This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

# Multiple-Description Speech Coding using Speech-Polarity Decomposition

Stephen Voran and Andrew Catellier
Institute for Telecommunication Sciences
Boulder, Colorado USA
{svoran,acatellier}@its.bldrdoc.gov

***Abstract*—We present and evaluate a new multiple-description coding extension to the international standard for pulse code modulation speech coding (ITU-T Rec. G.711). This extension is inserted between the G.711 encoder and decoder. It uses speech-polarity decomposition to spread the speech signal across two channels thus increasing robustness to channel losses. When both channels deliver their payloads the extension becomes transparent and bit-exact G.711 speech samples are produced—there is *no quality penalty*. Due to low inter-channel redundancy, block coding, and entropy coding, the average total speech payload bit-rate is no greater than the 64 kbps rate of conventional G.711—there is *no rate penalty*. When either channel fails to deliver, the remaining channel still produces intelligible speech with moderately reduced quality thanks to a compressed sine-pulse fill-in algorithm. We are not aware of any other viable multiple-description coding extension that simultaneously meets the opposing goals of *no quality penalty* and *no rate penalty*.**

## I. Introduction

Telecommunications systems often transmit speech signals over lossy channels. Prime examples of lossy channels include noisy and fading radio channels (wireless telephony) and congested packet data networks (Internet telephony). Packet loss concealment (PLC) algorithms can reduce the effects of short-duration channel losses on received speech. PLC is convenient because it can be implemented at the speech decoder alone. But when losses persist beyond 90 ms, these algorithms mute their outputs because concealing such losses is not possible.

An alternative that offers robustness to both short and long losses is multiple-description coding (MDC). MDC involves the source coder, the channel, and the source decoder. An MDC encoder forms multiple descriptions of a signal for transmission over multiple physical or virtual channels. If all descriptions arrive at the decoder, it will produce a high-quality reconstruction of the original signal. If any descriptions are lost in transmission due to poor channel conditions, a lower-quality reconstruction is still possible. The descriptions are usually partially redundant. In general, increasing this redundancy adds robustness but also increases the total transmitted bit-rate.

The theory of MDC was first described in [1], [2]. An application-driven tutorial is given in [3]. Analysis of actual Internet connections shows that when delays must be minimized (as in Internet telephony) MDC generally performs better than forward error correction [4]. Numerous applications of MDC to speech and audio transmission have been proposed and some examples can be found in [5]-[10].

Some multiple-description speech coders are designed from the ground up. Others extend existing standardized single-description (SD) speech coders for use over multiple channels. This approach adds additional robustness to and extends the applicability of SD speech coders that are already well-known, thoroughly tested, and widely deployed.

Such extensions have been proposed for FS-1016, MELP, G.723.1 [5], G.711 [6], G.729 [8], and AMR-WB [8], [7], [9]. The common thread throughout this work is the insightful division of the standardized (or slightly modified) encoder output into $m$ descriptions so that when taken together, those descriptions produce speech with a quality near that of the standardized coder, and when fewer than $m$ descriptions are available, they can still be decoded to produce somewhat lower-quality speech.

Three desirable properties for an MDC extension are evident. First, when all $m$ descriptions are received, it is desired that the output of the extended MDC decoder have speech quality equivalent to that of the original SD decoder. If this is attained, then when channel conditions are good, there is *no quality penalty* for having implemented MDC in place of SD coding. Second, it is desired that the total (across all $m$ channels) transmitted speech payload bit-rate be minimized—ideally it would be no greater than the rate of the original SD encoder. If this is attained, then there is *no speech payload rate penalty* for having implemented MDC. Finally, it is desired that when fewer than $m$ descriptions are received, the resulting speech quality be as high as possible. This maximizes robustness of speech quality to poor channel conditions.

