2.4 KB/SEC COMPRESSED DOMAIN TELECONFERENCE BRIDGE WITH UNIVERSAL TRANSCODER

Richard L. Zinser, Philip T. Choong† and Steven R. Koch**

*General Electric Corporate Research and Development
Niskayuna, New York

†Lockheed Martin Missiles and Space
Sunnyvale, California

ABSTRACT

Advanced new technologies, such as cellular-telephone-quality ultra-low-rate speech coders, model domain transcoders, and compressed domain conferencing algorithms provide an opportunity to develop a compressed domain conference bridge system for use in secure, survivable military communications environments. The new conference bridge will allow seamless interoperability with diverse voice terminals and enable full-duplex teleconference operation. Unlike users of half-duplex systems, conferencing participants will be able to talk at the same time and hear the two most relevant simultaneous talkers over a single 2.4 KBPS connection. This paper describes a system architecture that implements the features mentioned above. Compared to conventional multicast conferencing algorithms, the new system will consume a significantly smaller portion of the satellite resources; for N conference participants, conventional multicast requires N^2 channels, while the new system will use only 2N channels.

1. INTRODUCTION

Conference bridging technology has been available for many years to users of the Public Switched Telecommunications Network (PSTN). This technology enables multiple users in remote locations to participate in group discussions. Generally, a summation matrix that supplies an adaptive combination of the incoming signals to each conference participant accomplishes implementation of a conference bridge. The adaptive combination algorithm is designed to attenuate signals from incoming lines that are not actively carrying a voice signal.

In military scenarios, it is desirable to have conference bridge functionality in a secure, survivable communications environment. Because of the unique requirements of this environment, the design and implementation of a strategic teleconference bridge poses several challenges. Most of these challenges are caused by the requirement for digital transmission of speech at 2.4 kb/sec or below. The major issues are:

- Current generation 2.4 kb/sec vocoders are unable to transmit multiple talkers simultaneously. This precludes use of the summation matrix described above.
- Conventional conference bridge designs require decoding the incoming 2.4 kb/sec bit stream to a speech waveform for processing (such as speech activity detection). The speech must then be re-encoded for transmission to the participants. This encode/decode/encode/decode process is known as tandem operation and greatly decreases the subjective quality of the speech.
- Current Federal Standard 2.4 kb/sec vocoders can suffer severe degradation when used in high bit error rate environments. Under worst-case scenarios they require long interleaving lengths to achieve adequate intelligibility.
- The two 2.4 kb/sec Federal Standard vocoders currently in use (LPC-10 [1] and MELP [2]) are incompatible with each other. While new voice terminals will support the MELP algorithm with its vastly improved voice quality, connections between new and existing voice terminals must be made using the "lowest common denominator" (e.g. LPC-10) with the inherent loss in voice quality.

This paper describes a new conference bridge architecture that addresses these issues. Key technologies that enable the design include:

- A new ultra-low-rate flexible vocoder (*TDVC* [3]) that is designed to provide dual simultaneous talker capability at 2.4 kb/sec or highly robust operation in noisy channels.
- A *transcoder* that provides seamless interconnection between any combination of existing Federal Standard vocoders and TDVC with increased voice quality.
- A *compressed domain conference bridge* algorithm that provides full-duplex operation with minimal resource requirements.

This work was sponsored by Lockheed Martin Corporation.

2. SPEECH CODING: TIME DOMAIN VOICING CUTOFF (TDVC)

During the mid-nineties, Lockheed and GE began work on a new algorithm called TDVC (Time Domain Voicing Cutoff) [3]. By combining the best features of time-domain and frequency-domain based vocoders into an easy-to-quantize speech production model, TDVC is capable of maintaining high voice quality at a series of different rates between 1.2 and 2.4 kb/sec. Several innovations have been incorporated into TDVC, including:

- Combined time/frequency domain spectral enhancement for voiced speech;
- Improved low frequency modeling for low-pitched speakers;
- Zero-bit adaptive phase profile for voiced speech;
- A source-matched, bit-efficient channel coding algorithm that has been co-optimized using conditional inter-frame and absolute a-priori bit probabilities [4].

Today, TDVC is a mature algorithm that has been extensively tested. A total of 12 MOS (Mean Opinion Score) and 2 DAM (Diagnostic Acceptability Measure) tests have been completed using 2 independent testing laboratories and 4 different speech sources, including foreign languages. This testing has shown that 1.75 kb/sec TDVC has performance equivalent to the 13.0 kb/sec GSM digital cellular standard. At zero BER (Bit Error Rate), both TDVC and GSM produce the same MOS of 3.6. Furthermore, when the 1.75 kb/sec (source rate) TDVC algorithm is combined with its source-matched 1.25 kb/sec channel coder, the resulting 3.0 kb/sec aggregate-rate system is capable of operation in a 0.07 (7%) BER channel without significant degradation. With this very noisy channel condition, the resultant MOS is 3.45.

