A NOVEL BWE SCHEME BASED ON SPECTRAL PEAKS IN G.729 COMPRESSED DOMAIN

*Murali Mohan Deshpande * and K.R.Ramakrishnan**

* Indian Institute of Science, Department of Electrical Engineering, Bangalore, India
Email: dmmohan@gmail.com and krr@ee.iisc.ernet.in

ABSTRACT

This paper addresses the problem of interoperability of narrowband VoIP network with wideband wireless network. Bandwidth extension in G.729 compressed domain is proposed to address this problem. The proposed method is based on spectral envelope extension using classified codebook approach. The number of spectral peaks is used as the feature for classification in this method. The proposed method gives a performance improvement of around 0.8dB with respect to existing method of Decode Then Extend (DTE). This compressed domain method also reduces the computational cost by around 2.5 MIPS on TMS320C55x DSP with respect to DTE.

1. INTRODUCTION

The voice data in a wideband digital cellular network is compressed in WB-GSM AMR encoded format. The handling of VoIP calls on the cellular network requires the transcoding gateway, which takes in the G.729 (Most commonly used standard in VoIP) encoded packets, decodes it and then up-samples and reconverts it into the WB-GSM AMR stream as required by the wireless network. The up-sampling operation does not add any new spectral information. Hence, effectively trans-coder encodes the narrowband speech using the WB-GSM AMR algorithm (This is a speech compression standard which operates on wideband speech). There are various speech enhancement algorithms that enhance the speech quality by doing a Bandwidth Extension (BWE) [1] [2] of narrow band speech. These BWE schemes use the correlation or mutual information between the narrow band and the high band portions of speech [3]. Most algorithms for the extension of the bandwidth of speech signals are based on a simple linear model of the speech production process. According to the linear model of the speech production process, the algorithm for bandwidth extension can be divided into two tasks, which are, to a certain extent mutually independent. They are extension of the spectral envelope [4] [5] of the speech signal and excitation signal [4]. These bandwidth extension algorithms can be used for enhancing the speech quality in the interoperability of the VoIP with the cellular network by extending the bandwidth of G.729 decoded Narrow band (NB) speech to get the synthesized wide-band (WB) speech. This method is referred to as Decode then Extend (DTE). The DTE scheme introduces algorithmic delay and computational delay, which add to the overall delay in the network.

This paper proposes the Bandwidth extension scheme in the G.729 compressed domain by modifying the G.729 decoder for doing the BWE by partial decoding of the G.729 codewords. The modified decoder uses a novel method for spectral envelope extension based on number of spectral peaks.
The organization of the rest of this paper is as follows, In section 2, the proposed Classified codebook mapping scheme is discussed, section 3 provides the Experimental results and conclusion is presented in section 4.

2. CLASSIFIED CODEBOOK MAPPING SCHEME FOR BWE

This section discusses the proposed method for BWE in detail. The BWE scheme consists of the following steps,

- Classified Codebook Pairs Training (2.1, 2.2) - This requires the determination of spectral peaks creating three different training databases corresponding to different number spectral peaks and training of codebook pairs using Carl's method.
- Spectral envelope extension using classified codebook mapping (2.3).
- Excitation extension using pitch lag available in G.729 (2.4).
- Attainment of spectral continuity (2.5).
- Modifications to G.729 for doing BWE in compressed domain (2.6).

The further sub-sections discuss the above-mentioned steps in detail.

2.1 Determination of spectral peaks

The spectral envelope is obtained by evaluating over the unit circle from $(0,\pi)$ the polynomial,

$$\frac{G}{A_p(z)} \quad \text{where } G=\sqrt{e_p} \qquad (1)$$

and e_p is the prediction error from the LPC analysis filter $A_p(z)$. The log spectral envelope is determined using,

$$H(\omega)=20\log_{10}\left(\frac{G}{A_p(z)}\right) \qquad (2)$$

To determine the spectral peaks of $H(\omega)$ from discrete samples of the spectral envelope, differences between successive values are computed. A peak in the spectral envelope corresponds to a sign change in the spectral differences from negative to positive. Hence, the number of spectral peaks corresponds to the number of sign changes in one direction (negative to positive only).