This paper presents a new two-channel speech-polarity decomposition MDC (SPD-MDC) extension to the international standard for pulse code modulation (PCM) speech coding, ITU-T Recommendation G.711 [11]. The MDC extension is inserted between the G.711 encoder and decoder as shown in Fig. 1. The G.711 encoder maps each input speech sample $s(i)$ to a corresponding channel code $c(i)$. The SPD-MDC encoder processes groups of $M$ channel codes to produce two speech payload vectors, $\mathcal{N}$ and $\mathcal{P}$, that are transmitted on two separate channels. Depending on how many channels are working at a given instant, the SPD-MDC decoder may receive $\mathcal{N}$ or $\mathcal{P}$ or both or neither. In any of these cases, the decoder takes the appropriate action to produce a group of $M$ received channel

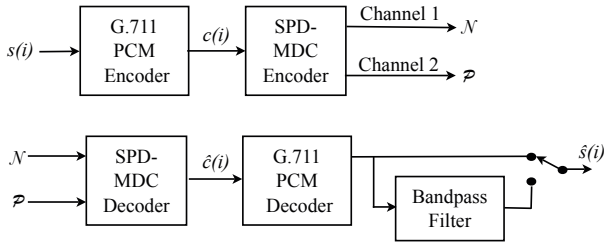


Fig. 1. A two-channel MDC extension to G.711 using a speech-polarity decomposition.

codes, $\hat{c}(i)$. The G.711 decoder translates each channel code $\hat{c}(i)$ to an output speech sample $\hat{s}(i)$, and a bandpass filter enhances speech quality when only one channel is working.

When both channels are working, SPD-MDC G.711 ("SPD-MDC" for short) becomes transparent ($\hat{c}(i) = c(i)$ for all $i$) and exact G.711 speech output is produced—there is *no quality penalty*. In addition, the average total transmitted speech payload bit-rate in SPD-MDC is no greater than the 64 kbps rate of conventional SD G.711 ("SD" for short)—there is *no rate penalty*. We are not aware of any other viable MDC extension to an existing speech coder that simultaneously meets these two opposing goals of *no quality penalty* and *no rate penalty*. The work in [5] has no rate penalty, but achieves this by doubling speech coding frame sizes thus incurring quality penalties. The results in [8], [7], [9] have no quality penalties, but the transmission of redundant information leads to rate penalties ranging from 9% to 20%. Reference [6] offers a family of designs that include one with a quality penalty and no rate penalty, one with a rate penalty and no quality penalty, and several intermediate cases that allow one to trade-off the two penalties.

Of course there are many ways to divide any bitstream into $m$ individual streams with no rate penalty and when all streams are received, there is no quality penalty. But these do not automatically lead to *viable* MDC speech systems because the individual bitstreams will not necessarily produce speech, let alone speech of usable quality.

SPD-MDC uses a sign-magnitude decomposition. Sign information is sent in both descriptions, but magnitude information is sent in only one description and this is determined by the associated sign. The sign information is the only redundancy between the two descriptions and this small redundancy would cause a small bit-rate penalty, but lossless compression is used to eliminate this penalty. Speech signals alternate in polarity so when only one description is received, time intervals of missing magnitudes and received magnitudes alternate and this is leads to a simple yet effective strategy for replacing the missing magnitudes. In Section II, we present the SPD-MDC system in full detail, and in Section III, we characterize its performance in terms of bit-rate and speech quality.

## II. Speech-Polarity Decomposition Multiple-Description Speech Coding

We view the G.711 encoder as a non-uniform quantizer that maps speech samples to channel codes. For each input speech sample, the G.711 encoder produces a channel code that tells what quantization cell the sample falls in. The quantization cell widths grow with speech sample magnitude resulting in a signal-to-quantization-noise ratio that is largely independent of speech level. This is a simple but effective form of perceptual coding. The sample rate is 8000 smp/s and an 8-bit channel code is produced for each sample. The resulting bit-rate is 64 kbps.

G.711 includes two variants called A-law and $\mu$-law and our work supports either of these. A-law divides the negative and positive sample amplitude ranges into 128 quantization cells, resulting in 128 negative channel codes and 128 positive channel codes. The $\mu$-law option has a quantization cell centered at zero. The remaining negative and positive speech amplitude ranges are each divided into 127 quantization cells and thus produce 127 channel codes each. In this work, each channel code must have a sign, and we have assigned the negative sign to the code for the quantization cell centered at zero. Thus, our $\mu$-law work has 128 negative channel codes and 127 positive channel codes.

In the following description of SPD-MDC, we assume packet-based or frame-based operation without loss of generality. The description works for any packet size $M$ that is a multiple of 8 samples (1 ms). This restriction is not critical to the algorithm and can be removed if desired.

