) was improved by 8.5 points. These significant improvements show the importance of the upper-band speech data.

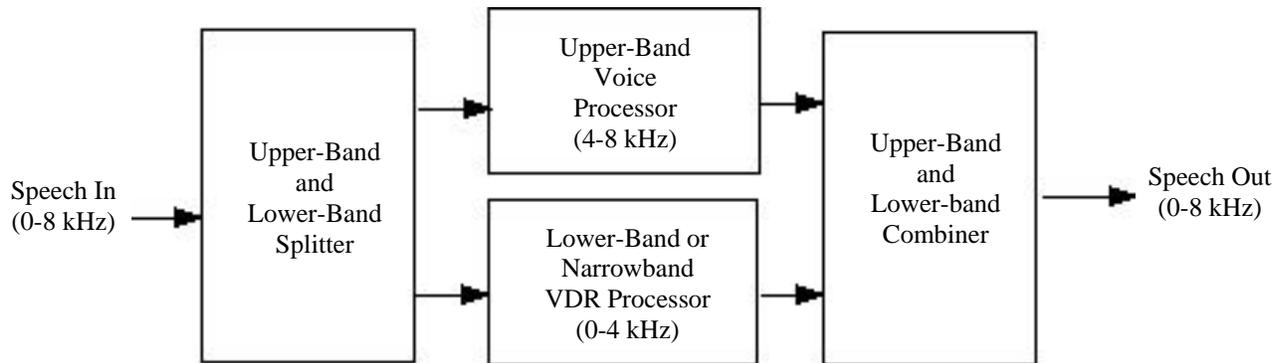


Fig. 12 — Conceptual flow diagram of wideband VDR, which consists of the existing narrowband VDR combined with an upper-band voice processor. The upper-band voice processor encodes speech from 4 to 8 kHz. The wideband VDR can readily be converted the narrowband VDR by dropping the upper-band data.

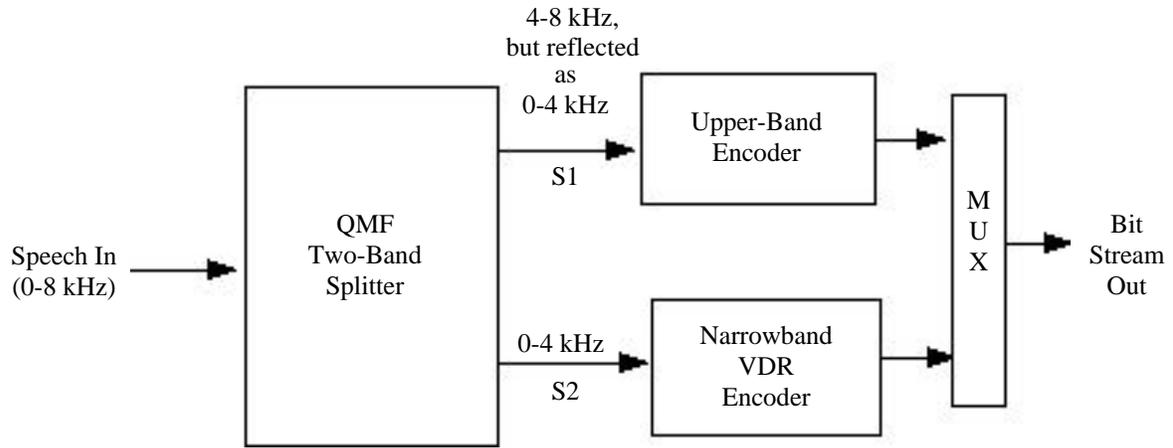
### 3.5.3 Block Diagram

Figure 13 shows the block diagram of the wideband VDR process. It is critical to note that the upper-band VDR process is actually performed at the lowband because the band splitter in Fig. 12 flips the upper-band spectrum into the lowband, as will be shown. Performing upper-band speech coding in the lowband is beneficial because after downsampling in the splitter, the speech sampling rate is 8 kHz, whereas the sampling rate for the original upper speech signal would be 16 kHz, requiring a higher data rate to encode the same amount of information.