2.2 Codebook Pair Design

Various codebooks are designed corresponding to frames having same number of spectral peaks. In this scheme three codebook pairs are trained for spectral peaks ranging from one to five. 10^{th} order LSP is used for narrow band speech and 18^{th} order LSP is used for wideband speech .The LSP parameters are calculated every 10ms frame of NB and the WB speech for the same corresponding speech frames after the classification is done to the training set based on the number of spectral peaks in the envelope. (Using the same frame structure as in G.729) The narrowband codebook for each of the classified set is generated for required size using LBG algorithm. The narrow band training data are quantized by predetermined NB codebook, Each wideband vector of index p is formed by taking the centroid of all the wideband LSPs whose corresponding narrowband LSPs were quantized by narrow band code-vector with index p. This method is conceptually similar to Non linear interpolated vector quantizer (NLIVQ). The NLIVQ results in the design of optimal WB codebook C_H over the training data. This algorithm is popular as Carl's method [4] and is given below,

- *Design a narrow band codebook C_L corresponding to NB speech for LSPs of order 10 using Generalized Llyods algorithm and codebook initialization is done by splitting. This is done using the entire training data T_L that consists of M vectors of dimension 10 each.*
- *Quantize T_L using C_L store the resulting indices i_n where $1 \leq n \leq M$.*
- *Calculate the J^{th} code-vector of C_H as mean of all vectors in T_H whose indices are equal to j.*

2.3 Mapping of LSP Parameters

Using the Partially decoded LPC parameters in the G.729 decoder, identify the number of spectral peaks in the envelope and choose the codebook based on this information. Mapping of LSPs can be described in the following steps, Let C_L Represent the 10^{th} order LSPs for the narrowband speech. N represents the number of NB code-vectors.

$$C_L = \{ \hat{a}_i \varepsilon R^P \mid 1 \leq i \leq N \} \qquad (3)$$

Let C_H Represent the q dimensional LSP parameters of the wideband speech. The number of wideband code-vectors is equal to NB code-vectors i.e., N.

$$C_H = \{ \hat{g}_i \varepsilon R^q \mid 1 \leq i \leq N \} \qquad (4)$$

Codebook mapping can be written as

$$F : R^P \to C_H \qquad (5)$$

$$F(a_i) = \hat{g}_i \text{ where } i = \arg\min_i d(a_i, \hat{a}_i) \qquad (6)$$

Here a_i represents 10^{th} order narrowband LSP vector F incorporates the partition of narrowband space into regions. Thus for an input narrowband parameter vector a_i the mapping $a_i \rightarrow \hat{a}_i \rightarrow \hat{g}_i$ is made by considering that index which corresponds to the minimum distance $d(a_i, \hat{a}_i)$ of these two vectors. d corresponds to distance measures which are optimally chosen for the LSP vectors by considering the LSP positions along with their Euclidean distance [6].

2.4 High band excitation extension

The residual is got as sum of the fixed and adaptive codebooks in the G.729 decoder. The pitch delay is present in the G.729 codeword. If the speech is unvoiced the pitch period is of little importance but the output of the pitch detector can still be used.

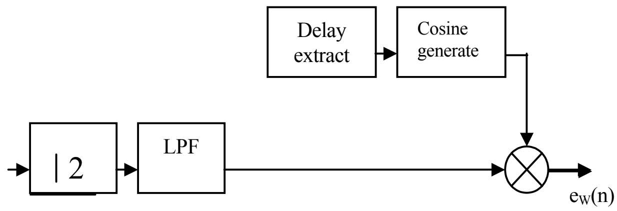


Figure 1: *Spectral shifting using pitch delay*

The various steps in the excitation extension is shown in Figure 1 and explained below,

