Subjective Effects of Cell Losses in Voice Over ATM

David W. Petr, Wei Gary Huang, Mark P. Mischler

Abstract—**We present the results of formal subjective listening tests of voice transported via ATM, emphasizing cell losses. The μ-law voice samples are divided into pairs of ATM cells, one containing most significant bits (MSB cells) and the other containing least significant bits (LSB cells). Based on isopreference testing, we conclude that (1) isolated blocks of 15 or fewer consecutive LSB cell losses (176 ms or less) are *indistinguishable* from the original speech and (2) LSB cell loss rates up to at least 10% (10^{-1}) are *indistinguishable* from the original speech.**

Keywords— **Packet Voice, Voice Quality Assessment**

I. Introduction

With the maturing of ATM technology and the introduction of commercial ATM network services, there has been a resurgent interest recently in voice over ATM (VOA) techniques. A recent overview of VOA alternatives is given in [1]. This paper concerns a simple VOA technique and the potential subjective effects of ATM cell losses on perceived voice quality.

We consider a system in which 64 kb/s digital voice (8 kHz sampling, 8 bit/sample) is packetized by first dividing the samples into the four most significant bits (MSBs) and the four least significant bits (LSBs). Every 11.75 ms, a pair of ATM cells is generated by placing MSBs from 94 samples into the 47-octet payload of one AAL1 cell (MSB cell) and filling another AAL1 cell (LSB cell) with the corresponding LSB data. If an LSB cell is lost in transit, the receiver substitutes zero-level LSB data for the lost payload, to produce (effectively) 4-bit companded PCM. MSB cell losses are not considered here. Using the cell loss priority (CLP) bit in the ATM cell header, MSB cells would receive preference over LSB cells if network congestion necessitated cell discarding. This system was chosen because of its simplicity and potential robustness to LSB cell losses. Other details of the VOA protocol are discussed in [2], [3].

II. Subjective Test Description

A. Isopreference Testing

Our primary goal for the subjective tests was to determine, in a near-ideal listening environment, the point at which ATM cell losses of various types caused the recovered speech to be subjectively *distinguishable* from speech that had sustained no cell losses. This is a much more conservative approach than attempting to determine the point at which the losses become *objectionable*. Hence we chose isoprefence testing as our methodology, instead of the more common mean opinion score (MOS) methodology. For an example of the latter in a packet voice context, see [4].

In isopreference tests, listeners (subjects) are presented with two versions of the same phrase, a "test" phrase and a "reference" phrase. The subjects are required to choose one as being preferable, even if they do not think there is any difference. Hypothesis testing is then used in post-processing the data to determine whether or not the two phrases are statistically distinguishable. Isopreference tests were used in an earlier study of voice packet losses [5].

B. Test Material

The test material was digitized speech sentences approximately three seconds long (approximately 250 MSB/LSB cell pairs) recorded through a standard telephone handset in low-noise surroundings and digitized using 8 kHz sampling and μ-255 PCM coding. The source material consisted of five phonetically balanced sentences spoken by three males and three females (one male and one female spoke the same sentence, and this pair was used exclusively for the test conditions involving added noise). All test phrases were generated by off-line computer processing.

C. Test Conditions

A total of 82 listening conditions was used, preceded by 5 practice conditions intended to familiarize the subjects with the test procedure. Each listening condition consisted of the pattern (Version 1–Version 2–Version 1–Version 2), where one version was the "reference" phrase and the other was the "test" phrase. The mapping from reference/test phrase to Version

This investigation was partially supported by Sprint Corporation.

1/2 was randomized, as was the order of the 82 listening conditions. All 87 conditions were recorded from workstations through analog I/O ports onto a Digital Audio Tape (DAT). The subjects listened to the conditions in a quiet room using standard telephone handsets connected to a DAT player. An auxiliary result of the testing was that the subjects and their listening environment were of sufficient quality that the subjects were able to distinguish original speech from speech with a 40 dB signal-to-noise ratio resulting from additive white noise, a result consistent with the testing in [5].

D. Subject Population

A total of 30 student volunteers participated as subjects in nine different sessions. There were 24 males and 13 native English speakers. Via hypothesis testing of the means, the listener population was determined to be statistically homogeneous at a 1% significance level in terms of sex, native language, and listening session. Similar hypothesis testing procedures were used to determine, for each test condition, a region of indistinguishability (at a 1% significance level) in terms of the proportion of listeners choosing the "test" phrase.