TDVC provides two unique features that are essential to the full functionality of the conference bridge design. First, in its 1.2 kb/sec mode of operation, TDVC is capable of transmitting two talkers simultaneously in a single 2.4 kb/sec channel. While the speech quality of TDVC is slightly lower when its rate is reduced from 1.75 kb/sec to 1.2 kb/sec, the magnitude of the loss is small, with an estimated MOS decrease of 0.1 to 0.2, based on testing performed in 1997 [5]. To minimize any degradation, the conference bridge control algorithm will place TDVC in ultra-low-rate operation only when two participants are talking simultaneously; otherwise, one of the higher-rate modes will be employed.

The second feature of TDVC that will be employed is its robustness to channel errors. To achieve full operational capability in a 0.10 BER channel at 2.4 kb/sec, the baseline TDVC encoder will be operated at 1.2 kb/sec with an additional 1.2 kb/sec of forward error correction (FEC) coding. This will yield performance that is even more robust than the 3.0 kb/sec aggregate-rate coder combination described above.

A powerful advantage of TDVC lies in the flexible design of its voice synthesizer (receiver). The TDVC synthesizer is very well suited for using analysis parameters generated by a MELP or LPC-10 transmitter to produce very high quality output speech. When used in combination with the transcoder described below, TDVC can provide remarkable interconnectivity for MELP and LPC-10, while actually improving the subjective quality of LPC-10 speech.

3. COMPRESSED DOMAIN UNIVERSAL TRANSCODER

As mentioned above, one of the issues in designing a military conference bridge is that the current standardized vocoder algorithms have incompatible bit streams. The overall usefulness and total speech quality provided to users can be greatly increased if the conference bridge system is able to "speak" in any vocoder format that will be used on the network.

Previous methods of translating between vocoder formats were accomplished by a primitive tandem connection. The incoming bits were fully decoded to speech and then re-encoded in the new format. This process requires significant computing resources and degrades the speech quality.

Transcoding technology greatly improves this translation process. The transcoder **directly** converts the speech coder parametric information in the compressed domain. There is no need to decode the incoming bit stream to speech samples with the transcoder. Instead, the parametric model parameters are decoded, transformed, and then re-encoded in the new format. The process requires significantly less computing resources than a full tandem decode/encode; in some cases, the CPU time and memory savings can exceed an order of magnitude. Figure 1 shows a block diagram of a MELP to LPC-10 transcoder.

Development of a transcoding algorithm requires more than the obvious transformation of parameters. In the MELP to LPC-10 example, the line spectral frequencies cannot be simply converted to refection coefficients and quantized. First, the preemphasis used in LPC-10 must be superimposed on the MELP spectral coefficients. This is performed via a convolution in the correlation domain. The addition of preemphasis requires an adjustment in the MELP RMS to compensate for the change in gain of the synthesis filter. Loss of spectral acuity in the conversion process necessitates a formant enhancement function for the preemphasized coefficients. In addition, voicing conversion is not straightforward. The MELP overall voicing bit cannot be used directly for LPC voicing decision because of peculiarities in the MELP bandpass voicing detection algorithm. Instead, the decision must include 1) the overall voicing bit; 2) the bandpass voicing strengths; and 3) the spectral tilt, as represented by the first reflection coefficient. Finally, the MELP half-frame RMS estimates cannot be simply averaged to arrive at the LPC RMS. In order to preserve unvoiced stops (such as p, t, and k), the movable RMS analysis window function of the LPC-10 analyzer must be emulated in the transcoder. This is accomplished by adaptively weighting the half frame gains based on the voicing state information.

