WIDEBAND SPEECH RECOVERY FROM NARROWBAND SPEECH USING CLASSIFIED CODEBOOK MAPPING

Yasheng Qian & Peter Kabal

Dept. of Electrical and Computer Engineering, McGill University

ABSTRACT: Speech sounds occupy 8 kHz or more of bandwidth. However, current public telephone networks limit the speech bandwidth to 300–3400 Hz. Telephone speech is characterized by thin and muffled sounds, and degraded speaker identification. We describe an algorithm which generates the missing highband components from the narrowband speech signal. The algorithm is based on three acoustic-phonetic classified narrowband-to-wideband linear prediction (LP) spectrum mapping codebooks to recover the missing highband spectrum. Subjective tests show that the reconstructed wideband speech improves the speech quality. LP spectrum bandwidth expansion is used to avoid sharp spectral peaks. The mean SD (log-spectrum distortion) decreases by 0.93 dB, comparing to non-classified codebooks without LP bandwidth expansion.

INTRODUCTION

Current public telephone networks provide voice services for narrowband speech (300–3400 Hz). Since the human voice occupies 8 kHz or more in bandwidth, telephone speech is characterized by thin and muffled sounds. That results in a degradation of the intelligibility of speech. For example, it is very difficult to distinguish between phonemes, /s/ and /f/, because their highband components (over 3.4 kHz) are important discriminators. Most other phonemes have lost, to some extent, their fidelity in telephony speech.

In the coming 3G wireless communications systems with the advanced speech coding technology, Adaptive Multi-Rate codec (AMR) will be employed to provide wideband speech coding. However, the connection between the existing PSTN and 3G wireless systems deteriorates the quality of the service because of the bandwidth bottleneck of PSTN. A promising solution is to use a wideband recovery technique to improve the service quality of PSTN when the signal passes to a wideband network.

Several attempts have been made to address the wideband recovery issue. These methods generate an excitation signal which passes through a synthesis filter. There are two main issues for the reconstruction of the missing highband components: the reconstruction of the highband excitation and the highband spectral envelope (via the synthesis filter). One approach is to use a mapping between a codebook of narrowband spectra and a codebook of wideband spectra [Epps, 1998], [Jax, 2002]. However, since different voice sounds demonstrate a large divergence of their spectrum envelope characteristics between the narrowband and wideband speech, a single mapping codebook can result in large spectrum mapping errors. Our new approach for highband spectrum recovery is based on three acoustic-phonetic classified narrowband to highband spectrum mapping codebooks. The missing highband excitation can be regenerated by a bandpass envelope modulated Gaussian white noise with a mapping modulation gain [Qian & Kabal, 2002].

We first introduce the acoustic-phonetic classification of speech sounds. Then, we describe the wideband recovery system with three classified lowband to highband spectral envelope mapping codebooks and three excitation gain mapping codebooks. The wideband reconstruction simulation results of objective measurements for mean spectrum distortion (SD) and spectrograms are given in the last section.

ACOUSTIC PHONETICS FOR CLASSIFICATION

Acoustic-phonetics describes distinctive waveform, spectrum and power properties of speech sounds, or phonemes. We have considered two parameters, the high band (4 kHz to 8 kHz) vs. lowband (below 4 kHz) energy ratio, g_r and the voicing periodicity, or the pitch predictor gain β, to classify phonemes. The g_r indicates, the feature of the energy of the missing highband. The parameter β is a measure of the degree of the waveform periodicity or the harmonic structure. It also correlates with the highband spectrum envelope. For our application, the 42 phonemes of American English can be classified into 3 groups using the two parameters.

- Five unvoiced fricative phonemes: /s/, /f/, /θ/, /sh/, /ʃ/, the whisper, /h/ and 2 affricatives, /j/, /tʃ/. They have large highband to lowband energy ratios and no harmonics in spectrum (small pitch gain). The parameter g_r varies between 9.0 and 25.0 dB. The parameter β is in the range 0.1–0.25. A typical spectral envelope of the phoneme /s/ is shown in Fig. 1.
- Four voiced fricatives and 6 stop explosive consonant phonemes: /v/, /th/, /z/, /zh/ and /p/, /t/, /k/, /b/, /d/, /g/. Their highband to low-band energy ratio is moderate, in the range of -13.0–9.0 dB. They have, to certain degree, periodicity in the waveform. The parameter β is in the range 0.25–0.90. The voiced phonemes have β close to the upper boundary. The spectrum envelope of the phoneme /k/ is depicted in Fig. 2.