III. Results

For all LSB cell loss conditions, the "reference" phrase was the original speech and the "test" phrase was one in which cell losses had been simulated. For the "sanity check" condition (below), both Version 1 and Version 2 were the original speech. Each of these test conditions consisted of four listening conditions, one for each of four different speakers (two male and two female), resulting in 120 "votes" being cast for each test condition. For all of these conditions, the "reference" and "test" phrases are statistically indistinguishable if the percentage choosing the "test" phrase is between 39% and 61%.

A. Sanity Check

The proportion of users choosing the "test" phrase when the two phrases were in fact identical was 52.5%. This is well within the region of indistinguishability, indicating that there was no bias in the condition ordering.

B. LSB Cell Losses

In the first set of conditions, blocks of N LSB cells are lost with infinite spacing between block losses ($N - inf$ conditions), i.e., only a single block of N LSB cells is lost. The starting position of the losses was the same for each phrase for all values of N, but the loss position was chosen to be different for each phrase (one at the beginning, two in the middle, one at the end of the phrase). All loss positions were deliberately chosen during active periods of the speech and in places that seemed to be sensitive to LSB cell losses.

Figure 1 shows the results of the $N - inf$ conditions for N=1, 3, 5, 10, and 15. All conditions are within the indistinguishable region, although the 15-inf condition is quite close to the distinguishability threshold. We conclude that under typical (less-than-ideal) listening conditions, blocks of consecutive LSB cell losses would be completely undetectable by VOA users for block lengths up to at least 15 cells (176 ms).

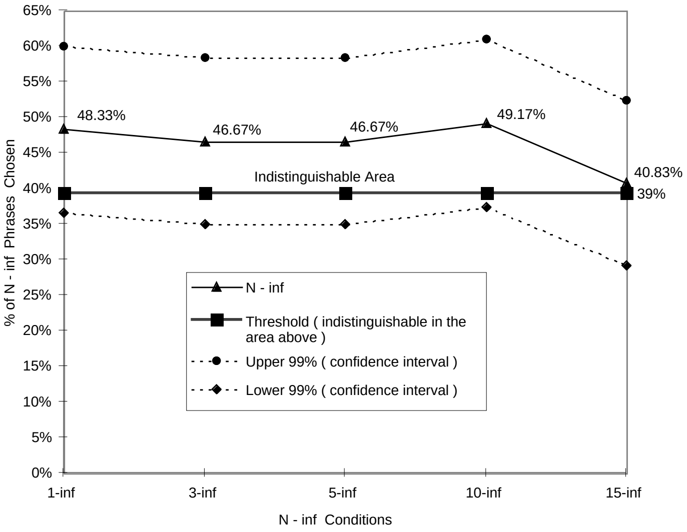

Fig. 1. Results for an Isolated Block of N Lost LSB Cells.

In the next set of conditions, blocks of N LSB cells are lost periodically with a period of m cell times ($N - m$ conditions). Figure 2 shows results for N=1 and various values of m. The results show that <= 10% cell losses are indistinguishable, but 20% losses can be detected. This conclusion also holds for the following $N - m$ combinations: 3-50 (6% loss, 44.2% chosen), 5-50 (10% loss, 45.0% chosen), 3-25 (12% loss, 42.5% chosen), and 5-25 (20% loss, 34.2% chosen). An interesting result is that for a given cell loss ratio, LSB cell losses that occur in larger blocks appear to be *less* distinguishable than losses occurring in smaller blocks, provided the block size is 5 or less.

A final condition was designed to determine whether, in an emergency congestion situation, it is better to simply discard all LSB cells or try to deliver some of them. The tests showed a clear preference for discarding all LSB cells (84.2% chosen) compared with

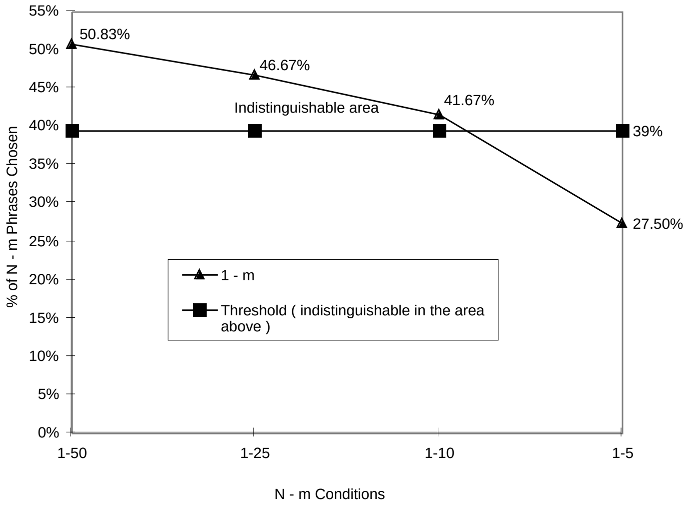

Fig. 2. Results for 1 Loss Out of Every m LSB Cells.