### A. Encoding

The functionality of the SPD-MDC encoder is straightforward. The motivation behind the decomposition becomes apparent when the decoder is described. SPD encoding hinges on a sign-magnitude representation of channel codes and classification of the magnitudes according to their signs. Given $M$ channel codes, the encoder forms two payload vectors $\mathcal{N}$ and $\mathcal{P}$. Each payload vector starts with $M$ sign bits, taken in order from each of the $M$ channel codes. This is the only redundancy between the two payloads. Following these sign bits $\mathcal{N}$ also receives the magnitudes of all negative channel codes. These range from 1 to 128 and thus can be directly coded with 7 bits each. Likewise, $\mathcal{P}$ receives the magnitudes of all positive codes, again using 7 bits each. $\mathcal{N}$ is sent on one channel and $\mathcal{P}$ is sent on a second channel. Because channel code polarity is directly related to speech sample polarity, we call this a speech-polarity decomposition. An example for the case $M = 8$ is given in Table I.

If $M_p$ of the channel codes are positive ($1 \leq M_p \leq M$), then $\mathcal{N}$ will contain $M + 7(M - M_p)$ bits and $\mathcal{P}$ will contain $M + 7M_p$ bits. The total number of bits transmitted will be

$$M + 7(M - M_p) + M + 7M_p = 9M \text{ bits} \qquad (1)$$

and the total bit-rate will be 9 b/smp. This is an increase from the original rate of 8 b/smp because sign bits have been

TABLE I
EXAMPLE SPEECH-POLARITY DECOMPOSITION OF $M = 8$ CHANNEL CODES.

| G.711 Channel Codes | $-3, +4, +28, +7, -11, +32, +67, -7$ |
|---|---|
| Contents of $\mathcal{N}$ | $-, +, +, +, -, +, +, -, 3, 11, 7$ |
| Contents of $\mathcal{P}$ | $-, +, +, +, -, +, +, -, 4, 28, 7, 32, 67$ |

transmitted twice. This redundancy increases the bit-rate when direct coding is used.

We use lossless coding to reduce the nominal average bit-rate from 9 b/smp to 7.5 b/smp. Because the probability mass function associated with the 128 or 127 magnitude values is non-uniform (smaller magnitudes are much more likely than larger ones) the entropy of the magnitudes is less than 7 b/smp. We have developed a Huffman code that results in an average magnitude coding rate near 6 b/smp.

In addition, speech signals often run for a full millisecond without crossing zero, so sign bits often show short runs of the same value. This motivates block coding the sign bits in blocks of 8. Each block of eight sign bits can be interpreted as 1 of 256 different symbols. Symbols associated with runs of a single sign are more common than other symbols. Entropy coding exploits this fact and our Huffman code for blocks of eight sign bits has an average rate near 6 b/block, which is $6/8 = 0.75$ b/sign bit. Larger block sizes give marginally lower average rates, but they require exponentially larger codebooks.

Because the sign bits are redundant, this rate reduction shows up twice in the total bit calculation:

$$0.75M + 6(M - M_p) + 0.75M + 6M_p = 7.5M \text{ bits} \quad (2)$$

or 7.5 b/smp. Using this coding scheme SPD-MDC has no rate penalty when compared to SD. In fact, SPD-MDC provides a small average rate reduction, nominally 0.5 b/smp. More detailed rate results are provided in Section III.

### B. Decoding

*1) Two Channels Working:* When both channels are working, $\mathcal{N}$ and $\mathcal{P}$ are both received. A single bit of header information identifies them properly as $\mathcal{N}$ and $\mathcal{P}$. Then the decoder can *exactly* reconstruct the original channel codes from the received signs, negative magnitudes, and positive magnitudes. When both channels are working, SPD-MDC becomes transparent and exact G.711 speech quality results.

*2) One Channel Working:* The motivation for the SPD comes from the case where one channel has failed to deliver data but the other channel is working. When that happens, the SPD allows for a low-complexity algorithm that fills in the missing channel codes with approximations, thus allowing a reduced-quality rendition of the speech signal. To maximize robustness the decoding algorithm is *stateless*, meaning that each packet is decoded without reference to any other packet.

Because speech signals regularly alternate polarity, the duration of missing contiguous channel codes is intrinsically limited. This is advantageous because shorter losses are easier to conceal than longer losses.