- Up-sample the residual by a factor of 2.
- The up-sampled residual is mixed with a cosine of amplitude 2 at a radian frequency ω_g, where ω_g is a multiple of fundamental frequency that is close to the cut-off frequency ω_c of NB speech as given in equation (7).

$$\omega_g = (2\pi/T_f) \,\text{floor}(T_f\, \omega_c/2\pi) \qquad (7)$$

2.5 Preserving NB and WB spectral envelope continuity

The spectrum of the NB and WB speech can be computed using the LPC synthesis filter that is computed for the NB and WB speech sub-frames. If there is a frequency domain overlap between input narrowband envelop and estimated high band envelop then the high band gain can be estimated by matching the two envelops over some overlapping range which yields a gain as given in equation (8),

$$g_M(dB) = \frac{1}{\omega_H - \omega_L} \sum_{\omega_l=\omega_L}^{\omega_H} 20\log_{10} \frac{|W(\omega_l)|\left|A_N\left(e^{j\omega_l T_N}\right)\right|}{\left|A_H\left(e^{j\omega T_H}\right)\right|} \qquad (8)$$

Where A_N and A_H are narrowband and high band envelope magnitude response, T_N and T_H are narrow band and wide band sampling periods respectively. ω_L and ω_H are the lower and upper frequency limits of the matching range and are set to the NB frequency range. $W(\omega)$ is a frequency dependent weighting function which is chosen to be unity inside and zero outside the range of the narrowband speech signal. After matching the highband spectra with the narrowband spectra the spectral continuity is attained by doing splicing [1], The NB envelope $A_N(\omega)$ and the high-band envelope $A_H(\omega)$ are spliced as follows,

$$A_W(\omega)=\begin{cases} A_N(\omega), & \omega \le \omega_{NH} \\ \dfrac{\omega_{HL}-\omega}{\omega_{HL}-\omega_{NH}} A_N(\omega) + \dfrac{\omega-\omega_{NH}}{\omega_{HL}-\omega_{NH}} A_H(\omega), & \omega_{NH} < \omega \le \omega_{HL} \\ A_H(\omega), & \omega > \omega_{HL} \end{cases}$$

2.6 Modified G.729 Decoder – Data flow

Figure 2a shows block diagram of various steps for the BWE in the G.729 compressed domain. Figure 2b shows the expanded view of codebook mapping block. The data flow in the proposed scheme is explained below,

- Extract the 10 interpolated LSPs available every subframe (5ms) which is available after the partial decoding at the decoder.
- The decoded and interpolated narrowband LSPs are converted to LP coefficients for each subframe.
- The LP coefficients are used to find the number of spectral peaks in the given spectral envelope.
- The extracted 10 LSPs are fed to the WB spectral envelope estimator which is based on the codebook mapping to get the WB estimated LSPs of order 18. Appropriate codebook pair is selected based on number of spectral peaks (from three codebook pairs) for doing the mapping (Figure 2a).
- The residual vector is tapped after partial decoding and is fed to high band excitation regenerating block. The 18^{th} order estimated WB LSPs are converted to get the WB LP synthesis filter. Match the original narrow-band spectral envelop with the wideband spectral envelop for spectral continuity.
- The wideband speech is reconstructed after passing the high band excitation through the WB LP synthesis filter. The regenerated WB speech is high pass filtered using a 10^{th} order chebyshev filter with 0.25dB pass-band ripple and 3400 Hz cut-off frequency.
- The Narrow-band speech is up-sampled by a factor of 2. Add the up-sampled narrowband speech and the reconstructed high band speech to get the WB speech.

