A ROBUST NARROWBAND TO WIDEBAND EXTENSION SYSTEM FEATURING ENHANCED CODEBOOK MAPPING

*Takahiro Unno and Alan McCree**

DSPS R&D Center, Texas Instruments, Dallas, Texas
E-mail: takahiro@ti.com, mccree@ll.mit.edu

ABSTRACT

It is well-known that wideband speech (0 - 7 kHz) provides better quality and intelligibility than narrowband speech (300 - 3400 Hz), but typically only narrowband speech information is available in current wireless communication systems. Narrowband to wideband extension technology has been recently investigated to artificially generate wideband speech from narrowband speech for better speech quality and intelligibility. This paper presents a robust split-band narrowband to wideband extension system based on algorithmic enhancements to the codebook mapping technique for high-band parameter estimation. Numerical measurements confirm the performance improvements of the codebook mapping process, and informal listening evaluations show the potential of the system and its robustness to input distortions and non-speech input signals.

1. INTRODUCTION

Traditionally, digital speech coders have operated with the bandwidth of the analog telephone network, from 300 to 3400 Hz, and therefore used a sampling rate of 8 kHz. However, the increasing use of wideband (0 - 7 kHz) speech coders in digital networks has raised awareness of the potential quality improvements with wider bandwidths. In narrowband communication systems such as digital cellular, the receiver in the handset does not have access to the wideband signal. However, recent work [1, 2, 3, 4, 5] has shown that by signal processing techniques, the receiver can artificially generate a wideband signal from the narrowband information using a process called narrowband to wideband extension (or simply wideband extension).

In this paper, we present our wideband extension baseline system, codebook mapping performance enhancements including predictive codebook mapping and optimal codebook interpolation, robust system features to input signal distortion, and informal listening evaluation results.

2. BASELINE SYSTEM

Our wideband extension baseline system consists of three parts: high-band parameter estimation, high-band signal generation, and synthesis filterbank. Figure 1 shows the block diagram of the baseline system. Our system uses a split-band synthesis method in which an artificial high-band signal (4 - 7 kHz, 8 kHz sampling) is generated to produce wideband speech. As shown in the figure, a codebook mapping first estimates the high-band parameters (gain and linear prediction coefficients (LPC)) from narrowband information. Then, artificial high-band speech is generated using a noise excitation LPC synthesis model in which modulated random noise is amplified by the estimated gain and synthesized with the estimated LPC. Finally, the input narrowband speech is up-sampled and added to the up-sampled artificial high-band speech to produce wideband speech using a synthesis filterbank. This process is performed every 10 ms, with 5 ms interpolation, to track rapid speech changes.

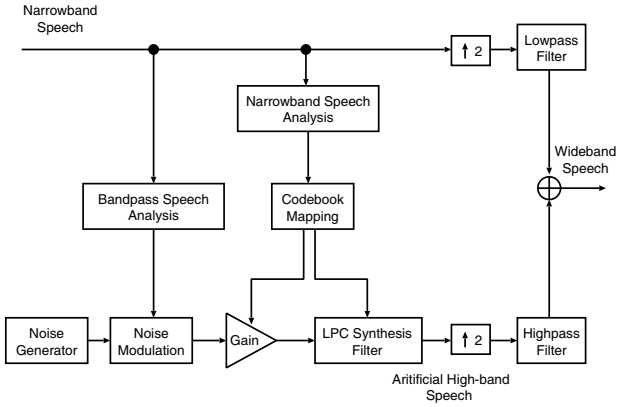

Fig. 1. Block diagram of wideband extension system

2.1. Excitation

The artificial high-band signal is generated with the LPC excitation model. In an embedded wideband speech coding system, pitch-modulated noise excitation was shown to provide good quality of high-band speech [6]; this has also been confirmed in a previous wideband extension system [5]. For this application, we use the envelope of the bandpass signal between 2.5 - 3.5 kHz to modulate the high-band excitation noise since the signal above 3.5 kHz may be cut off by telephone network filtering.