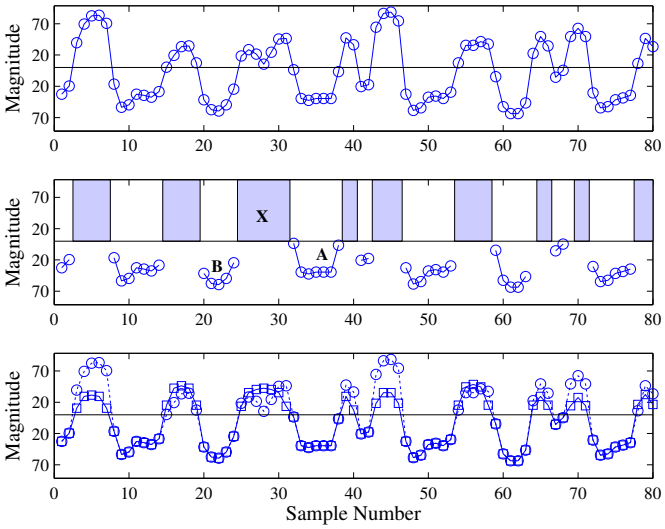


Fig. 2. Example SPD-MDC decoder operation, $\mathcal{N}$ payload received but $\mathcal{P}$ not received. Top: original channel codes. Middle: received channel codes and regions where missing channel codes must reside (shaded). Bottom: decoder output with original channel codes shown with circles and approximated channel codes shown with squares.

The top panel of Fig. 2 shows an example of $M = 80$ transmitted channel codes. Sign-magnitude decomposition is used in this figure. The negative channel codes occupy the lower half of the y-axis and positive channel codes occupy the upper half.

Consider the case where $\mathcal{N}$ is received but $\mathcal{P}$ is not received. The decoder will have sign bits for all $M$ channel codes and magnitudes for only the negative channel codes. The sign bits allow the decoder to assign those negative magnitudes to the proper time locations, as shown in the middle panel of Fig. 2. The sign bits also indicate the time locations of the missing positive magnitudes, these are indicated by the shaded regions in the figure.

Consider a time interval of $L$ missing positive magnitudes, $m(i), i = T+1, T+2, \cdots, T+L$ with $1 \le L \le M$. The leftmost shaded region in the graphical example of Fig. 2 has $L = 5$ and $T = 2$. For each such region of missing positive magnitudes, the decoder will look up or calculate a pulse of length $L$. When $5 \le L$, these pulses are half cycles of a sine wave that has been logarithmically compressed:

$$p(k) = \frac{1}{\alpha} \log_{10}\left(50 \sin\left(\Delta + (\pi - 2\Delta)\frac{k-1}{L-1}\right)\right),$$
$$\alpha = \log_{10}\left(50 \sin\left(\Delta + (\pi - 2\Delta)\frac{\lceil \frac{L}{2} \rceil - 1}{L-1}\right)\right),$$
$$k = 1, 2, \cdots, L, \quad \Delta = \frac{\pi}{45} = 0.070. \quad (3)$$

The amplitude normalization factor, $\alpha$, forces the peak value of this compressed half-cycle to 1.0 for both odd and even values of $L$. Note that the argument to the sine function runs from $\Delta$ to $\pi - \Delta$ thus producing nearly a half-cycle of logarithmically compressed sine wave. When passed through

the G.711 decoding that follows the SPD-MDC decoder, these pulses generate approximate half-cycles of sine waves. This choice produces a smooth, continuous output waveform that sounds better than other choices we explored (including using waveform shape information from neighboring half-cycles of the opposite polarity). In fact, if one averages all length-L half cycles of G.711 channel codes, the resulting, data-driven pulses are remarkably close to those described by (3).

When $1 \leq L \leq 4$, the pulses are given by

$$\begin{aligned} L = 1 \quad &\Rightarrow \quad p(1) = 1, \\ L = 2 \quad &\Rightarrow \quad p(1) = 1,\ p(2) = 0.25, \\ L = 3 \quad &\Rightarrow \quad p(1) = 0.5,\ p(2) = 1,\ p(3) = 0.5, \\ L = 4 \quad &\Rightarrow \quad p(1) = 0.5,\ p(2) = 1,\ p(3) = 1,\ p(4) = 0.5. \end{aligned} \tag{4}$$

We selected these pulse shapes empirically for their ability to approximate a smooth continuous output waveform, and for the sound they produce. They also roughly match shapes found by averaging data. Note that for any value of $L$ the pulse has a peak value of 1.0.

Once the proper pulse has been calculated or looked up, that pulse must be transformed to have the proper amplitude. This transformation is driven by the neighboring half-cycles of the opposite polarity. In the example of Fig. 2, the transformation for the positive pulse that will fill in the area marked $\mathbf{X}$ is based on the negative half-cycles marked $\mathbf{B}$ and $\mathbf{A}$ (short for "before" and "after").