discarding exactly every other LSB cell. This somewhat surprising result (that no LSB information is preferable to some LSB information, specifically LSB information for only every other cell) may be somewhat influenced by the periodic nature of the every-other-cell losses, which created a subjectively annoying ("buzzing") effect.

C. Noise Equivalency

To establish the speech quality during extreme congestion (all LSB cells lost), we determined the subjective equivalence between MSB-only speech and a more common and familiar speech quality measure: signal to noise ratio (SNR). To establish this subjective equivalence, the "reference" phrases were the original phrase with varying amounts of white gaussian noise added, and the "test" phrase was one in which all LSB cells were lost (but no other noise added). For the "reference" phrases, signal to noise ratio (SNR) was calculated as the ratio of the average signal power during active speech to the average noise power, as in [5]. The male and female phrases used for these conditions were the same as those used in [5], resulting in 60 "votes" being cast for each test condition and the indistinguishable region being between 35% and 65% choosing the "test" phrase. The results shown in Figure 3 indicate that speech with all LSB cells lost is subjectively equivalent to speech with an additive-noise SNR of approximately 29.5 dB.

IV. Conclusions

Our results indicate that this simple method of transporting voice over an ATM network should be subjectively very robust in the face of LSB cell losses. The highly conservative nature of the isopreference testing conducted here should give even the most quality-conscious network providers confidence that voice quality using this system would not be adversely affected even for LSB cell loss rates in excess of 10% or isolated bursts of up to 15 consecutive LSB cell losses. An auxiliary result is that the worst speech quality for LSB-only losses would be subjectively equivalent to speech with SNR (additive noise) of approximately 29.5 dB.

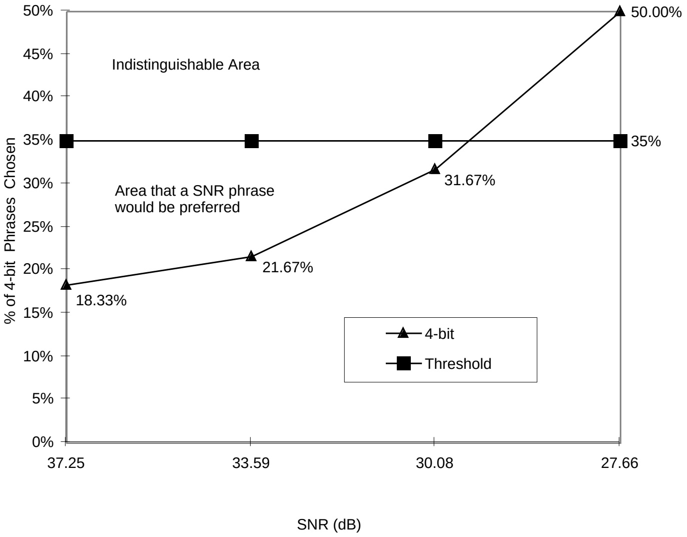

Fig. 3. Results of MSB-only vs. SNR Conditions.

References

[1] D. J. Wright, "Voice over ATM: An Evaluation of Implementation Alternatives", *IEEE Communications Magazine*, Vol. 34, No. 5, pp. 72-80, May 1996.

[2] D. W. Petr, M. Mischler, M. Shanableh, V. S. Frost, "Voice Transport via ATM Networks: Proposed System Solutions", Technical Report TISL-10610-01, Telecommunications and Information Sciences Laboratory, University of Kansas, July 1994.

[3] M. Shanableh, J. B. Evans, D. W. Petr, "Voice Transport via ATM Networks Using DS0 Packetization", Technical Report TISL-10610-02, Telecommunications and Information Sciences Laboratory, University of Kansas, July 1994.

[4] D. O. Bowker and C. A. Dvorak, "Speech Transmission Quality of Wideband Packet Technology", *Proceedings of IEEE Globecom '87*, pp. 47.7.1 - 47.7.3, November 1987.

[5] D. W. Petr, L. A. DaSilva and V. S. Frost, "Priority Discarding of Speech in Integrated Packet Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp. 644-656, June 1989.