BANDWIDTH EXTENSION OF SPEECH USING DATA HIDING METHOD

[1]P. Srinivasa Rao, [2] V.Sagar Reddy
[1]Assistant Professor, [2]Assistant Professor
[1] Department of Electronics and Communication Engineering,
[1]VNR VJIET, Hyderabad, India

ABSTRACT: The public switched telephone network (PSTN) systems are only able to deliver speech signals in a narrow frequency band about 300–3400 Hz. Such a bandwidth is so small that the intelligibility of speech suffers frequently poor subjective quality. Meanwhile, codes for Wide Band (WB) telephony (50Hz to 7 kHz) exist with significantly improved speech intelligibility and naturalness. However, the broad introduction of wideband speech coding will require strong efforts of both network operator and customers because many elements of the networks have to be modified that turns out to be time-consuming. In order to improve intelligibility and perceived quality of telephone speech, we propose data hiding method based on dither quantization is used for speech bandwidth extension. More specifically, the out-of-band information is encoded and embedded into the narrowband speech without degrading the quality of the band limited signal. At the receiver, when the out-of-band information is extracted from the hidden channel, it can be used to combine with the band limited signal, providing a signal with a wider bandwidth. To encode the out-of-band speech more efficiently. That the proposed approach is robust to quantization errors and channel noises. Al- Though we cannot physically extend the transmission bandwidth of PSTN, the telephony speech quality can be significantly improved by using the proposed data hiding technique.

Key words - Speech bandwidth extension, Data hiding, public switched telephone network (PSTN), wideband speech coding.

I. INTRODUCTION

The current public switched telephone network (PTSN), which has been a part of our daily life for more than century, is designed to transmit toll quality voice only. Quality, intelligibility and naturalness of the speech are the main factors in digital telecommunication systems. The speech quality can be degraded due to Limited bandwidth of the speech signal to the telephone frequency band 300-3400 Hz. Many noise reduction and error concealment techniques have been devised to improve the speech quality and intelligibility still it may sound unnatural and muffled. Due to historical and economic reasons, the audio frequency range of telephone speech is typically limited to approximately the traditional telephone band of 300– 3400 Hz. The mobile users are experiencing the muffled sound of speech due to its narrow audio bandwidth. The lack of low frequencies results in a thin voice and degrades the naturalness and presence, whereas the absence of high frequencies deteriorates especially fricative sounds such as [s] and [f] and causes the characteristic muffled speech quality. This constraint limits the quality of speech received by the listener on either side of the connection. Furthermore, some of the speaker-specific characteristics of speech are lost.

The technology and standardized methods exist for the transmission of speech of a wider bandwidth. Wideband speech typically refers to the acoustic bandwidth of 50– 7000 Hz, which gives a substantially better speech quality with brighter and fuller sound compared to narrowband speech. Until recently, the deployment of wideband speech transmission has been slow, and wideband speech has been mainly used in specific applications such as video conferencing and internet telephony. Wide band speech transmission requires the transmission network and terminal devices at both ends to be upgraded to the wideband that turns out to be a costly for all the users across the communication system.

Presently, majority of the network operators are offering wideband speech services, but it requires the transmission network and terminal devices at both the ends to be upgraded to the wideband that turns out to be prolonged changeover time. To improve the quality during the changeover, novel BWE techniques have been developed using digital signal processing techniques. One such method is bandwidth extension in which the missing spectrum is estimated from narrowband signal. Alternatively, the additional information about the missing frequency regions can be transmitted as side information to support the bandwidth extension. In general, this provides better output quality than artificial bandwidth extension, but the transmission of side information requires special arrangements that may be difficult to implement in practice.