The algorithm collects the magnitudes of the neighboring half-cycles into the sets $\mathbf{B}$ and $\mathbf{A}$ with mean values $\mu_B$ and $\mu_A$ respectively. Due to the stateless decoding requirement, packet boundaries can cause $\mathbf{B}$ or $\mathbf{A}$ or both to be empty sets. Define the gain $G$ as follows:

$$\begin{aligned} \mathbf{B} \neq \emptyset \wedge \mathbf{A} \neq \emptyset \quad &\Rightarrow \quad G = 0.9 \left( \frac{\mu_B + \mu_A}{2} \right), \\ \mathbf{B} \neq \emptyset \wedge \mathbf{A} = \emptyset \quad &\Rightarrow \quad G = 0.9\mu_B, \\ \mathbf{B} = \emptyset \wedge \mathbf{A} \neq \emptyset \quad &\Rightarrow \quad G = 0.9\mu_A, \\ \mathbf{B} = \emptyset \wedge \mathbf{A} = \emptyset \quad &\Rightarrow \quad G = 2. \end{aligned} \tag{5}$$

This gain is then applied to the pulse

$$\tilde{p}(k) = (G-1)p(k) + 1,\ k = 1, 2, \cdots L \tag{6}$$

and the resulting pulse, $\tilde{p}(k)$, provides a way to fill in the $L$ missing positive magnitudes: $m(T+k) = \tilde{p}(k), k = 1, 2, \cdots L$. The result of the process is shown in the bottom panel of Fig. 2 using square markers. The original channel codes are shown with circles. In some cases, the approximation is close, in other cases significant waveform detail is lost.

Note that (5) and (6) serve to match the peak value of the pulse to 90% of the mean of the surrounding magnitudes. We explored numerous amplitude transformations and this one provides the best speech quality. We have determined that bandpass filtering the reconstructed speech (after G.711 decoding) with a passband of 250-3000 Hz generally removes more distortion than speech, and thus improves the resulting speech quality. This filter is the final step in the speech reconstruction process when only one channel is working.

*3) Other Channel Working:* If $\mathcal{P}$ is available to the decoder and $\mathcal{N}$ is not available, then the procedure for filling in the missing negative magnitudes is analogous to the procedure described above. This approach does not acknowledge the asymmetries with respect to polarity that can be readily observed in many speech waveforms. Instead, it forces symmetry on signals that are often slightly asymmetric by nature. Our experiments to detect and preserve simple asymmetry descriptors in SPD-MDC did not yield higher reconstructed speech quality. We conclude that the speech quality of SPD-MDC is not limited by the fact that SPD-MDC forces symmetry but is instead limited by other factors.

*4) No Channels Working:* If both channels fail at the same time, then neither $\mathcal{N}$ nor $\mathcal{P}$ is available to the decoder and we apply the very effective PLC algorithm specified in [12]. Note that if the two channels have independent failure processes and each fail with probability $0 < P_{loss} < 1$, then the probability of both channels failing at the same time is ${P_{loss}}^2$. By spreading the data across two channels, MDC reduces the probability that PLC must be invoked from $P_{loss}$ to ${P_{loss}}^2$.

## III. Performance

### A. Speech Databases

Three speech databases were used in the development and evaluation of SPD-MDC. The *training database* contains recordings of four females and four males reading English language sentences from the Harvard phonetically-balanced sentence lists [13]. This is clean speech at the nominal G.711 input level with total duration of six minutes. It was used to create four variants: speech with levels 10 dB above and below nominal and speech with SNRs of 10 and 20 dB. Thus, the training database includes $6 \times 5 = 30$ minutes of speech covering a wide range of conditions.

The *testing database* contains recordings of five females and five males also reading English language Harvard sentences. The duration of this nominal-level clean speech is about seven minutes.

The *extended testing database* includes the testing database and 4 variants: speech with levels 10 dB above and below nominal, and speech with SNRs of 10 and 20 dB. This is $7 \times 5 = 35$ minutes of speech. It also includes about two minutes each of German, Italian, Japanese, and Portuguese language speech. Each of these languages is represented by two female and two male talkers. Thus the extended testing database is 43 minutes of speech that covers 26 talkers, 5 languages, clean speech and 2 noise levels, and 3 speech levels.

