
Jun Tang
Twin Sun Inc.
360 N. Sepulveda Blvd. El Segundo. CA90245

Abstract

Recovering missing speech is an important problem in packet voice communications, since missing speech are a major impairment in packet network communications. Speech quality is improved by substituting the gaps with estimates of the transmitted waveform in reconstructed speech by previous methods. We have proposed a DSWS (Double Sides Waveform Substitution) method and a DSPS (Double Sides Periodic Substitution) for recovering missing speech. In this paper, we have investigated all estimation techniques for packets containing 16 ms of speech and taken comparison tests of each technique by a subjective test (opinion test) and an objective test (LPC spectral distortion). These tests indicated that the most effective estimation is the DSPS method, which extends the acceptable ratio of missing packets of 10 percent by previous method, to 20 percent in the experimental results.

1. Introduction

Packet communication techniques were a kind of reservation communication mode, devised for digital data transmission. When these techniques are used in packet voice communications, it leads to the serious impairment of speech quality and the loss of speech information, because of the discarded packets and long delays in the delivery of information [1]. It is very important to recover the missing speech in the improvement of speech quality and the compensation of speech information.

It has been reported that packet delay beyond the receiving terminal, and good speech quality under two percent of missing packet rate can be achieved [2]. Many techniques have been proposed for recovering missing packet in packet voice communications. The simplest technique, packet repetition, uses the most recent received packet as an estimate of the missing one. Another technique, Pattern matching method [3], substitutes the past speech waveform, of the missing interval, that resembles the waveform of the missing interval, for the missing interval. A third approach is to estimate the pitch of the received speech and to replicate prior pitch waveform for the duration of a missing packet [5].

With the recovery of missing speech, we have proposed a DSWS (Double Sides Waveform Substitution) method [6], which is based on the idea that the missing speech waveform has resemblance to the past and the future waveforms of the missing interval, and uses the past and the future speech waveforms that resemble the speech waveform of the missing interval to substitute for the missing interval. But when the duration of missing speech is very long, it will lead to a loss in resemblance of the substituted waveform. So we only use the past and the future waveforms that have strong resemblances to the waveform of the missing interval, to substitute for the missing interval iteratively. This method is called double sided periodic substitution (DSPS) method [7].

In packet voice communications, without any signal processing, the acceptable missing packet rate is about 2 percent, with pattern matching technique, up to 5 percent. Pitch waveform replication technique extends this limit to 10 percent [1].

The purpose of this paper is to evaluate the effect of our DSPS method and investigate the maximum missing packet rate by our DSPS method, compared to other techniques, by a objective test (LPC spectral distortion) and a subjective test (Opinion Test), and we also evaluate the compensation of speech information by DSPS method.

2. Experimental Conditions

The duration range of packet is about 8-32 ms in packet voice communications. In our experiment, the packet duration was 16 ms, each packet containing 128 consecutive samples (8 kHz sampling rate). We performed all the signal processing on a mini-super computer (mv 6000) with 14 bit analog-to-digital and digital-to-analog converters. Thus, the PCM quantizing noise was disregard and the audible speech impairments were those due to missing packets.
In our experiments, the missing packets were distributed randomly over the duration of the speech. So the packet losses are bursty, there can be long periods with no missing packets, interspersed with high probabilities of packet loss. The duration of the missing interval is not always 16 ms. It may be 32 ms, 48 ms, 64 ms or more. The speech quality in these cases will be quite impaired.

In signal processing, five methods were used to recover missing packets. The simplest treatment of missing packet is packet repetition in which the receiving terminal stores the contents of the most recently received packet. This packet is substituted for the missing packet.

A. DSPS Method

The most recent waveforms preceding and following the missing interval, that have strong resemblances to missing speech, are used to substitute for the missing interval iteratively. The substitution waveform consists successively of the same waveform.

$X_0$ and $X_m$ are the vectors of preceding and following waveforms of missing interval with a one pitch length as shown in Fig.1. The waveform of missing interval $X_i$ is calculated in DSPS as follows:

$$r(i) = \frac{1}{\delta_x \delta_y N} \sum_{j=0}^{N-1} x(j)y(j)-\bar{x}\bar{y}$$

where

- $X_k(n) = w(n)x_0(n) + (1-w(n))x_m(n)$
- $w(n)$: triangular window
- $w(n) = \frac{l-n}{l-1}$: window length
- $k\cdot l = \frac{5}{8}m$
- $m$: the length of missing interval

$\delta(i)$ represents the power parameter such as standard deviation, maximum value and minimum value of waveform. By the experimental results, the standard deviation is the most effective to represent the power of speech waveform.

$$\delta_x = \sqrt{\frac{1}{N} \sum_{j=0}^{N-1} (x(j)-\bar{x})^2}$$

$$\delta_y = \sqrt{\frac{1}{N} \sum_{j=0}^{N-1} (y(j)-\bar{y})^2}$$

$\bar{x}$ and $\bar{y}$ are the mean value of waveform $\delta_x$ and $\delta_y$ are the standard deviations of waveforms.