3. EXPERIMENTAL RESULTS

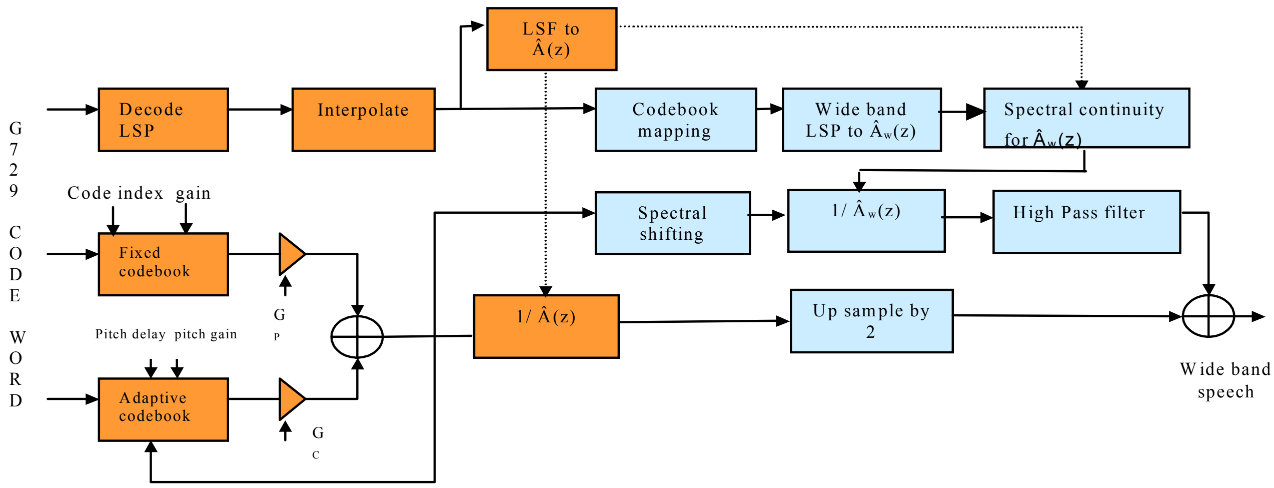
The proposed bandwidth extension in G.729 compressed domain was simulated using Matlab and C routines. The codebook training was conducted using TIMIT database consisting of wideband speech files. The speech files used consisted of around 165944 speech frames of size 10ms each. The speech files consisted of speech uttered by many speakers, male and female, with different native tongues and encompass as many utterances as possible. These wideband speech files were low-pass filtered using a Chebyshev filter of order 20 with cutoff frequency of 3.4 KHz and down-sampled to form the narrowband database. The narrowband and the corresponding wideband LSPs of order 10 and 18 respectively for the speech frames were computed. Three codebooks corresponding to one/two peaks , three and four/ five peaks respectively was trained with sizes 32, 128 and 256 respectively (This corresponds to optimal training ratio of 325). The codebooks were trained by making use of two appropriate distance measures suitable for the LSP vectors (Considering positions of LSPs aswell). The narrowband test vectors outside of the training set were passed through the G.729 encoder to create G.729 codewords. These code-words were given as inputs to the modified G.729 decoder that performs the bandwidth extension along with decoding. The output of the decoder was compared with the actual wideband speech testvector using the highband distortion measure [1]. The value of the average high band distortion is obtained by considering the average over all the frames in the test vector. The value of the average highband distortion along with the corresponding value for DTE scheme is shown in the Table 1. It is observed that the average high band distortion is approximately 0.8dB less in the case of bandwidth extension using the compressed domain ap-proach. Hence bandwidth extension in compressed domain using classified codebooks generates significantly better quality of speech than the DTE based scheme. The imple-mentations of the common routines in the bandwidth extension scheme and G.729 codec were done on the TMS320C55x processor from Texas Instruments Inc., to calculate the computational gain by doing the BWE in the compressed domain. The details are given in the Table 2 given below,

Bandwidth Extension Scheme	**Average Highband spectral distortion (dB)**
Raw speech based extension	3.8289
Compressed domain based extension using classified codebooks (# of spectral peaks) and distance 1 for training and mapping.	3.0624
Compressed domain based extension using classified codebooks (# of spectral peaks) and distance 2 for training and mapping.	3.0336