There is no overlap between the training database and the testing databases. In each database the speech recordings have a speech activity factor of 100% as measured by the standardized tool specified in [14]. Background noise types include bus, car, coffee shop, office, party, and street noise. All English speech is bandpass filtered (200 Hz to 3400 Hz) and non-English speech is filtered according to the "Intermediate Reference System" transmit characteristic specified in [14].

### B. Bit-Rate

SPD-MDC offers a fixed total coding rate of 9 b/smp (72 kbps). This is attained by simply mapping and repacking the G.711 channel codes in the SPD-MDC format. This total bit-rate is fixed but the individual rates for the $\mathcal{N}$ and $\mathcal{P}$ payloads will show variation across time since the number of negative and positive samples per unit time is not constant. These payloads always show a mean rate of 4.5 b/smp and in the worst case ($M = 80$), 95% of the payloads have a rate between 2.9 and 6.1 b/smp.

An added layer of variable-rate lossless coding allows SPD-MDC to operate with no rate penalty. Huffman codes were developed using the training database. The average total bit-rates and ranges of variation for the basic testing database are given in Table II. For example, the column headed "100%" shows the full range and the column headed "90%" shows the range excluding the smallest 5% and the largest 5%.

A- and $\mu$-law average bit-rates are 7.4 and 7.7 b/smp respectively. Compared to conventional G.711 (8 b/smp) there is no rate penalty and, in fact, a small rate reduction. Individual rates for the $\mathcal{N}$ and $\mathcal{P}$ payloads show additional variation across time. The case with the highest variation is $M = 80$, where 95% of the payload rates are between 1.7 and 3.9 b/smp.

Telecommunication environments extend beyond the scope of the basic testing database. To include the effects of different speech levels, background noise levels, languages, and filtering, we performed bit-rate measurements using the extended testing database. The resulting average rates were 7.7 and 7.8 b/smp for A- and $\mu$-law respectively. Thus even when considering the wide range of factors included in this extended testing database, there is no rate penalty rate but instead a small rate reduction.

### C. Speech Quality

Speech quality is most directly assessed through subjective listening experiments, commonly resulting in mean opinion score (MOS) values between one and five, with "five" indicating highest quality. An efficient surrogate is the objective speech quality estimation algorithm called PESQ [15] which produces estimated MOS (MOS-LQOn) values.

When both channels are working, SPD-MDC becomes transparent. The speech signal is the same as that of conventional SD G.711 and the estimated MOS value is 4.4. When only one channel is working, the decoder produces speech that is easily intelligible, but of reduced quality. Since PESQ is appropriate for testing clean speech, we processed the clean speech portion of the extended testing database (29 minutes, 26 talkers, 5 languages) with SPD-MDC. The estimated MOS values after averaging across the two cases of "only $\mathcal{N}$ received" and "only $\mathcal{P}$ received" are 2.2 ($M = 80$) and 2.3 ($M = 160$ and 320). A- and $\mu$-law give the same values.

Additional investigations with informal listening and with PESQ allow us to report that the average quality of SPD-MDC when only one channel is working is equivalent to speech processed with the modulated noise reference unit (MNRU) [14] with the SNR parameter Q set between 13 and 14 dB.

TABLE II
TOTAL SPEECH PAYLOAD BIT-RATES FOR SPD-MDC IN B/SMP. FINAL THREE COLUMNS GIVE RANGE OF RATES ASSOCIATED WITH THE SPECIFIED PERCENTAGE OF TOTAL PAYLOADS.

| Case | Mean | 90% | 95% | 100% |
|---|---|---|---|---|
| A-law, $M = 80$ | 7.4 | 4.8 - 7.4 | 4.5 - 7.7 | 3.8 - 10.0 |
| A-law, $M = 160$ | 7.4 | 4.9 - 7.7 | 4.6 - 7.7 | 3.9 - 9.7 |
| A-law, $M = 320$ | 7.4 | 5.0 - 7.7 | 4.7 - 7.7 | 4.0 - 9.5 |
| $\mu$-law, $M = 80$ | 7.7 | 5.6 - 7.7 | 5.0 - 7.8 | 3.8 - 9.8 |
| $\mu$-law, $M = 160$ | 7.7 | 5.7 - 7.8 | 5.1 - 7.8 | 3.8 - 9.6 |
| $\mu$-law, $M = 320$ | 7.7 | 5.9 - 7.8 | 5.3 - 7.8 | 3.8 - 9.4 |

In practice, at any instant the SPD-MDC decoder will have access to $\mathcal{N}$ or $\mathcal{P}$ or both or neither, according to the current state of the two channels. We have simulated several cases using a two-state Markov model. This model is commonly used to model Internet packet losses because it can reproduce actual measured loss patterns, as originally reported in [16].