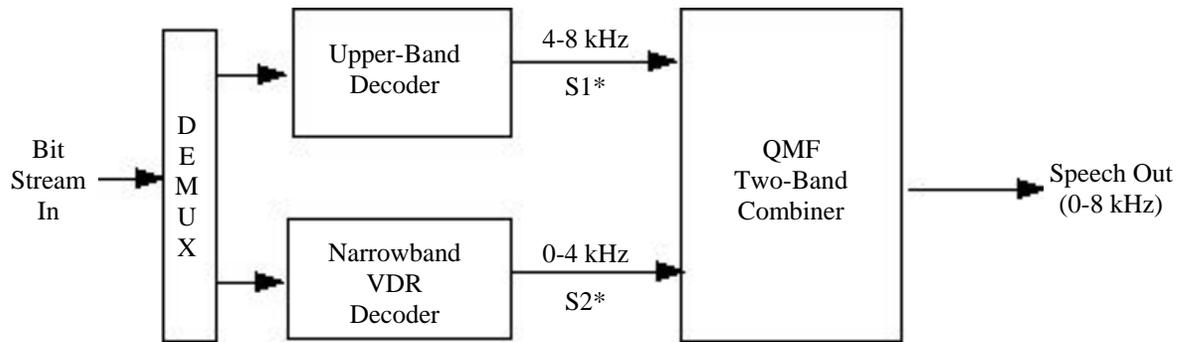
This remarkable technique for splitting a given frequency band into an even number of subbands was advanced by Estaban and Garlans [8]. It was originally developed for encoding the speech waveform from each sub-band with a different resolution to take advantage of the fact that human hearing sensitivity decreases with the increase of frequency. Estaban and Garlans once produced speech encoded at a data rate of 9.6 kbps that almost sounded like unprocessed speech.

### 3.5.4 Quadrature Mirror Filter

In the subband decomposition process, the upper-band signal spectrum is reflected into the lower-band. This is accomplished not by modulating the signal by sinusoidal functions, but by passively filtering the signal using the Quadrature Mirror Filter (QMF) technique, then upsampling and downsampling the filtered outputs [8]. The QMF filtering operation begins with a perfectly matched pair of a low-pass filter and high-pass filter (shown in Fig. 14) and exploits the aliasing phenomenon that flips the upper-band spectrum into the lowband frequency band, and vice versa. If the computation accuracy is high enough, back-to-back bandsplitting and recombination produces the ideal result of the output speech equaling the input.



(a) Wideband VDR Encoder



(b) Wideband VDR Decoder

Fig. 13 — A block diagram of the wideband VDR. The quantity  $S_2$  is the lowband speech waveform. It is identical to the input signal of the narrowband VDR. See Fig. 15 for spectral comparisons.  $S_2^*$  is the quantized version of  $S_2$ .  $S_1$  is the modified upper-band speech waveform in which its spectrum is flipped over into the lowband. See Fig. 15 again for spectral comparisons.  $S_1^*$  is the quantized version of  $S_1$ .