3. Signal Processing Examples

The original speech waveform, missing speech waveform, reconstructed speech waveform by DSWS method and by DSPS method are showed in Fig.2 respectively. This figure indicated that the reconstructed speech waveform by DSWS and DSPS are very similar to the original speech waveform, so DSWS and DSPS can be expected to be the very effective methods for recovering missing speech.

Fig.1 DSPS Algorithm

$$X_0 = \begin{bmatrix} x_0(1) \\ x_0(2) \\ \vdots \\ x_0(l) \end{bmatrix}, \quad X_i = \begin{bmatrix} x_i(1) \\ x_i(2) \\ \vdots \\ x_i(l) \end{bmatrix}$$

$$X_k = \begin{cases} \delta(i)X_0 & \text{if } i<l \\ \delta(i)X_k & \text{if } i=l \\ \delta(i)X_m & \text{if } i>l \end{cases}$$

$$\delta(i) = \delta_0 + \frac{\delta_m - \delta_0}{l}i$$

Fig.2 Speech Waveform Samples (a)Original (b) Missing (c) DSWS (d) DSPS
4. Experimental Results

In order to evaluate the recovered speech quality, we performed an objective test and a subjective test.

A. Objective Test

We measure the LPC spectral envelope difference of the original speech with the processed speech and calculate the LPC spectral distortion.

In this experiment, LPC spectral distortion is defined as the difference of spectral envelope of the original speech with the processed speech, extracted from the speech signal by using linear prediction analysis. The LPC spectral distortion is shown as follows.

\[ D = \frac{1}{F} \int_{T}^{F} (P - P') \, dt \]

where \( P \) and \( P' \) are spectral envelopes of the original speech and processed speech respectively, \( F \) is the upper limit frequency, \( T \) is the articulated duration. \( D \) is the LPC spectral distortion.

In the analysis of LPC, analysis order is 12, frame length and and frame shift are 32 ms and 6.4 ms respectively.

Table 1. LPC Spectral Distortion (dB)

<table>
<thead>
<tr>
<th>Rate</th>
<th>DSPS</th>
<th>DSWS</th>
<th>Pitch</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>0.37</td>
<td>0.38</td>
<td>0.59</td>
<td>0.46</td>
</tr>
<tr>
<td>8%</td>
<td>0.70</td>
<td>0.85</td>
<td>0.90</td>
<td>1.10</td>
</tr>
<tr>
<td>12%</td>
<td>0.83</td>
<td>1.11</td>
<td>1.03</td>
<td>1.21</td>
</tr>
<tr>
<td>16%</td>
<td>0.85</td>
<td>1.13</td>
<td>1.08</td>
<td>1.25</td>
</tr>
<tr>
<td>20%</td>
<td>1.02</td>
<td>1.33</td>
<td>1.43</td>
<td>1.61</td>
</tr>
<tr>
<td>24%</td>
<td>1.15</td>
<td>1.54</td>
<td>1.70</td>
<td>1.72</td>
</tr>
</tbody>
</table>

Table 1 indicates the LPC spectral distortion of the processed speech by each technique. The minimum spectral distortion is achieved by the DSPS method for all missing packet ratios, and the spectral distortion with a missing packet ratio of 20% are about 0.92 dB. The LPC spectral distortion for missing packet rate of 10% achieved by Pitch Waveform Replication is about 0.9 dB, and for 20% by DSPS is about 0.9 dB. So the processed speech quality for a missing packet ratio of 20% by DSPS is nearly equal to the speech quality in 10% by Pitch Waveform Replication confirming to the objective test results. It also means that our DSPS method extends the limit of acceptable missing packet rate from 10% by Pitch Waveform Replication to 20%.

B. Subjective Test

In this experiment, opinion test is used. Opinion test measures the degree whether the processed speech sounds satisfactory or not, directly for processed speech by all methods. Listeners do the absolute judgement by using a 5 category evaluation such as excellent, good, fair, bad, and unsatisfactory. Mean opinion score is calculated by averaging over scores for each method.

The experiment was carried out in a room with acoustic absorbing material on the walls, the total of subjects were 21. The speech data were presented randomly.