A further benefit from transcoding derives from using a proprietary vocoder synthesizer (receiver) with an existing vocoder analyzer (transmitter). In many cases, the new synthesizer is capable of producing better quality speech than the old synthesizer. For example, consider the combination of an

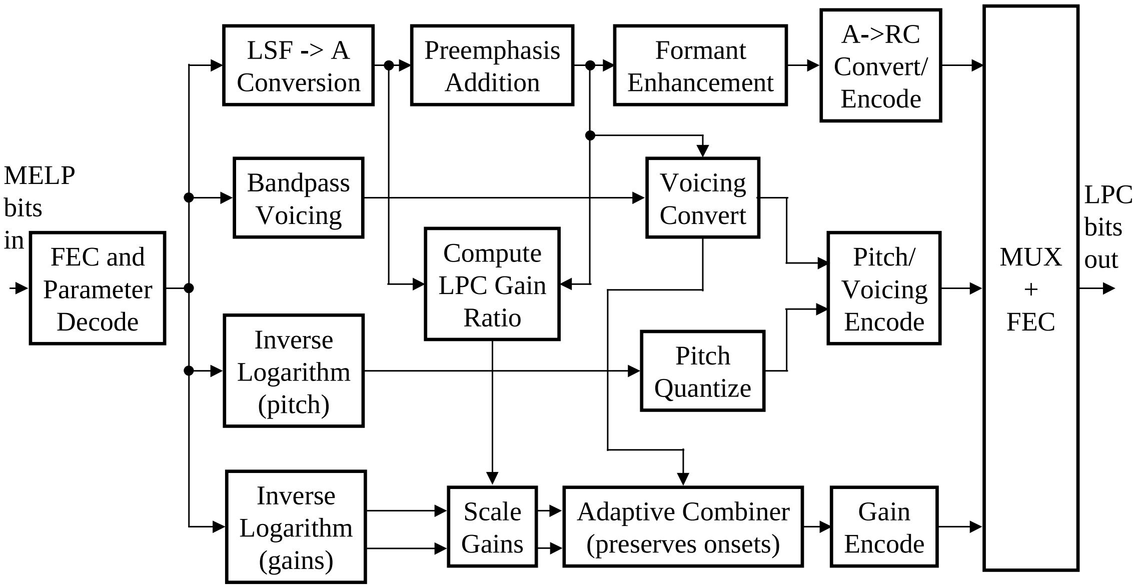

Figure 1: MELP to LPC-10 Transcoder

LPC-10 analyzer feeding data to an LPC/TDVC transcoder, which in turn feeds the transcoded bit stream to a TDVC synthesizer. Experienced LPC-10 users have reported that the output speech quality for this combination is superior to “straight” LPC-10 to LPC-10. This benefit allows users of new equipment to receive a higher QOS (quality of service) even when communicating with older terminals.

4. COMPRESSED DOMAIN CONFERENCE BRIDGE

Utilizing transcoder technology, we have developed a conference bridge algorithm that operates in the compressed domain. (Refer to the block diagram in Figure 2.) The incoming bit streams from each conference participant are first decoded into parametric model data. The parameters for each stream are then analyzed to determine which stream(s) carry an active voice signal by a compressed domain VAD (Voice Activity Detector). The parameters and the VAD decisions are fed to a crossbar weighting and switching matrix that selects and adaptively gain/delay-adjusts the signals that will be transmitted over each output stream. Finally, a dual speaker-capable transcoder transforms the selected speech model parameters into a single 2.4 kb/sec bit stream that is customized for each user. Note that the block diagram also contains switchable FEC encoding and decoding modules for protected mode operation (described below).

The bridge control logic will allow full customization of what is received by each individual user. Each user can receive one or two talkers simultaneously, as selected by a combination of the talker’s pre-set priority, VAD decision, and the receiving user’s pre-set preferences. Each user’s downlink channel will always carry the proper vocoder format; the transcoder will automatically switch input modes as the selected talker changes. Dual speaker mode will be initiated for any receiving user that has dual speaker capability when two of the input channels are active simultaneously. All of these functions will be performed transparently to the users.

The bridge supports the following bit stream formats (all at 2.4 kb/sec):

1. LPC-10
2. MELP
3. TDVC full rate
4. TDVC dual speaker
5. TDVC protected mode (for severe channel error environments)

TDVC protected mode utilizes the TDVC vocoder operating at 1.2 kb/sec with an additional 1.2 kb/sec of source-matched FEC code.

5. APPLICATIONS TO MILITARY STRATEGIC VOICE CONFERENCING (SVC)

Strategic Voice Conferencing (SVC) is a major communication service requirement supported by Milsatcom programs. The U.S. Government is currently shifting the SVC requirement to Milstar and its follow-on AEHF satellites. These are the only two satellite communications systems that are truly survivable under the worst case nuclear scintillation and jamming scenarios. Both systems have a very limited number of channels available at the “survivable” data rate of 2.4 kb/sec.