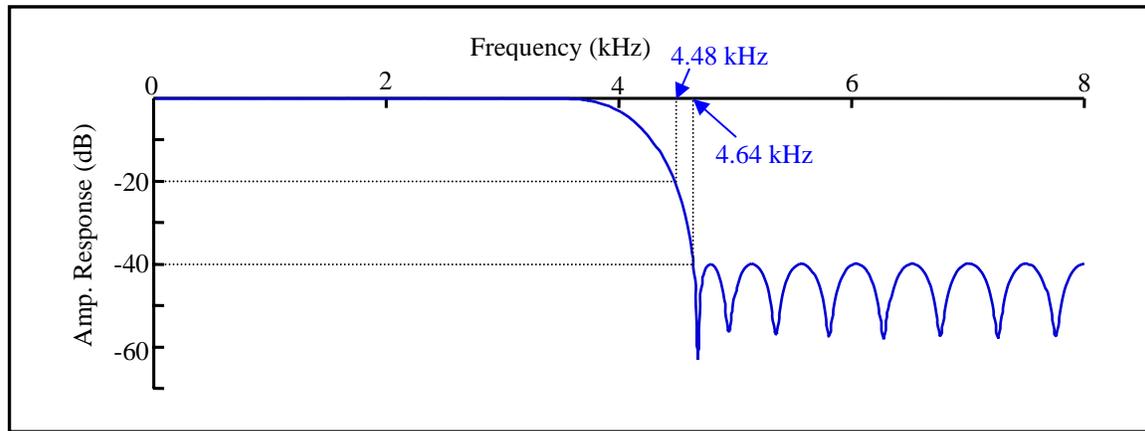
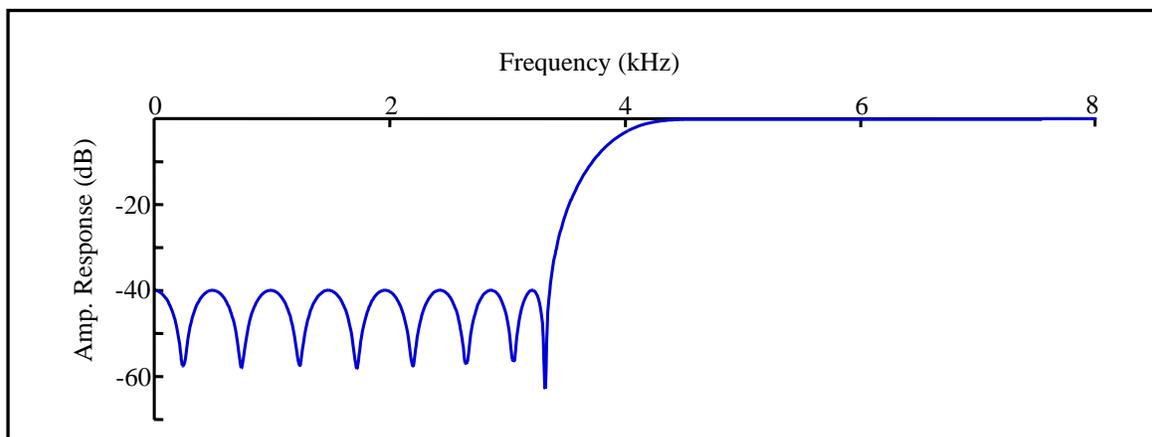
(a) Low-Pass Filter,  $H_1(z)$ (b) High-Pass Filter,  $H_2(z)$ 

Fig. 14 — Frequency response of the 32-tap low-pass filter and high-pass filter we used in the QMF frequency bandsplitting

### 3.5.5 Lower-Band and Upper-Band Decomposition of Speech

The lowband spectrum of the input signal remains as the lowband output of the Two-Band Splitter (compare Fig. 15(b) with the lower half of Fig. 15(a)). The upper-band spectrum of the input, however, is flipped over into the lowband frequency region (compare Fig. 15(c) with the upper half of Fig. 15(a)).