Table 2. Mean Opinion Score (MOS)

<table>
<thead>
<tr>
<th>Rate</th>
<th>DSPS</th>
<th>DSWS</th>
<th>Pitch</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>4.38</td>
<td>4.19</td>
<td>4.00</td>
<td>4.19</td>
</tr>
<tr>
<td>8%</td>
<td>4.00</td>
<td>3.33</td>
<td>3.62</td>
<td>3.23</td>
</tr>
<tr>
<td>12%</td>
<td>3.85</td>
<td>3.19</td>
<td>3.20</td>
<td>2.57</td>
</tr>
<tr>
<td>16%</td>
<td>3.62</td>
<td>3.00</td>
<td>2.57</td>
<td>2.38</td>
</tr>
<tr>
<td>20%</td>
<td>3.05</td>
<td>2.57</td>
<td>2.29</td>
<td>1.95</td>
</tr>
<tr>
<td>24%</td>
<td>2.95</td>
<td>2.38</td>
<td>1.81</td>
<td>1.52</td>
</tr>
</tbody>
</table>

Table 2 shows Mean Opinion Score (MOS) as a function of missing packet percentage for the five replacement schemes. The best speech quality is achieved by DSPS. In determining the acceptable missing packet rate, assume, for example, that the criterion of acceptability is a mean opinion score (MOS) of 3.5. Then, with Pitch Waveform Replication, the acceptable missing packet rate is about 7%, with DSWS, up to 8%. With DSPS, the acceptable missing packet rate is about 16%. It has been reported, however, that the limit of acceptable missing packet rate is about 14% with DSWS. DSPS extends this limit to 20%.

C. Speech Information Compensation

We also did the experiments in speech recognition to check the loss of information of missing speech
and the compensation of speech information after recovery.

The basic database used in the experiments consists of two sets of 550 Japanese city names spoken by five male speakers. The first set is used as the reference pattern and the second set is used as the test pattern. The analysis condition of speech data sets are as follows:

- Sampling frequency: 10 KHz
- Sampling accuracy: 12 bits
- Cut-off frequency of LPF: 4.5 KHz
- Window type: Hamming window
- Frame period: 10 ms
- Order of LPC analysis: 10

LPC cepstral distance measure is used in recognition experiment. The maximum cepstral number used in speech recognition is 30.

![Fig.3 Results of Speech Recognition](image)

In order to check the speech information compensation after recovery by DSPS, we take 4 speech recognition tests, of original speech recognition, missing speech recognition, reconstructed speech recognition by DSWS and by DSPS respectively. By comparing the recognition rate of original speech, missing speech and reconstructed speech, the effect of DSPS for speech information compensation can be obtained.

In Fig.3, the recognition rate for missing speech decreased by 20%-25% when the missing rate 16%-20%. The recognition rate for recovered speech by DSPS method is the same as the recognition rate for original speech. The recognition rate increased by 25% after recovery by DSPS method. We obtained very good results in compensation of speech information after recovery by DSPS method.

5. Conclusions

In order to evaluate the effect of DSPS method on the quality of packet voice communications, we performed an objective test and a subjective test. By the experiments, we also can evaluate the relative merits of each technique for reproducing speech signals with missing packets. We conclude as follows:

1. In the objective test, a very low spectral distortion was achieved by DSPS compared to other techniques. Spectral distortion for a missing packet ratio of 20% by DSPS is the same as that for 10% by Pitch Waveform Replication.

2. In the subjective test, if the criterion of mean opinion score (MOS) of 3.5 is used, with Pitch Waveform Replication, the acceptable missing packet rate is 7%, with DSWS, up to 8%, with DSPS, the acceptable missing packet rate is about 16%. On the other sides, if the limiting criterion of an acceptable missing packet rate of 10% as in [1] is used, with DSWS, the acceptable missing packet rate will be 14%, and with DSPS, up to 20%.

3. Regarding complexity of signal processing, the simplest technique is Packet Repetition, then Pattern Matching followed by Pitch Waveform Replication, followed by DSWS, approximately equal to DSPS. The speech quality increases with the complexity of signal processing, as expected.

4. Speech recognition rate increased by 25% after recovery by DSPS method, we also achieved a good results in compensation of speech information.

In conclusion, the most effective technique for recovering missing packets in packet voice communications is the DSPS method. The DSPS method extends the limit of acceptable missing packet rate from 10% by previous techniques to 20%.
References

(2) Iosh etc.: "Study of acceptable limit of transmission delay on two direction communication system", Research Conference of Hearing Conceptual Field (in Japanese), H-84-57(1984-12)