# MEDIUMBAND SPEECH PROCESSOR WITH BASEBAND RESIDUAL SPECTRUM ENCODING

George S. Kang
Lawrence J. Fransen
Evans L. Kline

NAVAL RESEARCH LABORATORY
WASHINGTON, D.C. 20375

ABSTRACT

The Navy has developed a Multirate Processor (MRP) which generates digitized speech at 2.4, 9.6, and 16 kb/s by the linear predictive coding principle. This multirate capability is achieved by embedding the 2.4 kb/s data in the 9.6 kb/s data stream and the 9.6 kb/s data in the 16 kb/s data stream. Conversion between the rates is accomplished by truncating a certain portion of the bits from the higher-data rate signal or appending extra bits to the lower-data rate signal. The MRP mediumband (9.6 kb/s or 16 kb/s) mode is a baseband residual excited LPC in which the baseband residual is transmitted in terms of Fourier spectral components. Under various operational conditions, the Diagnostic Rhyme Test (DRT) scores for the 9.6 kb/s rate of the MRP compare favorably to the DRT scores of an existing 16 kb/s rate Continuously Variable Slope Delta (CVSD) encoder.

INTRODUCTION

The MRP integrates both narrowband (2.4 kb/s) and mediumband resources into a single capability to provide satisfactory communicability over a wide range of operational conditions. The processor employs a single voice processing algorithm for generating both narrowband and mediumband speech data. The unique characteristic of the MRP is that the bit-stream of the 16 kb/s data contains the bit-stream of the 9.6 kb/s data as a subset. Likewise, the bit-stream of the 9.6 kb/s data also contains the bit-stream of the 2.4 kb/s data as a subset. This embedded data structure makes it possible to interconnect, without user intervention, narrowband and mediumband systems via a digital rate-converter located somewhere along the link. The direct rate conversion allows end-to-end encryption of speech data and eliminates the need of analog tandeming (and resulting speech degradation). During overloaded or disrupted channel conditions, communication survivability may be increased by rate reduction and/or re-routing through other available narrowband communication links.

Since 1977, the Navy has undertaken a number of projects related to MRP including: system analysis, technology development, demonstration models, and mediumband voice algorithm development. Other researchers also found useful applications of the MRP. For example, Bially and his associates applied the MRP concept to packetized speech communications [1]. Others investigated various mediumband voice algorithms based on the embedded data structure: a channel vocoder and sub-band coder model by Gold [2], an adaptive predictive coder model by Goldberg [3], an adaptive transform coder model by Berouti and Makhoul [4], and a hybrid model (a combination of the baseband speech and the upperband synthetic speech) by Sambur [5] which is conceptually similar to that of Watkins [6].

The MRP mediumband voice algorithm presented in this paper is a direct extension of a narrowband linear predictive coder (LPC). The only difference between the narrowband mode and the mediumband mode is in the nature of the excitation signal (Figure 1). The MRP performs the linear prediction analysis as defined for a narrowband mode, and it uses a common synthesis filter for all rates. In the narrowband mode, the