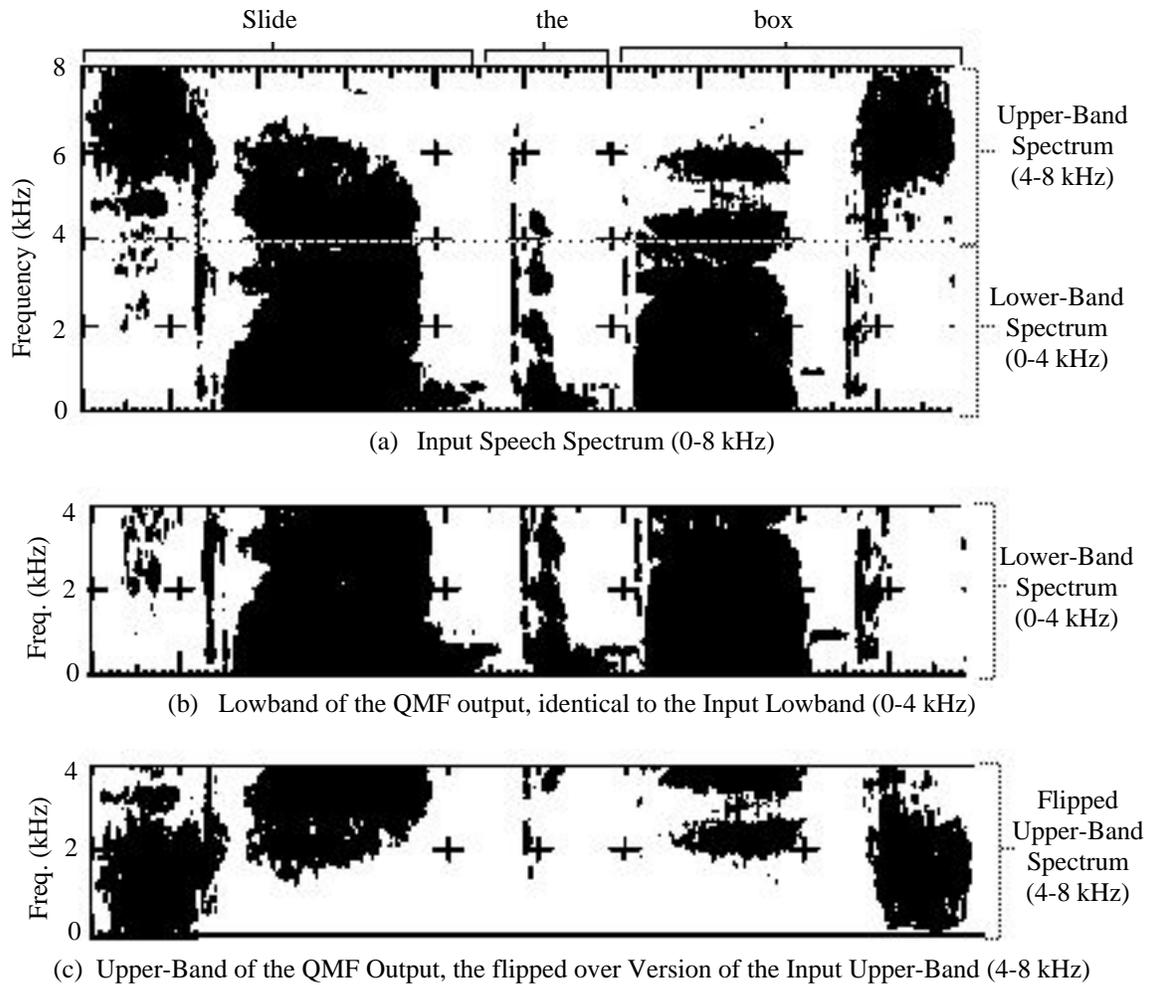


Fig. 15 — Spectral comparison of the input and output spectra of the QMF filters. The lowband of the input spectrum (0-4 kHz) remains as a 0-4 kHz spectrum after the decomposition process. Compare Fig. 15(b) with the lower half of Fig. 15(a). However, the upper-band of the input spectrum (4-8 kHz) is flipped over as a lowband spectrum after the decomposition process. Compare Fig. 15(c) with the upper half of Fig. 15(a). The flipped-over spectrum will be properly re-positioned to the original upper-band location after QMF recombination.

The wideband VDR speech is a sum of the lowband speech data and the upper-band speech. The VDR speech could be generated without the upper-band process but not without the lower-band (or narrowband) VDR processor. The lower-band speech (0-4 kHz) is encoded by the Narrowband VDR presented in Section 3.4 earlier.

### 3.5.6 Upper-Band Encoding (Noise Excited LPC)

With the band-splitter using the QMF filters, the upper-band speech spectrum is reflected in the lowband (where its spectrum is the mirror image of the upper-band spectrum as illustrated in Fig. 15.). To encode the upper-band speech, we use a noise excited LPC; a completely autonomous processor separate from the narrowband VDR. To reduce the data to encode the upper-band speech, the following modifications are implemented.