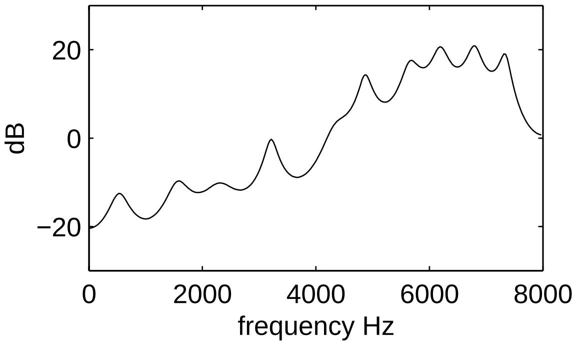

Fig. 1 The spectrum envelope of the phoneme, /s/.

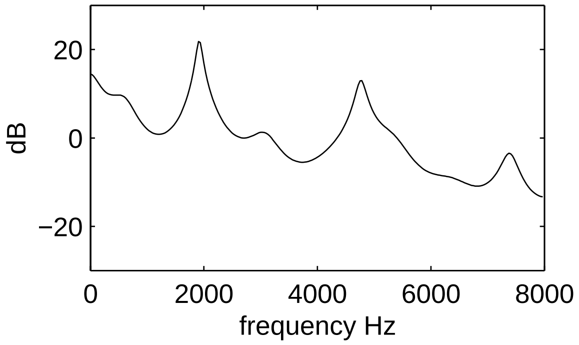

Fig. 2 The spectrum envelope of the phoneme /k/.

- The other 11 vowels, 6 diphthongs, 4 semivowels and 3 nasal consonant phonemes manifest a small lowband to highband energy ratio (g_r< –15 dB) and very good periodicity in the waveform (pitch gain $\beta > 0.9$). The phoneme /o/ spectrum envelope is illustrated in Fig. 3.

The voicing degree parameter β plays an important role in the classification. We have used the voicing periodicity parameter to approximately classify the speech frames into three groups:

$\beta \le 0.25$	unvoiced phonemes
$0.25 < \beta < 0.9$	mixed phonemes
$0.9 \le \beta$	voiced phonemes

HIGHBAND RESTORATION

The key part of a wideband reconstruction system is the highband spectrum envelope reproduction, as depicted in Fig. 4.

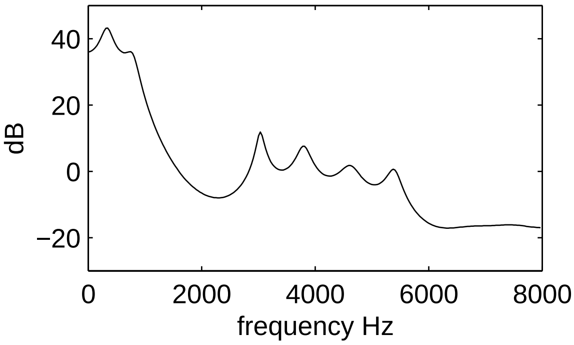

Fig. 3 The spectrum envelope of the phoneme /o/.

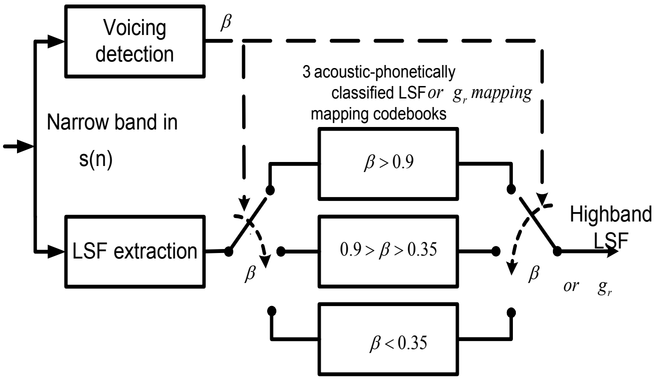

Fig. 4 The highband LSF regeneration.

The narrow band spectrum envelope parameterized by the Line-Spectrum-Frequencies (LSF) and the voicing periodicity β are extracted from the input telephone speech. Assuming the narrowband spectrum envelope LSF's are closely correlated to the highband spectrum envelope in each phoneme group, we are able to restore the missing highband spectrum envelope via a narrowband to highband LSF mapping. There are three mapping codebooks for the three different voicing periodicity β value ranges. The parameter β determines which mapping codebook is to be used. Each lowband LSF vector in the mapping codebook corresponds to a highband LSF vector. Based on the extracted lowband LSF and β of each frame, the highband spectrum envelope can be regenerated.