2.2. High-band Parameter Estimation

A number of techniques for high-band or wideband parameter estimation have been recently presented [1, 2, 3, 4]. In particular, codebook mapping has been shown to be a promising method. In codebook mapping, the high-band or wideband speech parameters are estimated from the narrowband speech information using codebooks of narrowband and the corresponding high-band or wide-

*Currently with MIT Lincoln Laboratory, Lexington MA

band parameters. For each input speech frame, the narrowband information is mapped into the nearest codebook entry and the corresponding high-band or wideband parameters are extracted from the mapped codevector in high-band/wideband codebook. Codebooks are generated from narrowband and high-band/wideband speech training data using Generalized Vector Quantization design techniques [7].

2.3. Synthesis Filterbank

Our system uses a Butterworth IIR filterbank to provide low delay [8]. This IIR filterbank provides typical delay of less than 0.3 ms, and peak delay of less than 1 ms. As there is no lookahead required in the system, the algorithm delay of our system is equivalent to the filtering delay.

2.4. Performance

In informal listening, this high-band synthesis model with unquantized parameters was shown to provide very high-quality wideband speech. However, there is significant degradation due to the codebook mapping process.

Codebook mapping performance depends on various factors such as codebook size (the number of codebook entries), the type of narrowband information and the codebook partitioning method. Table 1 shows Spectral Distortion (SD) of narrowband and high-band for different codebook sizes. For this measurement, the narrowband parameters were 10th-order line spectral frequencies (LSF) and high-band parameters were 6th-order LSF and high-band gain (relative gain against bandpass speech (2.5-3.5 kHz)). Narrowband SD is computed over 0 - 4 kHz with Bark weighting [9], and high-band SD is computed as follows:

$$D_{HB} = \sqrt{\frac{1}{7000-4000}\frac{G_{HB}^2}{\hat{G}_{HB}^2}\int_{4000}^{7000}\left|10\log\frac{|\hat{A}_{HB}(f)|^2}{|A_{HB}(f)|^2}\right|^2 df} \quad (1)$$

where $\hat{A}_{HB}(f)$ and $A_{HB}(f)$ are quantized and unquantized high-band LPC spectrum, and $\hat{G}_{HB}$ and G_{HB} are quantized and unquantized high-band gain. As the narrowband input speech is reproduced in the artificial wideband speech, only the high-band SD is relevant for wideband extension performance; we compute the narrowband SD for information only. Details of training data are described in Sec 4.1. Although Table 1 shows that increasing the codebook size provides better SD, 256-level codebooks are used in our system to maintain reasonable data ROM size and codebook search complexity.

Codebook Levels	NB SD (dB)	HB SD (dB)
32	3.59	10.02
64	3.31	9.78
128	3.05	9.62
256	2.82	9.45
512	2.62	9.33
1024	2.44	9.19
2048	2.28	9.09
4096	2.13	9.02

Table 1. Codebook mapping performance for different codebook sizes.

In our preliminary experiments, the addition of low-pass energy to the narrowband parameters was shown to improve high-band SD. This may be because low-pass energy is smoothly changed in vowel segments and also includes some voicing information. However, as low-pass energy causes input gain level dependency of codebook mapping performance, we added open-loop low-pass energy prediction error instead of instantaneous low-pass energy to the narrowband parameters. Table 2 shows the high-band SD improvement from adding the low-pass energy prediction error.

NB Information	NB SD (dB)	HB SD (dB)
LSF	2.82	9.45
LSF and LP Energy Pred. Error	2.89	9.13

Table 2. Codebook mapping performance for different narrowband information with codebook size 256.