**No pitch prediction:** the noise excited linear predictor for the upper-band speech has no pitch prediction because upper-band speech is mostly aperiodic waveforms (fricatives and

consonants). Even if there are some periodic speech waveforms (upper end of vowel spectra for female voice), pitch prediction becomes less effective for frequencies above 4 kHz.

**Reduced number of prediction coefficients:** The number of filter weights for the upper-band speech will be reduced to 6 from 10 because there are no sharp resonant frequencies or complex spectral patterns above 4 kHz.

**Reduced quantization range for each filter coefficient:** Each coefficient will have a more limited quantization range (see Fig. 16) in comparison with lowband filter coefficients. The preferred range for each coefficient was determined by the analysis of a large quantity of male and female speech samples. In this case, each of the filter coefficients is quantized with 4 bits.

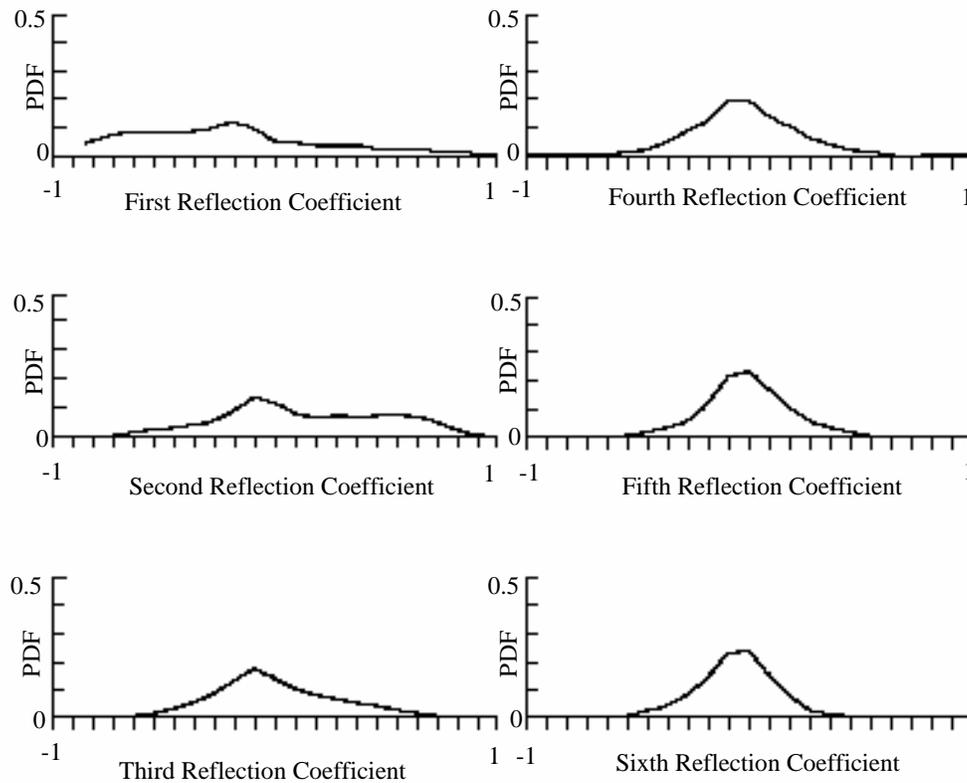


Fig. 16 — Probability Density Functions (PDFs) of six reflection coefficients derived from the upper-band speech at the QMF filter output. These PDFs are generated from speech uttered by 54 males and 12 females, each uttering five 3-second sentences.

**Reduced dynamic range of the prediction residual:** Since the dynamic range of the upper-band speech is nearly always lower in comparison with the lowband speech, the upper-band residual amplitude will only be quantized with 4 bits. However, to ensure good upper-band speech quality, it will be estimated and encoded 10 times per frame (every 2.25 ms). This amplitude parameter will be used as a gain coefficient 10 times per frame when synthesizing the noise at the receiver.