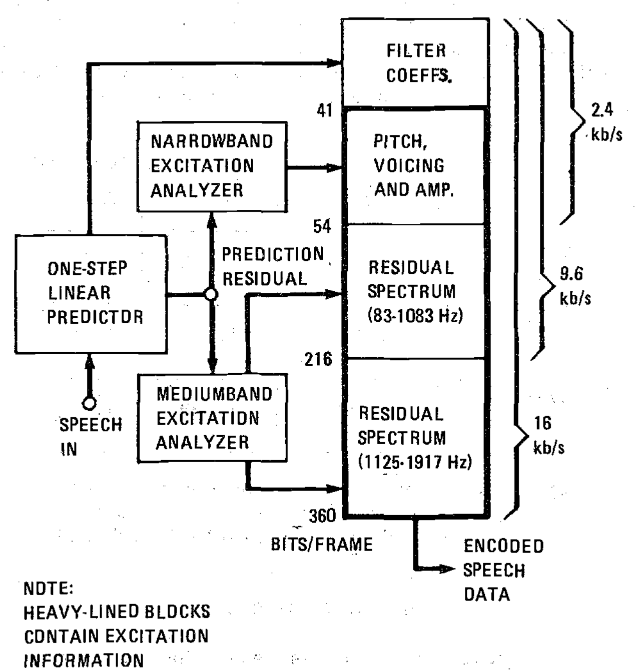


Figure 1 Block Diagram of MRP transmitter

synthesis filter is driven by a conventional pitch-excitation signal. Filter coefficients and excitation parameters are so encoded that the MRP is interoperable with other Department of Defense (DOD) narrowband voice processors. In the mediumband mode, the pitch-excitation signal is replaced by a close replica of the prediction residual. Although only two mediumband data rates are considered in this paper, the MRP voice processing algorithm can generate other data rates greater than 9.6 kb/s at a 0.25 kb/s increment. The algorithm, however, is not well suited for data rates between 2.4 kb/s and 9.6 kb/s.

The MRP was implemented for real-time operation and was tested extensively by DOD in early 1980. In addition, the Naval Research Laboratory (NRL) conducted the Diagnostic Communicability Test (DCT) to assess subjective user reactions. Results from these tests are summarized in this paper.

## OVERVIEW OF MRP MEDIUMBAND VOICE ALGORITHM

The MRP mediumband mode is a baseband residual-excited LPC in which the baseband residual is transmitted in terms of Fourier spectral components, and spectral flattening is accomplished by spectral replication.

A baseband excitation method was selected for the MRP because it produced a generally more favorable 9.6 kb/s speech quality than that obtainable by a procedure using the entire residual and a one-bit quantizer. Contrary to previous baseband residual-excited LPCs which encode time-samples [7,8], the MRP mediumband mode encodes Fourier components of the baseband residual. The spectral encoding approach yields several benifits: (1) low-pass filtering is not required, (2) high-pass filtering is not necessary, (3) down-sampling is not needed, (4) extreme low-frequency components, not essential to speech communications, can be omitted from encoding to save bits , (5) the data rate can be changed at a small increment equal to the transmission rate of each spectral component (6 bits/frame), (6) the overhead data (sync bits, error protection bits, and side information bits) may be incremented at 6 bits/frame, which makes the bit-tradeoff between the speech data and the overhead data flexible, (7) each complex spectral component quantized vectorially into 6-bits allows for more error-resistant coding, and (8) spectral flattening can be realized by simple spectral replication.

The spectral flattening method has a drawback similar to other known methods. Since the baseband spectrum is not replicated at a multiple of the fundamental pitch-frequency, the composite spectrum is not expected to have evenly spaced pitch harmonics for voiced speech. The human ear is sensitive to this kind of pitch deformation. The unnatural tonal quality, however, may be supressed to an acceptable level by making the baseband bandwidth large enough (i.e., 1083 Hz in the 9.6 kb/s mode of the MRP). The human ear is somewhat deficient in crosscorrelating the upperband and the lowerband as demonstrated by coders of Sambur [5] and Watkins [6], where the output is a superposition of the lowband speech and high-pass filtered narrowband speech. The upperband pitch-frequency is not only approximate, but it is also phase incoherent with that of the lowband. Nevertheless, both devices give satisfactory performance in a conversational mode.

The maximum amplitude spectral component within the baseband is used as a frame-by-frame spectral normalization factor. It is semi-logarithmically encoded to six bits with an additional three bits protecting the three most significant bits. Each complex frequency component is encoded to 6 bits by selecting one of 64 points within a unit circle. The 64 points are distributed in a similar pattern as the scatter diagram of the residual complex spectrum. Each point is so indexed that a single-bit error leads to decoding of a neighborhood point. Table 1 lists the design parameters of the MRP.