3. PERFORMANCE ENHANCEMENT

We made two performance enhancements on the codebook mapping method: predictive codebook mapping and optimal codebook interpolation. Predictive codebook mapping exploits time history and smoothes high-band parameters over time. The optimal codebook interpolation method interpolates neighboring codevectors to improve codebook mapping performance without increasing codebook size.

3.1. Predictive Codebook Mapping

One problem with conventional codebook mapping methods for wideband extension is the inability to exploit time history information, since the codebook mapping is based only on the current speech frame. Intuitively, utilizing additional information from the past speech frame should enable a more accurate codebook mapping. In addition, the high-band parameters may not be consistent from one frame to the next, resulting in perceptually-annoying artifacts. The method proposed in [2] applies smoothing to the decoded wideband parameters to improve subjective quality, but it would be preferable to incorporate the smoothing into the codebook design process. We have developed such an approach using predictive codebook mapping, an extension of predictive VQ to the case of generalized VQ.

In traditional VQ, the quantized vector is reconstructed from the quantization codebook of the input vector:

$$\hat{X} = Q(X) \quad (2)$$

with the quantization codebook represented by $Q()$. In predictive VQ, the codebook is applied to the difference between the current input vector and its estimated or predicted value based on the previous value:

$$\hat{X} = Q(X - \alpha\hat{X}_{prev}) + \alpha\hat{X}_{prev} \quad (3)$$

where α represents a prediction coefficient between 0 and 1. $\hat{X}_{prev}$ can be the previous quantized or unquantized vector (closed-loop or open-loop prediction). The use of prediction has been shown to improve quantizer performance and also results in smoother output signals.

In generalized VQ, the estimated output vector $\hat{Y}$ is in a different domain than the input vector X, based on a codebook mapping $C()$:

$$\hat{Y} = C(X) \tag{4}$$

We have developed a method called predictive generalized VQ, given by:

$$\hat{Y} = C(X - \alpha_x \hat{X}_{prev}) + \alpha_y \hat{Y}_{prev} \tag{5}$$

where the prediction coefficients in the X and Y domains may be different. While in general vector spaces there may be no useful relationships between the time histories in the two different domains, for the particular case of wideband extension we expect similar evolution of the low-band and high-band spectral information.

Experimentally we have found that predictive codebook mapping method improves codebook mapping performance. For our 10 ms frame rate, setting both prediction coefficients to 0.5 works well. This improvement is shown in Table 3.

	NB SD (dB)	HB SD (dB)
Non-Predictive	2.89	9.13
Predictive	2.42	8.68

Table 3. Codebook mapping performance for non-predictive and predictive codebook mapping methods.

3.2. Optimal Codebook Interpolation

Another problem with the conventional codebook mapping approach is the need for significant storage and computational complexity for large codebooks to provide better codebook mapping performance. In the method of [1], the N nearest codevectors are averaged together to provide higher resolution codevector. While this provides some improvement, it is not optimal and can actually degrade performance for speech frames where the nearest neighbor provides a close match.

We estimate the optimal interpolation coefficents between nearest neighbors in the codebook mapping approach for wideband extension by assuming that the narrowband and high-band parameter spaces are locally similar. For example, if a narrowband input vector is halfway between two codevectors, then the corresponding high-band vector should be very close to halfway between the two corresponding high-band codevectors.

So we first find the optimal coefficients in the narrowband domain (where the precise vector is known) and apply the same interpolation between candidate codevectors in the high-band domain. For the narrowband parameter vectors $\mathbf{x}$, the optimal interpolation coefficients can be shown to satisfy a form of normal equations. Suppose we derive optimal coefficients from $N+1$ nearest codevectors. The interpolated narrowband parameter vector $\hat{\mathbf{x}}$ is shown as follows.

