THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 8, October 1977
Printed in U.S.A.

B.S.T.J. BRIEFS

Use of Variable-Quality Coding and Time-Interval Modification in Packet Transmission of Speech

By S. A. WEBBER, C. J. HARRIS, and J. L. FLANAGAN

(Manuscript received December 14, 1976)

Speech transmission by switched digital packets offers several opportunities for increasing the utilization of transmission capacity. We comment here upon a combination of variable-quality coding and time-interval modification that can efficiently load a transmission facility and accommodate fluctuating demands on it.

Consider, typically, that a conventional voice switch detects speech energy bursts and demarks each as a packet. A time stamp is given to each packet, and the interburst silences are discarded. Each packet is digitally encoded with a quality that reflects service demands being made on the transmission facility at the moment. Coding bit rate and time-stamp are written in the header data for each packet, along with necessary supervisory information, such as destination and source addresses. Successive packets are assembled in a transmit buffer and are transmitted when capacity is available. Figure 1 illustrates the process.

At the receiver, arriving packets are accepted into a receive buffer. The receiver decodes each packet (in accordance with the header bit rate), reassembles the packets in temporal order (according to the time-stamp), and reinserts the silent intervals, not necessarily exactly as in the original, but with a variation that is perceptually acceptable.

Relevant design questions include: (*i*) how much saving in transmission capacity can be achieved by discarding the silent intervals, (*ii*) what range of signal quality is acceptable in digitally coding the packets, (*iii*) what latitude is perceptually acceptable in reconstructing the speech silent intervals, (*iv*) what total round-trip delay time is allowable in a packet system, (*v*) what transmit and receive buffer sizes are required, and (*vi*) what packet sizes are attractive for transmission economy.

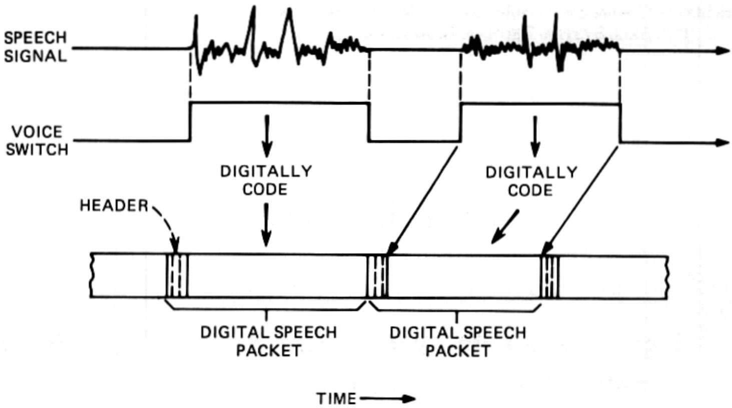

Fig. 1—Speech energy bursts detected by a voice switch, digitally coded, and formed into packets.

Question (*i*) is thoroughly addressed in the extensive literature on Time Assignment Speech Interpolation (TASI) systems. We will add here one more bit of confirmatory data. Questions (*ii*) and (*iii*) dramatically influence the buffer requirements for the system. Our purpose here is to remark about preliminary observations on these issues, and to emphasize these points as candidates for quantitative study.

Extensive data from satellite transmission and echo canceller technology relate to question (*iv*) and suggest that round-trip delays of 0.6 sec, and in some cases up to 1.2 sec, can be used. Questions (*v*) and (*vi*) can properly be addressed only in the context of a complete system design and its optimization. Our remarks, therefore, relate to the points (*i*), (*ii*) and (*iii*). For convenience, all our observations assume that each speech burst is coded as one packet. We implement our experiment by simulation on a laboratory computer, and we process sentence-length signals.