We use two independent models, one for each channel. The model inputs are the average loss rate $P_{loss}$ and the loss correlation $\rho_{loss}$ (or loss "burstiness" factor.) When $\rho_{loss} = 0$, each packet is lost with probability $P_{loss}$, independent of the fate of the surround packets. As $\rho_{loss}$ increases, the packet losses become more clustered. Fig. 3 shows estimated MOS values (from PESQ) for both SPD-MDC and conventional SD G.711 with the PLC in [12], for the case $\rho_{loss} = 0$ as $P_{loss}$ ranges from 0 to 0.2.

Both options show the expected decreasing trend as $P_{loss}$ increases. As expected, SPD-MDC is more robust to packet loss than SD because the speech signal is spread across two channels. Fig. 3 also indicates that MOS estimates for SPD-MDC increase as the packet size is increased, and we have confirmed this trend (at $P_{loss} = 0.1$) with a small paired-comparison listening experiment.

The fraction of time the SPD-MDC decoder spends in each state ($\mathcal{N}$ received, $\mathcal{P}$ received, both received, neither received) is unchanged by packet size and thus time-averaged speech quality measures will also be unchanged by packet size. Thus the quality vs. packet size relationship in Fig. 3 must be related to the *temporal variation* in speech quality. In other words, we might find that fewer variations per unit time (longer packets) are associated with higher quality while more variations per unit time (shorter packets) are associated with lower quality. As $P_{loss}$ approaches 0.2 this second-order effect diminishes, likely being overcome by the first-order effect of lost packets.

SD shows the opposite (estimated) trend. Here, the likely explanation is that it is harder to generate high-quality concealment of longer losses than shorter losses.

Fig. 4 shows estimated MOS values for SPD-MDC and SD, for the case $\rho_{loss} = 0$ and two cases of bursty packet loss as well. Here again quality decreases as $P_{loss}$ increases and SPD-MDC is more robust to packet loss than SD. Increasing the burstiness of the loss pattern increases the average burst length and this means fewer speech quality transitions per unit time. For SPD-MDC, this reduction in variation increases the

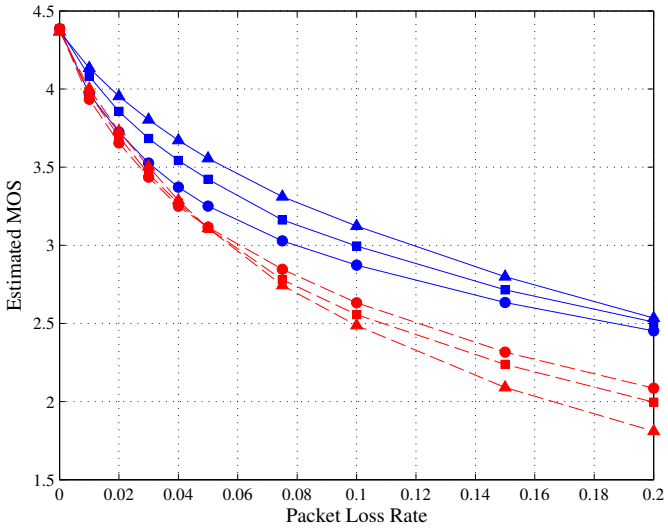


Fig. 3. Estimated MOS for SPD-MDC (solid) and conventional SD G.711 (dashed) speech coding vs. packet loss (independent losses). Circles: 10 ms packet size. Squares: 20 ms. Triangles: 40 ms.

