

Speech Signal Recovery in Communication Networks

V. Zagursky, A. Riekstinsh

Interpolation approaches to the shape recovery of a speech signal in transmission over packet switched communications networks are proposed. The samples of signal fragments are mixed and transmitted in correspondence with standard procedure for packet-switched transmission. After reception a reverse permutation is made. In the case of packet losses missing samples are separated by several samples of the source signal. Correlation properties of the signal are used for the recovery samples due to first- and second-order non-adaptive and adaptive interpolation. For the loss of 25 % packets and second order adaptive interpolation a 2–4 % error distribution range has been achieved.

1 Introduction

In packet speech transmission over a network, part of the information is lost. If we are to preserve an acceptable quality of speech, the permissible percentage of losses is limited. The authors of [1] and other papers admit 1–3 % percentage loss. As a result, the average network load is limited. The authors of [2], [3] propose the use of a variable speech packet-encoding rate to enable smoothing of the effect of network overloads on the received speech quality. Paper [4] proposes the classification of speech segments in accordance with their structure. Packets belonging to different classes are assigned different priorities of delivery. When network overloads appear, packets with lower priority are discarded first. At the receiving end, regeneration of lost packets is performed.

The articles mentioned above deal mainly with information transfer based on the encoding of speech signal parameters. This paper deals with transfer based on waveform encoding using pulse-code modulation.

Section 2 considers the main principle of loss recovery in waveform encoding. Section 3 analyses first-order interpolation methods. Section 4 analyses the possibilities of second-order interpolation. Section 5 discusses some experimental results on the recovery of phrases and suggest some future investigations.

2 Basic interpolation principle

For channels with error bursts, the sequence is often subjected to an alternation prior to transmission and is recovered at the receiving end. In this case the errors are distributed in a more uniform manner [5]. We will use this principle for packet speech transmission. The source sequence of the samples of a signal segment is memorized, and some permutation of the samples is made, which is followed by separation into packets, and then transfer. At the receiving end a reverse procedure is performed. If a packet is lost during transmission, the lost samples are separated by one or several samples of the source signal. Such a procedure enables the recovery of losses as a result of correlation.

Interpolation of samples was applied in [6]. A signal is divided into two groups of even and odd samples. Each group is shaped into a packet, following reverse permutation, and in the case of the loss of one packet, one source sample appears between the missing samples. This allows the applicability

extrapolation and first-order interpolation. The “Odd-even” alternation allows only the correlation of the neighbouring samples to be used for the recovery. Using the information on a greater number of samples, the value of the lost samples can be recovered more accurately. To do this, the samples of the source signal have to be interchanged on a segment containing more than two packets. For example, using a block encoder, the sequence of samples is written into an $n \times m$ matrix column-wise and is read row-wise. Having defined the length of row n as being equal to the packet length, after the reading we will obtain m packets. If one of the m packets is lost, then after the reverse permutation the lost samples will be separated by $m-1$ samples of the source signal. In this case interpolation procedures ranging from the zero order and up to the $(m-1)$ -th order may be applied for the recovery.

3 First-order interpolation

We will assume that not more than one packet is randomly lost from the sequence of signal samples consisting of packets. The sequence X_i ($i = 1, 2, \dots$) of the centered signal $X(t)$ is transmitted. The reception and the reverse permutation follows. In the case of the loss of one packet, the received signal $\bar{X}_i$ differs from X_i only in the points $t_i = t_j$, where the samples are missing. In this case the probability is $P(t_i = t_j) = 1/m$. The transmission error is $\varepsilon_i = X_i - \bar{X}_i$. The mean square error of the transmission of a sequence consisting of m packets is

$$E(\varepsilon_i^2) = \sum_i (X_i - \bar{X}_i)^2 P(t_i = t_j) = E \left[(X_i - \bar{X}_i)^2 \right] / m \quad (1)$$

If the estimate $\bar{X}_i$ of the values of the missing samples is equated to zero, then:

$$E(\varepsilon_0^2) = E(X_i^2) / m = R_{XX}(0) / m, \quad (2)$$

where $R_{XX}(0)$ is the value of the correlation function for a zero shift.