Within-sentence silence time. Our particular voice switch utilizes the Hilbert envelope of the speech signal, and includes a hysteresis logic for positive switch action. The total within-sentence silent time made available by the switching is of course a function of the switch threshold. Too low a threshold provides too little silent time, and too high a threshold eliminates too much signal. Our laboratory observations suggest that within-sentence silent time equal to about 15 percent (of the total sentence duration) can be usefully eliminated. This figure also appears consistent with related studies on voice switching. (Additionally, of course, there are substantial between-sentence silences and natural pauses in conversation flow that can be eliminated.) The sound spectrogram of Fig. 2a shows an input sentence with the significant silent

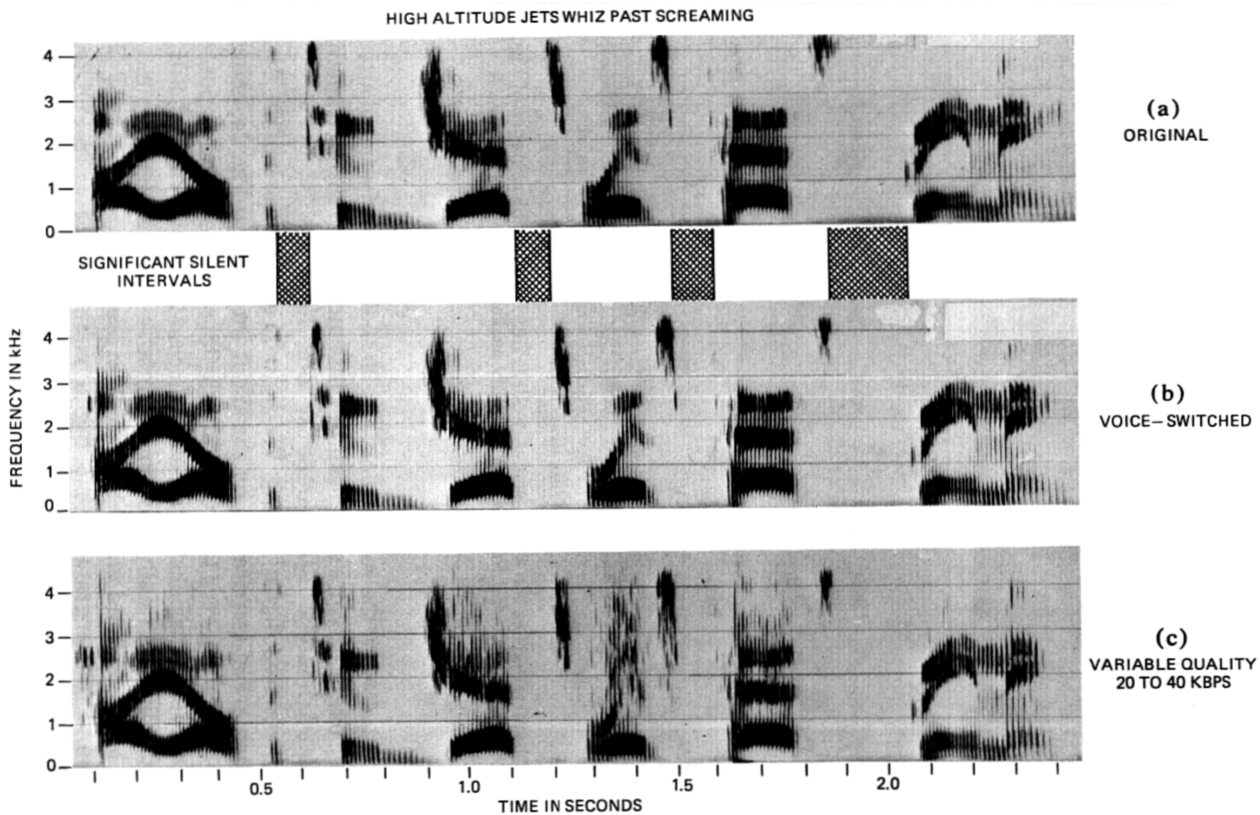

Fig. 2—Experimental transmission of packetized digital speech with variable-quality coding.