Table 1 MRP Design Parameters

| | |
|---|---|
| **GENERAL INFORMATION** | |
| Speech samples | |
| Sampling rate (kHz) | 8 |
| Frame rate (Hz) | 44.444 |
| Frame size (samples) | 180 |
| Residual samples | |
| Time overlap (samples) | 12 |
| Window size (samples) | 192 |
| Complex frequency components (samples) | 96 |
| Frequency component spacing (Hz) | 41.67 |
| Baseband spectral indices | |
| 9.6 kb/s mode | 3-27 |
| 16 kb/s mode | 3-47 |
| Baseband bandwidth (Hz) | |
| 9.6 kb/s mode | 83-1083 |
| 16 kb/s mode | 83-1917 |
| **ENCODED DATA (bits/frame)** | |
| 2.4 kb/s mode | |
| Sync bit | 1 |
| Excitation parameters | 12 |
| 10 filter weights if voiced, or 4 filter weights with error protection if unvoiced | 41 |
| Total ... | 54 |
| 9.6 kb/s mode | |
| All of the above (embedded) | 54 |
| Additional sync bits | 3 |
| Maximum amplitude spectral component | 6 |
| Error protection for maximum amplitude | 3 |
| 25 complex frequency components | 150 |
| Total .. | 216 |
| 16 kb/s mode | |
| All of the above (embedded) | 216 |
| Additional sync bits | 4 |
| 20 additional complex freq. components | 120 |
| Error protection for filter coefficients | 20 |
| Total .. | 360 |

In addition to error protection for the excitation signal, filter coefficients are also protected. The 9.6 kb/s mode relies on the error protection furnished by the 2.4 kb/s mode; namely, the more sensitive bits of the first four reflection coefficients are protected if speech is unvoiced [9,10]. At the 16 kb/s mode, 20 additional bits perform the same error protection for the filter weights independent of voicing decision.

## COMPUTATIONAL PROCEDURES

At the transmitter, 180 prediction residual samples are generated for each speech frame by using the quantized reflection coefficients and the 180 preemphasized speech samples. The preemphasis factor and the coefficient quantization rules are as specified for the narrowband mode [9,10]. In order to provide time overlapping, the resulting residual samples are loaded into a 192-word first-in-first-out buffer. The 192 samples are trapezoidally windowed with linear amplitude adjustment over the 12 overlapped samples.

The time-to-frequency transform is carried out by a 96-point complex fast- Fourier-transform (FFT). The 192 windowed residual samples are alternately loaded into the real and imaginary parts of a 96-word complex buffer. The resulting FFT output is in the form of

$$C(k)=A(k)+jB(k) \qquad k=1,2,\ldots,96$$

Only the baseband of the Fourier spectrum is computed by descrambling A(k) and B(k) in the frequency domain. Thus,

$$\begin{bmatrix} R(k) \\ X(k) \end{bmatrix} = \begin{bmatrix} A1 \\ B1 \end{bmatrix} + \begin{bmatrix} B1 & -A2 \\ -A2 & B1 \end{bmatrix} \begin{bmatrix} COS(\theta(k)) \\ SIN(\theta(k)) \end{bmatrix} \qquad k=3,4,\ldots,47$$

where
$$A1=A(k)+A(98-k)$$
$$A2=A(k)-A(98-k)$$
$$B1=B(k)+B(98-k)$$
$$B2=B(k)-B(98-k)$$
$$\theta(k)=\pi(k-1)/96$$

and R(k) and X(k) are respectively the real and imaginary components of the windowed residual. Each normalized complex spectral component is encoded to 6 bits by a table look-up.

At the receiver, the baseband residual spectral components are decoded (R'(k), X'(k), k=3,4,...,N), and they are replicated to form the upperband. N is the frequency index corresponding to the upper end of the baseband. As listed in Table 1, N=27 for 9.6 kb/s, and N=47 for 16 kb/s. Thus, for a 9.6 kb/s mode,

$$R'(k+25i)=R'(k)$$
$$X'(k+25i)=R'(k)$$

where i=1,2,3 and k=3,4,...,27. For the 16 kb/s mode,

$$R'(k+45)=R'(k)$$
$$X'(k+45)=X'(k)$$

where k=3,4,...,47. The reverse transform (a complex conjugate operation of the forward transform including the descrambling process) generates 192 reconstructed residual samples, 12 of which are overlapped with the 12 trailing reconstructed residual samples of the preceeding frame, to form the excitation signal for the MRP mediumband mode.

