Information hiding in G.711 encoded speech, based on substitution of LSBs and estimation of tolerable distortion

Presented by
Karthik Ram

Outline

- Introduction
- LSB substitution for G.711 coded speech
- Aoki's method for selective embedding
  - improved quality of embedded speech by reducing distortion.
- Novel data-hiding based on estimation of tolerable distortion
  - Improving subjective quality by determining the number of bits for embedding.
- Comparison
- Conclusion

Introduction

- Proposed by Akinori Ito, *Shun'ichiro Abe, Yˆoiti Suzuki.*
- Enhancing the conventional LSB substitution technique.
- A low-bitrate encoder, G.726 ADPCM, is used as a reference for deciding how many bits can be embedded in a sample in order to maintain acceptable speech quality.

The big picture !

- VOIP – Voice over Internet protocol
  - Free calls or communication to any place irrespective of distance.
  - Replace traditional phone system.
  - Uses G.711 codec (simplest and most common codec).
  - Various attempts to optimize this codec.
    - ✓ Packet loss concealment
    - ✓ Bandwidth expansion.

Problems with Optimization :

- Requires side information to be added to the speech data.
- Ensure that the encoded data is downward compatible with ordinary speech data.

- Solution - Data hiding.

LSB substitution technique :

- Simplest method for embedding information into speech data.
- *x1, . . . , xn be a sequence of scalar values, where xi* is encoded into *M bits.*
- *b1, . . . , bn be the data we want to* embed, where *bi* $\in$ *{0, 1}*
- *m(x) is a bitmask function, clears the LSB of x.*
- LSB of x_i *acts as* payload for embedding data b_i.
- *yi = m(xi) + bi ; bi = yi − m(yi).*
  - *yi -> Embedded speech data.*

Drawbacks in simple LSB:

- G7.11 applies non linear functions such as *μ-law* or A-law.
  - Decoding requires non linear transformation function.
- The distortion for such a nonlinear quantized sample is calculated using :
- $\varepsilon_i 2 = (g(x_i) - g(y_i))2$, When *xi is large, we* obtain larger $\varepsilon_i 2$
  - Distortion depends on the magnitude of the input speech.

Selective embedding by Aoki :

- Embeds a fixed number of information bits in samples that have small magnitudes.
- *Here i1, i2, . . . , in, is a permutation of 1, 2, . . . , n(scalar values).* We define a number *k ≤ n.*
- yi=> Embedded signal

$$y_{i_j} = \begin{cases} m(x_{i_j}) + b_j & \text{if} \quad j \leq k \\ x_{i_j} & \text{otherwise} \end{cases}$$

- Reduces the total distortion of the embedded signal.
- Bitrates of the embedded data can be controlled
- up to 8 kbit/s (i.e. the bitrate of LSB) by changing *k in balance* with the speech quality.

Drawbacks with Aoki's method:

- Reducing distortion does not improve subjective quality.
- Aoki's method (and the original LSB substitution) only uses the least significant one bit for embedding.

Novel data hiding method :

- Based on estimation of tolerable distortion.
- To judge the limit of subjective quality, an encoder which encodes the input speech at less than 64 kbit/s by taking human hearing characteristics into account is used as a reference.
- It is assumed that speech is encoded sample-by-sample, considering the past input.
- A low bit rate encoder maximizes subjective quality.

Procedure :

for $i \leftarrow 1$ **to** n
  for $j \leftarrow 1$ **to** 8
    $e_i \leftarrow j - 1$
    $I_0(i,j) \leftarrow I(g(m_j^-(x_i))|\hat{o}_1, \ldots, \hat{o}_{i-1})$
    $I_1(i,j) \leftarrow I(g(m_j^+(x_i))|\hat{o}_1, \ldots, \hat{o}_{i-1})$
    if $I_0(i,j) \neq I_1(i,j)$ **then**
      break
    end if
  end for
  Embed data into the least e_i bits of x_i
  $\hat{o}_i \leftarrow g(m_{e_i}^-(x_i))$
end for

Fig. 1. Procedure for embedding

Calculation of number of Embeddable bits :

- Let *m_{-j} (x) be a function that clears the least j bits of x, and m_{+j} (x) be a function that sets the least j bits of x.*
- Let *I(oi|o1, . . . , oi−1) be an output of the i-th sample* using a low-bitrate encoder, given the past input *o1, . . . , oi−1.*
- Number of embeddable bits *ei is given as:*

$$\hat{o}_k = g(m^-_{e_k}(x_k))$$

$$I_0(i,j) = I(g(m^-_j(x_i))|\hat{o}_1, \ldots, \hat{o}_{i-1})$$

$$I_1(i,j) = I(g(m^+_j(x_i))|\hat{o}_1, \ldots, \hat{o}_{i-1})$$

=>

$$e_i = \max_j\{j|I_0(i,j) = I_1(i,j)\}.$$

Calculation of Embeddable bits :

- Determination of *ei is not affected by the embedded* content (*ei is determined* using the past signal whose embedded bits are cleared).
- Use G.726 as our low bit rate encoder
  - Determines the code sample-by-sample.
  - Offers several bit rate configurations, which allow us to control the amount of embedded information.
- Let *N be the bit length of a* sample coded by ADPCM. N depends on the bit rate of ADPCM
  - *N = 5 for the 40 kbit/s configuration, N = 4* for 32 kbit/s and *N = 3 for 24 kbit/s.*

Calculation of Embeddable bits :

- We cannot trust *I0(i, j) = I1(i, j) as an index of subjective* quality, so we use the following *ei' instead of ei.*

$$e'_i = \begin{cases} 0 & \text{if} \quad |I_0(i,0)| = 2^{N-1} - 1 \\ e_i & \text{otherwise} \end{cases}$$

- When |*I0(i, 0)*| = 2(*N−1)−1, it means that the distortion between the true sample* and the predicted value is maximum.