The speech bandwidth can be extended to frequencies lower than or higher than the narrow band or both. The missing frequency band higher than the narrow band typically ranges from 3.4 kHz up to 7 kHz and is referred to as the high band. The missing frequency band below the narrow band typically covers frequencies below 300 Hz and is referred to as the low band. Bandwidth extension methods towards high frequencies create new content in the frequency band from 3.4 kHz to 7 kHz. There are also bandwidth extension methods towards low frequencies, 50-300 Hz. The low band effects the quality and speech naturalness but has only a minor influence on intelligibility, whereas the HB influences the speech intelligibility and quality.

Quality, intelligibility and naturalness of the speech are the main factors in digital telecommunication systems. The speech quality can be degraded due to Limited bandwidth of the speech signal to the telephone frequency band: 300-3400 Hz. Many noise reduction and error concealment techniques have been devised to improve the speech quality and intelligibility still it may sound unnatural and muffled. Especially to distinguish between certain unvoiced or plosive utterances, such as /s/ and /f/ or /p/ and /t/ when applied only a narrowband speech signal. This is due to the fact that the considerable portion of their energy is located in higher frequency components, while the low-frequency characteristic can easily be confused among these sounds. Human speech contains considerably more frequency components than it is being utilized for NB telephone speech coding. It is due to the limitation in storage, coding complexity and bandwidth provided by telephone networks. Since the inception of pulse code modulation (PCM), a speech coding algorithm that has been used in telecommunications for more than 30 years, the frequency bandwidth has been limited to 300 Hz to 3.4 kHz. Implementing wideband system yield the experience of wideband higher signal quality and many more new applications like hands free speaking and teleconferencing

II. WIDE BAND CODER IMPLEMENATION

Implementing wideband system yield the experience of wideband higher signal quality and many more new applications like hands free speaking and teleconferencing. Several wideband speech codes have been standardized in the past. A first wideband speech codec (G.722) was specified by CCITT for ISDN and tele-conferencing with bit rates of 64, 56 and 48 Kbit/sec. It is mainly applied in context with radio broadcast stations by external reporters using special terminals and ISDN connections from outside to the station. In 1999, a second wideband codec (G.722.1) was introduced by ITU-T that produces almost comparable speech quality at reduced bit rates of 32 and 24 Kbit/sec. Most recently, the adaptive multi-rate wideband (AMR-WB) speech codec was standardized by ETSI and 3GPP for CDMA cellular networks such as UMTS.

The AMR-WB codec has also been adopted for fixed network applications by ITU-T (G.722.2) [25]. By the AMR-WB standard a family of wideband codes with nine data rate modes between 6.6 and 23.85 Kbit/s is defined together with control mechanisms to adapt the codec mode to channel conditions. Further research has been extended to the AMR-WB+ codec that support general audio in mono/stereo with frequency bandwidths from 7 to more than 16 kHz and bit rates of between 6.6 and 32 bit/s.

The WB coders can only be used if the user-end terminals, the network and protocols all have the improved WB capabilities in terms of hardware and software compatibility. In addition, signaling procedures are needed for detection and activation of the wideband capability considering the fact about better speech quality performance offered by WB coders; still sudden replacement of entire NB coding and transmission systems is not feasible because of tremendous infrastructure expenses incurred to operators and also is a case with customers. Current speech transmission system is a mixture of traditional narrowband terminals and new wideband terminals. It will take longer time to replace all the equipment, protocols and whole transmission link supporting wideband transmission. The long transitional period, between up gradation of narrowband to wideband system, demands to enhance speech quality without much modification of already existing network infrastructure. It has motivated the approach of bandwidth extension. During this transition period different technical solutions may be employed. All of these solutions produce WB speech at the near-end terminal.

III. IMPLEMENTATION OF BWE IN NB CODER

Many technical solutions can be considered during the long transitional period of NB and WB telephony for generating wideband speech at the near end terminal. One alternative solution is to implement bandwidth extension algorithm in legacy narrowband coder. Bandwidth extension artificially adds the missing frequencies of the signal at the receiver, using only the information contained in the narrowband signal or either using the side information transmitted. This produces more natural sounding speech, and the user can benefit from the improved wideband capabilities of the terminal.