In the recovery of the values of $\bar{X}_i$ on the basis of the first-order prediction $\bar{X}_i = \alpha \cdot X_{i-1}$. It is usually assumed that coefficient $\alpha = 1$. Then, in accordance with (1), the mean square error will be:

$$E(\varepsilon_{p1}^2) = E \left[(X_i - X_{i-1})^2 \right] / m = 2 [R_{XX}(0) - R_{XX}(1)] / m, \quad (3)$$

where $R_{XX}(1)$ is the value of the correlation function of signal $X(t)$ with the shift equal to Δt .

For the first-order interpolation:

$$\bar{X}_i = a_{-1} \cdot X_{i-1} + a_{+1} \cdot X_{i+1}. \quad (4)$$

For the first-order interpolation it is usually assumed that $a_{-1} = a_{+1} = a$. For the most commonly used procedure $a = 0.5$ (linear interpolation). This procedure was referred to in [6] as non-adaptive, as well as the prediction with $\alpha = 1$. For $\bar{X}_i = 0.5(X_{i-1} + X_{i+1})$ the mean square error will be:

$$E(\varepsilon_{i1}^2) = [1.5R_{XX}(0) - 2R_{XX}(1) + 0.5R_{XX}(2)]/m. \quad (5)$$

Let us compare expressions (3) and (5):

$$E(\varepsilon_{r1}^2) - E(\varepsilon_{i1}^2) = [R_{XX}(0) - R_{XX}(2)]/m. \quad (6)$$

Expression (6) yields that interpolation provides a smaller recovery error, as compared with prediction, since for any random signals $R_{XX}(2) < R_{XX}(0)$.

The error can be minimized by applying the adaptive approach, as was demonstrated in [6], by calculating the error for interpolation with the use of (4) and by determining the minimal error depending on the coefficients a_{-1} and a_{+1} , we will obtain:

$$a_{\text{opt}} = a_{-1} = a_{+1} = r_{XX}(1) \cdot [1 + r_{XX}(2)]^{-1}, \quad (7)$$

where $r_{XX}(n) = R_{XX}(n)/R_{XX}(0)$ is the correlation coefficient.

The figure 1 shows the charts of the errors of recovering a signal which corresponds to the sounds *a*, *d*, *c*, *sh*. The vertical

axis are laid off, in percent, the value of the normalized mean square errors $\gamma = E(\varepsilon^2)/R_{XX}(0)\gamma$.

The horizontal axis displays the ordinal numbers of the appropriate signal segment with length $n = 128$ samples. A real speech signal was used in the experiments.

A situation was simulated in which every fourth packet is lost ($m = 4$). When no recovery is applied, the error at the receiving end equals 25 %, since from (2) we have $\gamma = 1/m$. It follows from the figure that for practically all signal segments $\gamma_{1.2} < \gamma_{1.1}$.

4 Second-order interpolation

The estimate of the values of the lost samples $\bar{X}_i$ using the known samples with numbers $i \pm 1$, $i \pm 2$ will be made using the expression:

$$\bar{X}_i = b_1(X_{i-1} + X_{i+1}) + b_2(X_{i-2} + X_{i+2}) \quad (8)$$

The normalized mean-square error of the second-order interpolation will be determined in a way similar to (3).

$$\gamma_2 = \left[(1 + 2b_1^2 + 2b_2^2) - 4b_2(1 - b_2)r_{XX}(1) + 2(b_1^2 - 2b_2^2)r_{XX}(2) + 4b_1b_2r_{XX}(3) + 2b_2^2r_{XX}(4) \right] / m. \quad (9)$$