$$\hat{\mathbf{x}} = (1 - \sum_{i=1}^{N} a_i)\mathbf{x_0} + \sum_{i=1}^{N} a_i \mathbf{x_i} \tag{6}$$

where $\mathbf{x_i} (i = 0...N)$ are the $N+1$ nearest codevectors for the input narrowband parameter vector $\mathbf{x}$, and a_i is interpolation coefficient. We rewrite this equation in terms of difference vectors $\mathbf{d_i} = \mathbf{x_i} - \mathbf{x_0}$ and $\hat{\mathbf{d}} = \hat{\mathbf{x}} - \mathbf{x_0}$:

$$\hat{\mathbf{d}} = \sum_{i=1}^{N} a_i \mathbf{d_i} \tag{7}$$

Minimizing the squared error between $\hat{\mathbf{x}}$ and $\mathbf{x}$ is equivalent to minimizing the error between $\hat{\mathbf{d}}$ and $\mathbf{d} = \mathbf{x} - \mathbf{x_0}$, and results in the normal equation:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} |\mathbf{d_1}|^2 & \dots & < \mathbf{d_1}, \mathbf{d_N} > \\ \vdots & \ddots & \vdots \\ < \mathbf{d_N}, \mathbf{d_1} > & \dots & |\mathbf{d_N}|^2 \end{bmatrix}^{-1} \begin{bmatrix} < \mathbf{d}, \mathbf{d_1} > \\ \vdots \\ < \mathbf{d}, \mathbf{d_N} > \end{bmatrix} \tag{8}$$

Finally, high-band parameter $\hat{\mathbf{y}}$ is derived by interpolating $N+1$ corresponding codevectors $\mathbf{y_i}$ in high-band codebook as follows:

$$\hat{\mathbf{y}} = (1 - \sum_{i=1}^{N} a_i)\mathbf{y_0} + \sum_{i=1}^{N} a_i \mathbf{y_i} \tag{9}$$

Table 4 shows the performance improvement from optimal codebook interpolation using the three nearest codevectors.

	NB SD (dB)	HB SD (dB)
No interpolation	2.42	8.68
Optimal interpolation	2.00	8.53

Table 4. Codebook mapping performance for no interpolation and optimal codebook interpolation methods.

4. ROBUSTNESS TO INPUT DISTORTIONS

Practically, a wideband extension system has to be robust to input signal distortion such as telephone network filtering, varying input gain levels, background noise, and music input. This section describes our robust system design.

4.1. Robust Codebook Design

Three types of pre-filtering, modified IRS (M-IRS) filtering, IRS filtering and flat filtering, were applied to our codebook training database. This prevent the system from providing poor performance for a particular telephone network filtering. In addition, our training database includes eight different languages and three different gain levels (-26, -16 and -36 dBov).

4.2. Background Noise and Music Signal Handling

Since the wideband extension system is based on a speech model, it can introduce perceptual artifacts for non-speech input signals. Therefore, our system includes a two-stage detection algorithm to identify music or background noise input. In the first stage, a music detector classifies signals into two classes: music or other signals. In the second stage, a Voice Activity Detector (VAD) classifies the other signals from the first stage into two classes: speech or

background noise. If the narrowband signal is classified as background noise or music, smoothing and/or muting is applied to the artificial high-band signal to prevent undesirable high-frequency noise. The music detector and VAD output soft decision classifications, and the level of smoothing and muting is determined based on these soft decision results to minimize switching artifacts and the effect of detection errors.

4.2.1. Music

Our low-complexity music input detector exploits the different energy contour behavior between speech and music. The output of the music detector is a soft decision parameter (likelihood of music signal). For music input, the wideband extension system mutes the artificial high-band signal and modifies the narrowband input vector by shifting it towards the narrowband codebook mean vector. This indirectly shifts the high-band parameters toward the high-band mean codebook vector and avoids generating strong high-frequency sounds. The levels of muting and shifting depend on the soft decision results of the music detector. For instance, if the music detector result shows strong likelihood of music input, the system will replace the input narrowband vector with the narrowband codebook mean vector and completely mute the high-band signal.