IV. DATA HIDING SCHEME

Before introducing the data hiding scheme, some terminologies have to be clearly defined. In this paper, we denote the bandwidth of 300–3400 Hz, which is used by the PSTN channel, as “NB”, and the bandwidth of 50–7000 Hz as “WB”. “EB” is referred to the higher frequency part in WB but beyond NB.

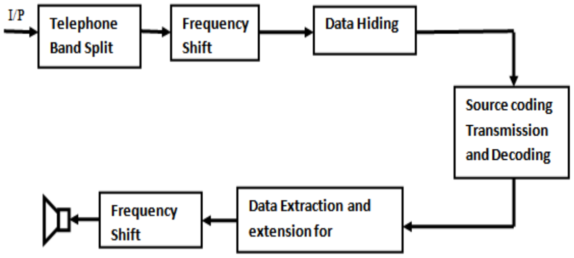
I/P
Telephone Band Split
Frequency Shift
Data Hiding
Source coding Transmission and Decoding
Data Extraction and extension for
Frequency Shift

Fig. 1.The block diagram of the proposed data hiding in speech transmission.

To encode the EB part and embed it into the NB signal, the original WB speech has to undergo band splitting initially, as shown in Figure 1. The WB speech is filtered by a low-pass and a high pass filter respectively. The output of the low-pass filter is down-sampled to provide the NB signal $x_{nb}(k)$, $1 \leq k \leq N$, where N is the total number of samples. The output of the high-pass filter is shifted to the NB frequency range, and also decimated to provide an NB version of x_{eb}. We can denote it as $x_{ne}(k)$, $1 \leq k \leq N$. Based on x_{ne}, the AR coefficients a_{eb} are estimated and encoded by LPC. It should be noted that data hiding requires imperceptible embedding. However, the more data bits are to be inserted, the more distortion will be introduced to the original NB speech. Therefore, it would be desirable to minimize the number of data bits used to encode x_{ne}. We do not encode the excitation signal s_{eb} for minimizing the number of bits to be embedded. This is because, according to the observations from [8], human ear is amazingly insensitive to distortions of the excitation signal at frequencies above 3.4 kHz. For example, spectral gaps of moderate width as produced by band stop filters are almost inaudible. Furthermore, a misalignment of the harmonic structure of speech at high frequencies does not significantly degrade the subjective quality. Therefore, an estimate of s_{eb} from the NB signal at the receiver is good enough for the reconstruction performance. Many data hiding approaches have been proposed in literatures. According to whether knowledge of the original host data will enhance the detection performance, most of them would fall in non-blind or blind categories. Conventional spread spectrum approaches are in the non-blind category, since they treat the host data as a source of interference. However, in this application, there is no chance for the original NB speech to be available at the receiver. We thus propose to use a simple blind scheme based on dithered quantizes.

V. PERFORMANCE ANALYSIS

1. Embedding Distortion

Since the embedding of $\{b_j\}$, $j = 1,\ldots, 23$, is based on quantization, the introduced distortion comes mainly from quantization errors. In the high signal-to-noise ratio (SNR) scenario, we can assume the quantization cells are small enough that the host speech sample can be modeled as uniformly distributed within each cell.

2. System Complexity

At the transmitting side, the proposed scheme employs a part of G.729 to perform LPC. The encoded AR coefficients are inserted into the NB speech by a simple quantization and perturbation process. At the receiver side, the hidden data is extracted by applying a minimum-distance decoder, which is widely used in telecommunications. The decoding of the extracted data bit is also performed as a part of G.729. That is, no complicated computation is involved. Therefore, the (6) proposed scheme is feasible for real-time processing.