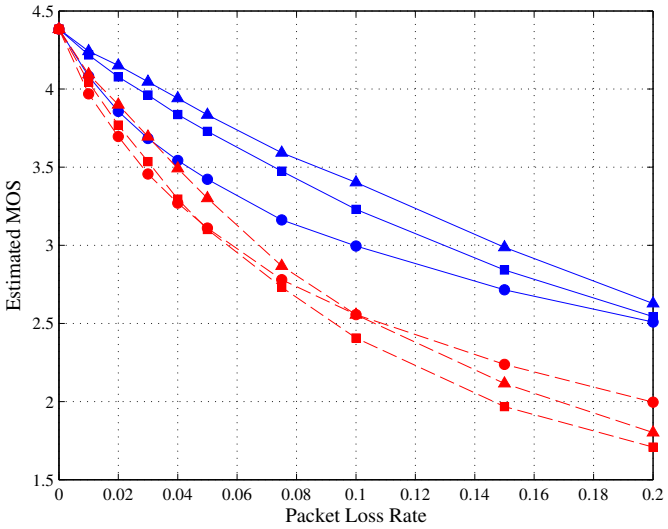


Fig. 4. Estimated MOS for SPD-MDC (solid) and conventional SD G.711 (dashed) speech coding vs. packet loss with 20 ms packets. Circles: independent losses. Squares: average loss burst length is 4 packets. Triangles: average loss burst length is 8 packets.

estimated speech quality and we have confirmed this increase (at $P_{loss} = 0.075$) with a small paired-comparison listening experiment. For SD, the response to this parameter is mixed, perhaps reflecting a trade-off between concealability (favoring shorter, more disbursed losses) and variation (favoring longer, more clustered losses).

## IV. Conclusion

SPD-MDC provides an efficient and effective two-way decomposition of G.711 channel codes. It is efficient from a rate perspective because the redundancy in the decomposition is only 1 b/smp, due to the inclusion of the sign bit of every sample in both descriptions. Lossless coding techniques more than compensate for this redundancy, resulting in a nominal average total speech payload bit-rate near 7.5 b/smp. This is 0.5 b/smp lower than the 8 b/smp rate of conventional single-description G.711 PCM.

SPD-MDC is effective from a quality perspective because it becomes transparent when both channels deliver speech payload, and a simple compressed sine-pulse fill-in technique allows the reconstruction of the original speech with reduced quality when only one channel delivers. Estimated speech-quality values for SPD-MDC improve on those of single-description G.711 with high-quality PLC [12] by up to 0.9 MOS units for the case of 20 ms packet size with bursty packet loss and an average loss rate of 0.15. For independent packet losses, the largest improvement is 0.7 MOS units in the case of 40 ms packets and an average packet loss rate of 0.2.

## References

[1] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, pp. 1909 – 1921, Dec. 1980.

[2] A. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Information Theory*, vol. 28, no. 6, pp. 851 – 857, Nov. 1982.

[3] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74 – 93, Sept. 2001.

[4] M. Y. Kim and W. B. Kleijn, "Comparative rate-distortion performance of multiple description coding for real-time audiovisual communication over the Internet," *IEEE Trans. Communications*, vol. 54, no. 4, pp. 625 – 636, April 2006.

[5] B. W. Wah and D. Lin, "LSP-based multiple-description coding for real-time low bit-rate voice over IP," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 167 – 178, Feb. 2005.

[6] S. Voran, "A multiple-description PCM speech coder using structured dual vector quantizers," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 129 – 132, Montreal, May 2005.

[7] M. Kwong, R. Lefebvre, and S. Cherkaoui, "Multiple description coding for audio transmission using conjugate vector quantization," *$112^{nd}$ AES Convention Paper*, no. 7083, Vienna, May 2007.

[8] J. Balam and J. D. Gibson, "Multiple descriptions and path diversity for voice communications over wireless mesh networks," *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 1073 – 1088, Aug. 2007.

[9] J. Balam and J. D. Gibson, "A transcoding-free multiple description coder for voice over mobile ad-hoc networks," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 3128 – 3132, Las Vegas, April 2008.

[10] J. R. Jensen, M. G. Christensen, M. H. Jensen, S. H. Jensen, and T. Larsen, "Robust parametric audio coding using multiple description coding," *IEEE Signal Processing Letters*, vol. 16, no. 12, pp. 1083 – 1086, Dec. 2009.

[11] ITU-T Rec. G.711, "Pulse code modulation (PCM) of voice frequencies," Geneva, 1988.

[12] ITU-T Rec. G.711, Appendix I, "A high quality low-complexity algorithm for packet loss concealment with G.711," Geneva, 1999.

[13] "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, Sept. 1969.

[14] ITU-T Rec. P.191, "Software tools for speech and audio coding standardization," Geneva, 2005.

[15] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Geneva, 2001.

[16] J. C. Bolot, "Characterizing end-to-end packet delay and loss in the Internet," *J. High-Speed Networks*, vol. 2, no. 3, pp. 305 – 323, Dec. 1993.