4.2.2. Noise

Input noise segments are identified by a VAD algorithm. The VAD makes a soft decision (speech, noise or in-between) by comparing signal energy with the estimated noise energy. As is done for music input, our system modifies a noisy narrowband input vector by shifting it towards the narrowband codebook mean vector. In addition, the level of noise modulation for the high-band excitation signal is decreased depending on the VAD soft decision result [8]. This reduces busy and annoying high-frequency noise in the artificial high-band signal.

5. LISTENING EVALUATION

In addition to objective evaluation using SD, we evaluated our system with American English materials in informal listening. The following is a summary of our conclusions.

- Our wideband extension system provides a clear benefit over narrowband speech in the clean condition. It provides better intelligibility and quality of artificial wideband speech for both headphone and speaker listening environments.
- We tested our system on various background noise conditions including car, street, office, babble, interfering talker and artificial high-frequency noise. Our system provides better intelligibility and quality for background noise conditions although the enhancement is not as significant as that in the clean condition. There are very few annoying artifacts for background noise because of noise smoothing.
- The listening impression varied over different listening environment and listeners. The wideband extension enhancement is less audible with loudspeakers or less expensive headphones. Also, some listeners consistently preferred narrowband speech because it sounded less noisy to them. This indicates that a wideband extension system might need an external tuning parameter so that listeners can set up the high-frequency sound level as they want.

6. CONCLUSION

We have presented codebook mapping performance enhancements and robust algorithm design in our wideband extension system. Predictive codebook mapping and optimal codebook interpolation method were shown to provide significant SD improvement for high-band parameter estimation. Robust system design to input signal distortion allows the wideband extension system to perform well for different types of telephone network filtering, background noise and music input. Informal listening evaluation results showed that our wideband extension system improved intelligibility and quality of narrowband input signal while it might require external tuning for different listening environment and listener preference.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge valuable feedback by Laurent Le-Faucheur.

8. REFERENCES

[1] J. Epps and W. H. Holmes, "A New Technique for Wideband Enhancement of Coded Narrowband Speech," in *IEEE Workshop on Speech Coding Proceedings*, (Porvoo, Finland), pp. 174–176, 1999.

[2] N. Enbom and W. B. Kleijn, "Bandwidth Expansion of Speech Based on Vector Quantization of the Mel Frequency Cepstral Coefficients," in *IEEE Workshop on Speech Coding Proceedings*, (Porvoo, Finland), pp. 1953–1956, 1999.

[3] Y. M. Cheng, D. O'Shaughnessy and P. Mermelstein, "Statistical Recovery of Wideband Speech from Narrowband Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 544–548, 1994.

[4] P. Jax and P. Vary, "Wideband Extension of Telephone Speech Using a Hidden Markov Model," in *IEEE Workshop on Speech Coding Proceedings*, (Delavan, Wisconsin), pp. 133–135, 2000.

[5] Y. Qian and P. Kabal, "Dual-Mode Wideband Speech Recovery from Narrowband Speech," in *Proc. 8th European Conf. Speech, Commun. Tech.*, pp. 1433–1437, 2003.

[6] A. McCree, "A 14 kb/s Wideband Speech Coder with a Parametric Highband Model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Istanbul, Turkey), pp. 1153–1156, 2000.

[7] A. Rao, D. Miller, K. Rose and A. Gersho, "A Generalized VQ Method for Combined Compression and Estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Atlanta, Georgia), pp. 2032–2035, 1996.

[8] A. McCree, T. Unno, A. Anandakumar, A. Bernard and E. Paksoy, "An Embedded Adaptive Multi-Rate Wideband Speech Coder," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Salt Lake City, Utah), pp. 761–764, 2000.

[9] J. S. Collura, A. McCree, and T. E. Tremain, "Perceptually Based Distortion Measurements For Spectrum Quantization," in *IEEE Workshop on Speech Coding Telecommunications*, pp. 49–50, 1995.