Perhaps the most significant benefit of the new system is its efficient use of these “survivable” satellite system resources. Current Milstar-supported SVC techniques use a multicast scheme to simulate full duplex operation. Consequently, a teleconference on the existing system consumes valuable

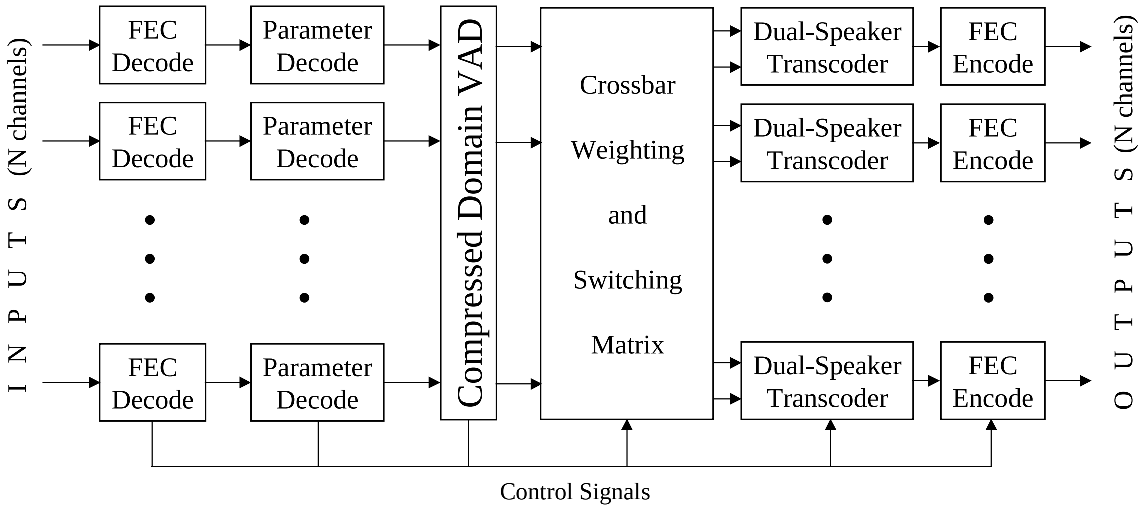

Figure 2: Compressed Domain Conference Bridge

communication channels at a rate that increases with the square of the number of participants. Because the new system design uses only one uplink and one downlink per conference participant, it will require significantly fewer channel resources. To a first order approximation, the channel usage for N conference participants is reduced from O(N^2) to O(2N). For example, a TVDC conference of 10 participants requires 20 channels vice a current conference of 10 that requires 90 channels.

Another benefit can be derived from the superior robustness of TDVC operating in protected mode. Significant intelligibility improvements and reduced delays due to shortened interleaving can be achieved over existing DoD standards during periods of heavy channel data loss.

The system design is intended to provide maximum flexibility with minimum disruption of existing infrastructure. Because of its low computational requirements, the conference bridge can be physically located in a spacecraft or adjacent to a ground command post. The latter location is particularly suited for currently deployed systems. However, locating the TDVC conference bridge function in the satellite fully minimizes the channel resource requirements. It cuts the required capacity for the downlink by a factor of (N-1).

6. SUMMARY

In this paper, we have described a new conference bridge architecture that can be advantageously applied to secure, survivable military communications. The architecture addresses many of the problems posed by the application of conventional bridge technology to the military environment, while maintaining much of the functionality of a commercial product. The new system is expected to deliver significantly higher QOS with a (N-1)/2 reduction in satellite resources consumed. The concept can be readily retrofitted into existing military satellite systems by locating the conference bridge on the ground. However, the maximum performance gains can only be realized if the conference bridge is implemented in space.

7. REFERENCES

[1] T. Tremain, “The Government Standard Linear Prediction Coding Algorithm: LPC-10,” *Speech Technology Magazine*, pp. 40-49, April 1982.

[2] A. McCree, K. Truong, E. George, T. Barnwell, and V. Viswanathan, “A 2.4 kb/sec MELP Coder Candidate for the new U.S. Federal Standard,” *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, pp. 200-203, 1996.

[3] R. Zinser, M. Grabb, S. Koch, and G. Brooksby, “Time Domain Voicing Cutoff (TDVC): A High Quality, Low Complexity 1.3-2.0 kb/sec Vocoder,” *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pp.25-26, 1997.

[4] J. Ross, N. Van Stralen, M. Grabb, S. Koch, R. Zinser, and J. Anderson, "Channel Decoding Short Frames of Voice Data", *Proc. Wireless '98*, July 1998.

[5] R. Zinser, M. Grabb, and S. Koch, “Multiple Source MOS Evaluation of a Flexible Low-Rate Vocoder,” *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, pp. 521-524, 1998.