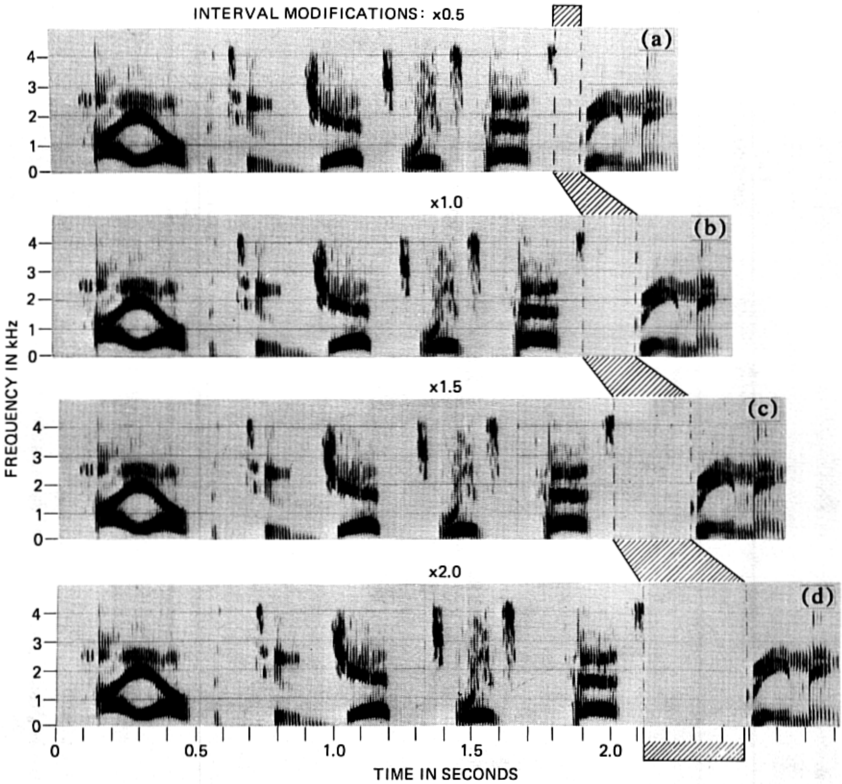

Fig. 3—Modifications of the within-sentence silent intervals for the sentence of Fig. 2. The output is reconstructed from five high-quality packets.

intervals detected by the voice switch. Fig. 2b shows the same signal after passing through the voice switch. In this instance the eliminated silent time is approximately 17 percent of the total duration.

Variable-quality digital coding. We digitally encoded each of the five packets demarked (separated) by the silent intervals in Fig. 2a, using adaptive-differential PCM (ADPCM). The digital coder was also computer simulated. We let the packets be coded successively at bit rates of 40K, 30K, 20K, 30K, and finally back to 40K bits/second, simulating a momentary heavy demand on the transmission system.

The sound spectrogram for this digital coding is shown in Fig. 2c, where the signal packets are reconstructed with silent intervals identical to the original input. One sees that the greatest quantizing noise appears for the momentary quality dip to 20K bits/sec in the third packet. The overall subjective impression of this coding is that the quality is reasonably acceptable.† The perceptual palatability of ADPCM coding also contributes to this result. In this particular instance, the average bit rate for the transmission is 28.6K bits/sec.

Time-interval modification. Latitude in reconstructing the silent intervals in the signal at the receiver can significantly relieve buffer requirements. What modifications in time intervals might be perceptually acceptable? Figure 3 shows receiver reconstruction of high-quality packets with constant, multiplicative modifications of the silent time intervals of 0.5, 1.0, 1.5, and 2.0. (0.0 and 4.0 were also examined, but are not shown here.) One silent interval (the last) is selected and marked for comparison across the signals. Perceptual assessment of these reconstructed packets suggests that interval modifications of the order of ±50 percent are tolerable. This latitude is also large enough to be advantageous in buffer design.* Interval lengthening of more than 200 percent, and shortening down to 0 percent, are clearly not acceptable.

---

† Extensive current work on TASI-D also gives insight about this coding range.

* Additionally, the possibility exists for modifying the durations of the active signal packets (by spectrum-preserving techniques such as the phase vocoder).

Technical material in this note was presented orally to the 93rd meeting of the Acoustical Society of America (J. Acoust. Soc. Am. *61,* S69, June 1977).