## INTELLIGIBILITY TESTS

Table 2 shows that the average score of the 9.6 kb/s MRP is essentially identical to that of a 16 kb/s CVSD. The MRP performs as good as or better than a 16 kb/s CVSD under acoustic noise interference except helicopter noise where the sound pressure level is approximately 115 dB. In this case, both the 9.6 kb/s MRP and the 16 kb/s CVSD do not perform at an acceptable level of intelligibility.

The intelligibilities of both the MRP 9.6 kb/s and the CVSD 16 kb/s are not impaired by errors as much as 1%. The error performance between these processors, however, cannot be compared directly because of the difference in

Table 2 DRT Scores of 9.6 kb/s MRP and 16 kb/s CVSD

| TEST CONDITIONS | No. of Spkrs | DRT Scores | |
|---|---|---|---|
| | | MRP 9.6 | CVSD 16 |
| Back-to-back mode | 9M/9F | 90 | 89 |
| WITH ACOUSTIC BACKGROUND NOISE | | | |
| Office noise | 3M/3F | 88 | 90 |
| Airborne command post noise | 3M/3F | 85 | 86 |
| Shipboard noise | 3M/3F | 84 | 85 |
| Helicipter noise | 3M/3F | 64 | 70 |
| E3A noise | 3M/3F | 90 | 91 |
| P3C turbo prop noise | 3M/3F | 86 | 85 |
| Destroyer noise | 3M/3F | 78 | 76 |
| Helicipter carrier noise | 3M/3F | 87 | 83 |
| Jeep noise | 3M/3F | 87 | 84 |
| Tank noise | 3M/3F | 86 | 83 |
| WITH TRANSMISSION ERROR | | | |
| 0.5% | 3M/3F | 89 | 90 |
| 1.0% | 3M/3F | 89 | 90 |
| 2.0% | 3M/3F | 85 | 87 |
| 5.0% | 3M/3F | 75 | 85 |
| UNDER TANDEM ARRANGEMENT | | | |
| Self tandem | 3M/3F | 86 | 87 |
| Output into 2.4 kb/s LPC | 3M/3F | 79 | 75 |
| Input from 2.4 kb/s LPC | 3M/3F | 79 | 82 |
| AVERAGE | | 84 | 84 |

data rates. If the error is 5% at 16 kb/s for a given channel, the error rate at 9.6 kb/s is surely less for the same channel. If the error rate is as much as 5% at 9.6 kb/s, the MRP has an option to use a 2.4 kb/s mode.

## COMMUNICABILITY TESTS

While the DRT is an excellent tool for testing the initial consonant, it is not intended to examine user's subjective opinions of speech quality or communicability. Recently, Voirs of Dynastat developed the DCT under the sponsorship of the Navy [11], which was further refined by Schmidt-Nielsen of NRL. Figure 2 shows a partial listing of DCT scores for clear text (indicated by ↓ ), 9.6 kb/s MRP (▲), 2.4 kb/s DOD LPC (■), and a 2.4 kb/s channel vocoder (●). As noted, the user acceptability for the 9.6 kb/s MRP is much greater than that of a 2.4 kb/s LPC.