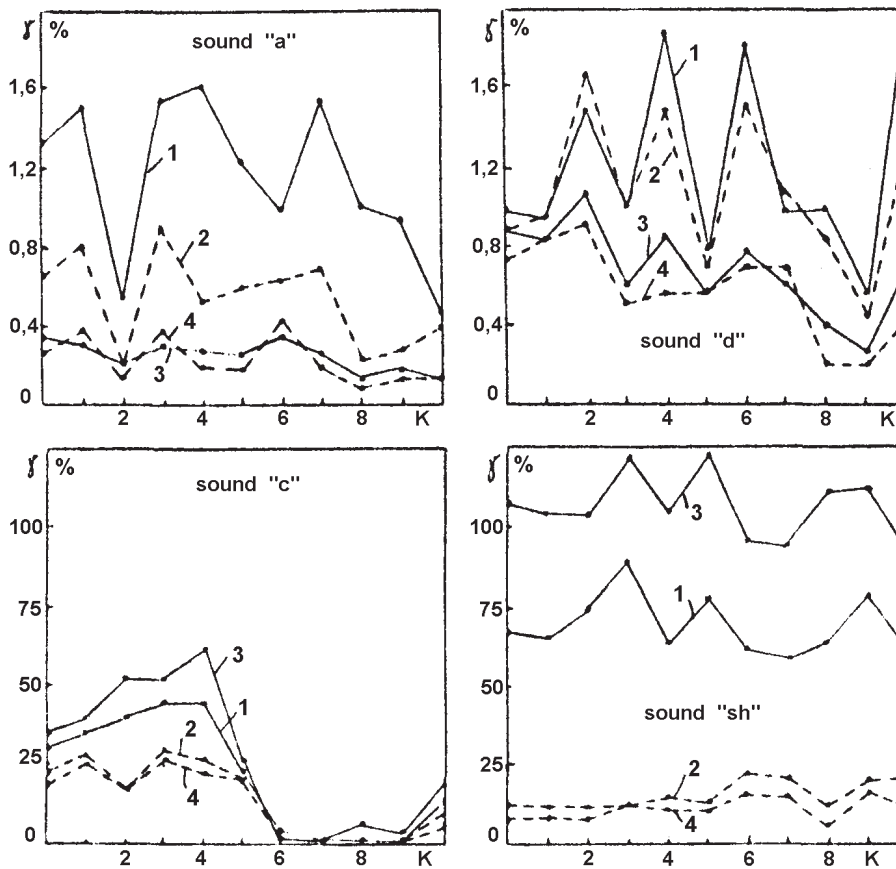


Fig. 1: Errors of recovering a signal which corresponds to the sounds *a*, *d*, *c*, *sh*

- 1 – linear interpolation ($\gamma_{1.1}$)
- 2 – 1-st order adaptive interpolation ($\gamma_{1.2}$)
- 3 – 2-nd order Chebyshev's interpolation ($\gamma_{2.1}$)
- 4 – 2-nd order adaptive interpolation ($\gamma_{2.2}$)

Like for the first-order interpolation, the procedure for determining coefficients b_1 , b_2 may be non-adaptive or adaptive. For a non-adaptive procedure some known family of polynomials can be used, for example, the family of second-order Chebyshev polynomials. It can be shown that $b_1 = 0.667$ and $b_2 = -0.167$. Then (9) is converted to the following form:

$$\gamma_{2,1} = [1.94 - 3.1 r_{XX}(1) + 1.56 r_{XX}(2) - 0.44 r_{XX}(3) + 0.056 r_{XX}(4)]/m. \quad (10)$$

Let us compare (10) and (5) for the normalized error $\gamma_{1,1} = E(\varepsilon_{1,1}^2)/R_{XX}(0)$:

$$\gamma_{1,1} - \gamma_{2,1} \approx [(r_{XX}(1) - 0.44) - (r_{XX}(2) - 0.44 r_{XX}(3))]/m \quad (11)$$