VI. EXPERIMENTAL RESULTS

The assessment of the perceptual quality for the composite signal x_{nb} is carried out by mean opinion score. The subjects are asked to compare x_{nb} with x_{nb} and give their opinions. α is set as small as 0.04 to reduce embedding distortion. Scaling of MOS is 4 grades and their instructions are as follows:

1. Two signals are too different
2. Two signals are similar, but the difference is more
3. Two signals sound very similar, only little difference exists
4. Two signals sound identical

The two signals are played in a random order, and the sound pressure level is set as 63 dB. Obtained by averaging ratings of all subjects over all testing signals, the resultant score comes to 3.06. Therefore, although the embedding of extra does have a negative impact on the perceptual quality, such a small degradation is still acceptable.

The BWE algorithm presented in is also implemented for the comparison purpose. The state number of the hidden Markov model is set as 2, 4, 8 and 16 respectively. In each state, 16 Gaussians are used. Both speaker-independent and dependent training are tried. For the proposed scheme, α is still set as 0.04. Figure 4 plots the LSD results for both schemes under μ-law coding. It is seen that the proposed data hiding scheme consistently outperforms the BWE method with speaker-independent training. It is also superior to the BWE method with speaker-dependent training when the state number of HMM is less than 8. Only when a large number of states is involved, the BWE with the speaker-dependent training is better than the proposed scheme. However, the more states are involved in the training, the more time is required to compute the emission and transition probabilities, etc., for HMM. The proposed scheme uses the true EB signal to estimate and encode a_{eb}. As long as the hidden data is robust to μ-law or A-law coding, a_{eb} can be recovered precisely without high computation load.

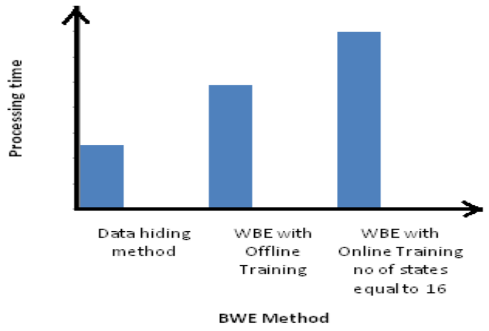

Fig.2 The processing time for the data hiding scheme and the BWE scheme.

We further investigate the processing time required for implementing the proposed data hiding scheme, the BWE with speaker-independent (offline) and speaker-dependent training. A PC with Intel Pentium IV 1.5 GHz CPU is used to run Matlab programs for the schemes to be compared. Although using C or C++ may be more appropriate and faster than using Matlab, the relative time difference between the tested schemes will not change much. Compared to conventional BWE schemes, the proposed data hiding scheme is found to have an absolute superiority in the processing speed, especially to the BWE with speaker-dependent training.

VII. REFERENCES

[1] International Telecommunication Union, “7 kHz audio coding with 64 kbit/s,” ITU-T Recommendation G.722, 1993.

[2] J. D. Markel and A. H. Gray, Linear Prediction of Speech, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[3] C. Avendano, H. Hermansky and E. A. Wan, “Beyond Nyquist: Towards the recovery of broad-bandwidth speech from narrow-bandwidth speech,” Proc. of European Conference on Speech Communication and Technol-ogy, Madrid, Sept. 1995

[4] P. jax and P.vary –Artificial bandwidth extension of speech signals Using MMSE estimation based on a hidden markov model‖ Institute of communication systems and data processing (ind), Aachen University (RWTH), 52056 Aachen, Germany,IEEE(ICASSP) 2003.

[5] A. V. Oppenheim, R. W. Schafer and J. R. Buck, Discrete-Time Signal Processing, 2nd edition, NJ: Pren-tice Hall, 1999.

[6] N. S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Engle-wood cliffs, NJ: Prentice-Hall, 1984.

[7] P. Jax and P. Vary, “On artificial bandwidth extension of telephone speech,” Signal Processing, vol. 83, pp. 1707–1710, 2003.

[8] N. S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Englewood Cliffs, NJ: Prentice-Hall, 1984.