Comparison of all methods :

- MOS-LQO (Mean Opinion Score-Listening Quality Objective)

Table 1. Experimental results using 32 kbit/s ADPCM

Method	Bitrate [bit/s]	MOS-LQO (st.dev.)
LSB substitution	8000	4.05 (0.07)
Aoki's method	2000	4.36 (0.04)
	4000	4.32 (0.04)
	5500	4.14 (0.05)
	6000	4.10 (0.06)
Proposed	5585	4.40 (0.05)

Experimental results :

Table 2. Ratio of embeddable bits

e'_i	0	1	2	3	≥ 4
Ratio [%]	51.6	30.0	13.1	4.1	1.3

- From this result, no information is embedded in more than half of the sample. On the other hand, we can embed more than two bits of information into almost 20% of the samples.

Experimental results :

- Using 24 kbit/s, 32 kbit/s and 40 kbit/s ADPCM configurations, changing the maximum number of bits for embedding.
- Figure3 illustrates better performance when the method used Alpha for bit rate control.

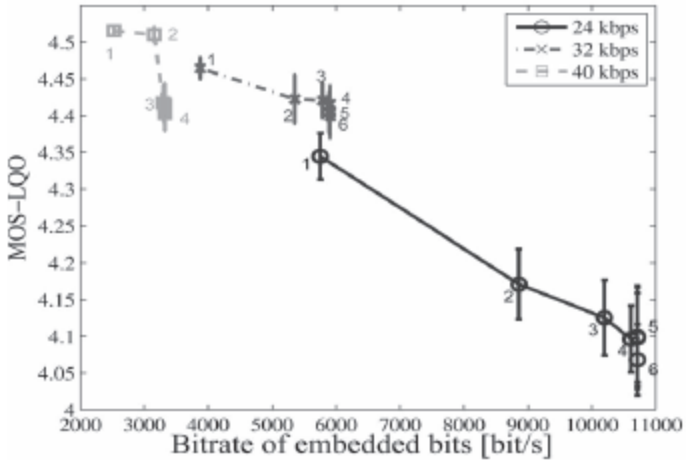

Fig. 2. Bitrate vs. MOS for various configurations

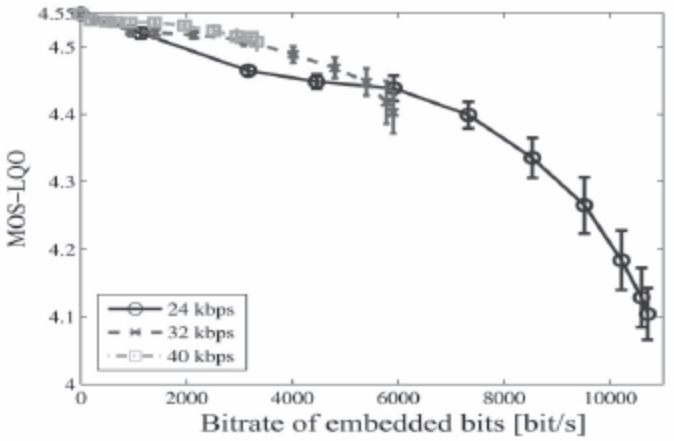

Fig. 3. Bitrate vs. MOS when using α.

Performance optimizations :

- Instead of changing the maximum number of bits for substitution, we can use a different way of controlling the balance between bitrate and speech quality for a given ADPCM configuration.
- $D_i = (g(m^+_{e_i}(x_i)) - g(m^-_{e_i}(x_i)))^2$ =>Tolerable distortion
- *α is a control parameter where 0 < α ≤ 1.*

$$\tilde{e}_i(\alpha) = \max_j \{ j | g(m_j^+(x_i)) - g(m_j^-(x_i))^2 < \alpha^2 D_i \}$$

Drawback of *proposed method* :

- The bitrate of the hidden information is not constant.

Conclusions

- This method outperforms the simple LSB substitution method and the selective embedding method proposed by Aoki.
- Determining the number of bits for embedding, considering the output of a low-bit rate encoder, we can ensure that the quality of the embedded speech is as good as that by the low-bit rate encoder.

Thank You

- Questions ??..

References :

- http://ieeexplore.ieee.org/Xplore/guesthome.jsp
- http://ieeexplore.ieee.org/Xplore/login.jsp?url=http://ieeexplore.ieee.org/iel5/4912736/4959496/04959857.pdf%3Farnumber%3D4959857&authDecision=-203
- http://voip.about.com/od/voipbasics/a/voipcodecs.htm
- http://en.wikipedia.org/wiki/G.726
- http://www.britannica.com/bps/additionalcontent/18/26474667/Packet-Loss-Concealment-for-G722-using-Side-Information-with-Application-to-Voice-over-Wireless-LANs
- http://www.vde-verlag.de/data/prcd.php?docid=442955013