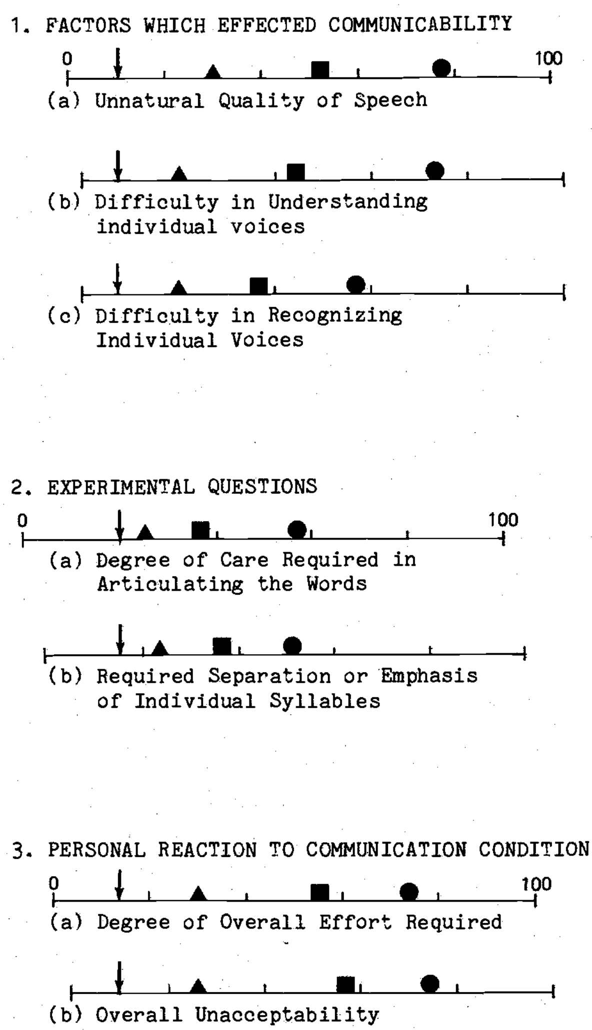


Figure 2 DCT Scores (After Schmidt-Nielsen)

## CONCLUSIONS

The purpose of the MRP is to provide more flexibility in the communication system design by generating several data rates simultaneously. Furthermore, lower-rate data are embedded in higher-rate data to make a direct rate-conversion possible by bit-stripping at a network node if the channel is overloaded or disrupted by natural or man-made interference.

The MRP voice algorithm presented in this paper is one approach to mediumband speech encoding with an embedded data structure. In certain respects, the embedded data structure is a constraint that reduces flexibility in the mediumband encoding. Nevertheless, a 9.6 kb/s mode of the MRP performed as robust as a 16 kb/s CVSD. The ultimate objective of a voice terminal is to provide a reliable, survivable, and robust performance under all operational conditions, particularly in an emergency. The MRP is a step toward reaching this objective.

## ACKNOWLEDGEMENTS

The authors express their appreciation to TRW members, Drs. Zibman, Kau, and Tsou who generated the software for the NRL-owned special processor.

## REFERENCES

1. T. Bially, A. McLaughlin, and C. Weinstein, "Voice Communication in Integrated Digital Voice and Data Networks," IEEE Trans. Commun, Vol. COM-28, No. 9, pp. 1478-1490, Sept. 1980.
2. B. Gold, "Multiple Rate Channel Vocoding," EASCON '78 Record, pp. 693-697.
3. A.J. Goldberg, "2400/16,000 bps Multirate Voice Processor," ICASSP-78, pp. 299-302.
4. M. Berouti and J. Makhoul, "An Embedded-Code Multirate Speech Transform Coder," ICASSP-80, pp. 356-359.
5. S. Sambur, private communication
6. H.E. Watkins, "Description of a Hybrid 7.2 kbps Vocoder," ICASSP-79, pp.546-549.
7. R. Viswanathan, A. Higgins, W. Russell, and J. Makhoul, "Baseband LPC Coder for Speech Transmission Over 9.6 kb/s Noisy Channels," ICASSP-80, pp. 348-351.
8. M. Dankberg and D. Wong, "Development of a 4.8-9.6 kbps RELP Vocoder," ICASSP-79, pp. 554-557.
9. T. Tremain, J. Fussell, R. Dean, B. Abzug, M. Cowing, and P. Boudra, Jr., "Implementation of Two Real Time Narrowband Speech Algorithms," EASCON '78 Record, pp. 698-708.
10. G. Kang, L. Fransen, and E. Kline, "Multirate Processor (MRP) for Digital Voice Communications," NRL Report 8295, Mar. 1979.
11. W. Voirs and M. Clark, "Exploratory Research on the Feasibility of a Practical and Realistic Test of Speech Communicability," Final report prepared by Dynastat, Inc. for Naval Electronic Systems Command, Apr. 1978.