### 3.5.7 Application of Wideband to All Modes of Narrowband VDR

Table 11 below gives the overall bit allocation for wideband VDR. Because the upper-band analysis is a completely autonomous process that is separate from the narrowband VDR data, we can apply the upper-band data to any mode of the narrowband VDR. We do not have to just add the wideband data at the end of the full rate bitstream. Instead, by only setting the wideband flag in the bitstream, we can add upper-band data to only the 2.4 kbps MELP mode, or any other modes, etc., depending on the requirements of the system. If desired, this flexibility can give users the option of choosing wideband speech quality without having to use the highest data rate mode possible.

Table 11 — Overall Bit Allocation for Wideband VDR

2.4 kbps MELP standard	54
Pitch gain	3
Residual peak amplitude	5
Data rate selector	3
Variable number of spectral components given in Tables 4-8	variable
6 upper-band reflection coefficients (4 bits each)	24
10 upper-band residual amplitudes (4 bits each)	40
1 upper-band flag switch	1

## 4. CONCLUSIONS

Many voice compression algorithms have been developed in recent years. Unfortunately, few of these algorithms are mutually interoperable. Many of these new compression algorithms are being used in commercial telephony, especially cellular communications. In these implementations interoperability is achieved by translating between compression algorithms at the connections to the larger public telephone network. When speech is highly compressed, as in many military environments, transcoding between algorithms results in poor speech quality. When security is required, transcoding prevents end-to-end encryption of the speech and adds complexity and cost to the system design. Having one voice compression algorithm which operates at a wide range of data rates greatly simplifies system design, enhances security, and makes it possible for diverse military communications systems and devices to directly interoperate.

In this report, a remarkable concept for a voice communication architecture is presented. The key element, called the Variable-Data-Rate (VDR) voice processor, makes it possible to replace a mass of incompatible voice processors with a single interoperable voice communication resource. This voice processing algorithm is designed to function over all network types but achieves the greatest efficiency over packet networks. It is based on some novel ideas and some that date back to the early 1980s.

## 5. ACKNOWLEDGMENTS

The authors wish to thank Yuh-Ling Su of SPAWAR, Program Executive Office C<sup>4</sup>I and Space, for supporting the voice processing research and much of the secure voice system development program conducted at NRL.

The authors express their thanks to Stuart Shoup, Kevin Thompson, and their associates of the SPAWAR Engineering Facilities in St. Juliens Creek, Chesapeake, VA, for promoting the VDR concept to the Navy and other service organizations.

**REFERENCES**

1. Joint Staff (J6I), "C<sup>4</sup>I for the Warrior," Pentagon, Washington, DC 20318-6000, 1992.
2. G.S. Kang, "Variable-Data-Rate Voice Encoder for Voice Over Internet Protocol (VoIP)," NRL/FR/5555--01-10,016, December 2001.
3. P. Lodefoged, "Elements of Acoustics Phonetics," The University of Chicago Press, Chicago and London, 1974.
4. G.S. Kang and D.A. Heide, "Transcoding Between Two DoD Narrowband Voice Encoding Algorithms (LPC-10 and MELP)," NRL/FR/5550--99-9921, October 1999.
5. G.S. Kang, L.J. Fransen, T.M. Moran, and D.A. Heide, "Voice Processing Techniques for C<sup>4</sup>I Applications," NRL/FR/5550--97-9852, September 1997.
6. G.S. Kang, L.J. Fransen, and E.L. Kline, "Multirate Digital Voice Communication Processor," United States Patent 4330689, May 18, 1982.
7. G.S. Kang and S.E. Everett, "Improvement of Narrowband Linear Predictive Coder, Part 1: Analysis Improvements," NRL Formal Report 8645, December 1982.
8. D.C. Estaban and C. Garlans, "Application of Quadrature Mirror Filters to Split Band Voice Coding Scheme," *Proc. 1977 IEEE Int. Conf. Acoust Speech Signal Process.* May 9-11, Hartford CT. 191-195 (1977).