Analysing (11), it is easy to see that the efficiency of the interpolation is determined by the type of correlation function of the signal. Specially, $\gamma_{2,1} < \gamma_{1,1}$ for the signals for which the correlation between the adjacent samples is high, and afterwards it rapidly decreases. This is true for many speech sounds [7]. However, for hushing sounds and fricatives, the value of $r_{XX}(1)$ may not be high. Then it is possible that $\gamma_{1,2} < \gamma_{1,1}$. In selecting coefficients in (8) the second-order adaptive interpolation allows us to take account of the effect of all values of the correlation function. The formulas for the optimum values of coefficients b_1 and b_2 will be obtained by equating the partial derivatives $\partial \gamma_2 / \partial b_1$ and $\partial \gamma_2 / \partial b_2$ to zero in (9). This is illustrated by Fig. 1, which shows the experimental results of signal recovery for 25 % losses for the first- and second-order non-adaptive and adaptive interpolation. For the sounds "a" and "d", the second-order interpolation provides better results than the first-order interpolation. For the sounds "c" and "sh" this is true for adaptive interpolation, while with the use of the second-order Chebyshev interpolation the recovery error increases.

5 Experimental results and conclusions

Experiments have been made on the recovery of losses in fused speech. Signal samples are divided into packets, 128 readings each. The packets are combined into groups of 4 packets. Within one group permutations are made so as to ensure that all 4 packets are interrelated. Each fourth packet is discarded, following which a reverse permutation and recovery are performed. Use was made of first- and second-order interpolation – non-adaptive and adaptive. The table shows the integral error estimates for all recovery procedures for different speech phrases up to 5 sec. in length.

Our investigations testify to the efficiency of approaches that involve waveform recovery. The first-order adaptive interpolation yielded results, acceptable in terms of sound

Table 1: Fused speech recovery error

Type of interpolation	Error distribution range
Linear interpolation	5–12 %
1-st order adaptive interpolation	3–5 %
2-nd order Chebyshev interpolation	5–15 %
2-nd order adaptive interpolation	2–4 %

quality, for the loss of 25 % packets. The second-order adaptive interpolation yields better results in terms of both sound quality and root-mean-square error.

The adaptive procedure requires additional processing of the signal at the receiving end in order to calculate the correlation instants and coefficients, followed by transmission of the calculated information in the packet. The authors intend to investigate other approaches – determining the relation between the interpolation coefficients and the sign characteristics of the signal, which are easier to determine than the correlation ones, as well as calculating the characteristics at the receiving end directly from the signal with losses.

References

- [1] Chlamtac I.: "An Ethernet compatible protocol for real-time voice/data integration." *Computer Networks and ISDN Syst.*, Vol. **10** (1985), No. 2, p.81–96.
- [2] Bially T., Crold B., Seneff S.: "A Technique for Adaptive Voice Flow Control in Integrated Packet Networks." *IEEE Trans. on Commun.*, Vol. **COM-28** (March 1980), No. 3, p. 325–333.
- [3] Forat V. S., Friedman E. M., Minden G. I.: "Multi-rate voice coding for load control on CSMA/CD local computer networks." *Computer Networks and ISDN Syst.*, Vol. **11** (1986), p. 99–110.
- [4] Petr P. W., DaSilva I. A., Forat V. S.: "Priority Discarding of Speech in Integrated Packet Networks." *IEEE Journal on Selected Areas in Commun.*, Vol. 7 (June 1989), No.5, p. 644–656.
- [5] Clark G. C., Cain J. B.: "Error-Correction Coding for Digital Communications." New York: Plenum Press, 1982, p. 352.
- [6] van den Bos A.: "Complex Electron Wave Reconstruction Using Parametr Estimation." *IEEE Transact. on Instrum. and Measurement*, Vol. **46**, No. 4, p. 826–830.
- [7] Rabiner L. R., Schafer R. W.: "Digital processing of speech signals." New Jersey 07632: Prentice-Hall, 1978, p. 436.

V. Zagursky
phone: +371 755 8448
fax: +371 755 5337
e-mail: zagursky@edi.lv, aigars@egle.cs.rtu.lv

Institute of Electronics and Computer Science
of Latvian University

A. Riekstinsh

Institute of Automation and Computer Engineering
of Riga Technical University

14 Dzerbenes str., Riga, LV-1006, Latvia