Table 1: *Objective distortion measure comparison table*

Function	MIPS on TMS320C55x
Windowing	0.04
Autocorrelation Calculation	0.46
Levinson Durbin	0.41
A(z) to LSP mapping	0.96
Pitch detection using AC	0.70

Table 2: *MIPS of common routines on TMS320C55x*



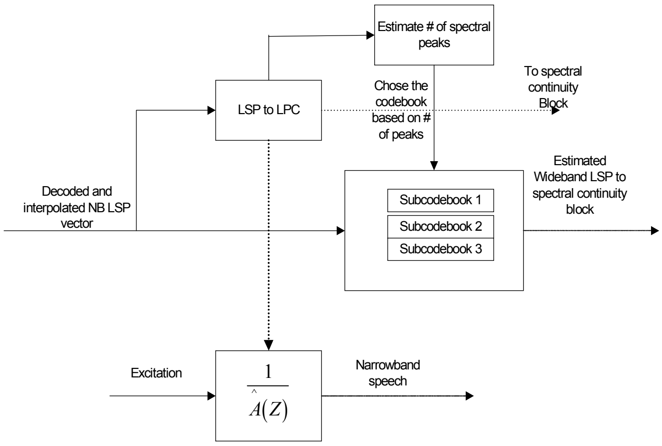


Figure 2a, 2b : *a. Data flow in G.729 decoder. b. Expanded view of codebook mapping.*

There is a saving of 2.57 MIPS on C55x by doing the BWE in the compressed domain. This saving is significant in the multichannel applications such as gateways as this enables us to support higher number of channels compared to DTE scheme. Any Bandwidth Extension scheme has an inherent delay of 10 to 30 ms depending on the frame size we use to extract the parameters. In the proposed scheme this algorithmic delay is totally eliminated, as the parameters are readily available in the compressed domain. Blind A/B/C Test was conducted for evaluating the subjective performance. The speech generated by the proposed scheme was preferred 90% of the times with respect to the speech generated by DTE.

4. CONCLUSION

This paper has proposed bandwidth extension of narrowband speech in G.729 standard based networks like VoIP. The proposed method removes the impairments that are present in the DTE scheme and is based on classified codebook approach using number of spectral peaks as feature. The average highband spectral distortion was observed to be around 3dB in the proposed scheme i.e., an improvement of 0.8dB in the highband spectral distortion measure compared to DTE scheme. The speech generated by the proposed scheme was found to be preferred 90% of the time in subjective listening tests. Further the proposed scheme reduces the algorithmic delay as the parameters for BWE are extracted from the G.729 codeword and achieves computational gain of around 2.57 MIPS that is critical in multi-channel applications.

REFERENCES

[1] Epps, J., and Holmes, W.H. (1999). "A new technique for wideband enhancement of coded narrowband speech", , IEEE Workshop on Speech Coding Proceedings, pp.174 - 176 , June.

[2] Peter Jax, Peter Vary (2003) "On Artificial Bandwidth Extension of Telephone Speech", SIGNAL PROCESSING, vol. 83, no. 8, pp. 1707-1719, August.

[3] Nilsson, M., Andersen, V., and Kleijn, W.B. (2000). "On the Mutual information between frequency bands in Speech", in Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing, vol 3, pp. 1327 –1330.

[4] Carl,H., and Heute, U. (1994)."Bandwidth enhance ment of narrowband speech signals " , Signal Process ing VII, Theories and applications, EUSIPCO, vol2, pp. 1178-1181

[5] Avendano, C., Hermansky, H and Wan, E. A. (1995). " Beyond Nyquist: towards the recovery of broadband width speech from narrow-bandwidth speech", in Proc. 4th European Conf. On speech Commun. And Technol., EUROSPEECH (Madrid, Spain), vol. 1, pp.165-168,

[6] Murali Mohan Deshpande, K.R.Ramakrishnan (2004). "A Novel LSP mapping scheme for speech enhancement by BWE", International Conference on DSP applications (DSPA 2004), Moscow, Russia, vol. 1, pp.213-216.