We have utilized the Generalized Lloyd-Max algorithm to construct three classified narrowband spectrum mapping codebooks. The iterative algorithm is carried out over three acoustic-phonetic classified narrowband spectrum training data subsets. The three classified mapping codebooks are generated by calculating both the centroids of the highband and lowband LSF vectors of the clusters of the training subsets for wideband speech. The lowband clusters are LSF vectors that have minimum distances to lowband VQ codevectors. Each vector has a time index of the training data. The highband LSF vectors are clustered using the time indices of the lowband LSF clusters. All the vectors in the clusters have the same time indices as the vectors in the lowband LSF clusters. Each lowband LSF codevector matches to a highband vector. However, the highband LSF vectors in the training data no longer minimize the mean-square error. That results in additional spectral distortion in the regenerated highband LSF's. The acoustic-phonetic classification will help the highband LSF vectors in each phoneme group to have smaller MSE. Since each group of phonemes has small average Euclidian distance of the spectrum features, that reduces the mean-square-error (MSE) of the reconstructed spectrum.

Another issue for recovery of the missing components in the highband is to recreate the excitation signal for the highband synthesis filter. The reconstruction of the excitation is shown in the Fig. 5. We have adopted a bandpass (2–3 kHz) envelope modulated Gaussian white noise as the excitation signal. The excitation signal is multiplied by a modulation gain to have an appropriate energy for the highband component. The modulation gain is defined as follows:

$$g = \sqrt{\frac{\| s_{hp}(n) \|^2}{\| s_{res}(n) \|^2}}.$$

where $s_{hp}(n)$ is the highband signal and $s_{res}(n)$ is the normalized resynthesized highband signal of a 20 ms frame. The three unquantized gains can be calculated from the three training data subsets of wideband speech. We have employed the same Lloyd-Max algorithm to train three classified modulation gain mapping codebooks for the three different β value intervals**.**

The quantized modulation gain factors are retrieved from those three modulation gain mapping codebooks. Since the modulation gain factor depends on the spectrum characteristics, it is reasonable to use three mapping codebooks for the different phoneme groups**.**

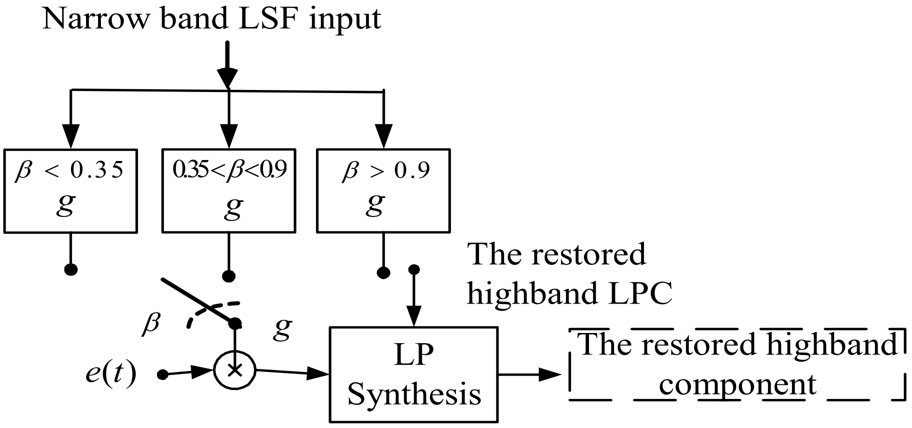

Fig. 5 The missing highband component reconstruction.

The restored highband component is combined with the narrowband input to reconstruct the wideband speech. LP analysis, particularly for high pitch values, can give narrow peaks for the LP synthesis filter in some frames. That may lead to unnatural (metallic) sounds in the resynthesized highband components. We have introduced the commonly used LPC spectrum bandwidth expansion method to avoid those artefacts. The LPC spectrum bandwidth is expanded for the LP coefficients of the highpass-filtered signal of the training data set of wideband speech in the VQ mapping codebook generation. The 10 original highband LP coefficients are expressed as $a(i)$ for $i = 0,\ldots,10$. The bandwidth expanded LP coefficients are

$$a'(i) = \gamma^i a(i).$$

The expansion factor γ = 253/256 is applied to have a fixed 60 Hz bandwidth expansion of each pole of the highband LPC spectrum. The expanded bandwidth ΔB can be calculated from [Kleijn et al, 1995]

$$\Delta \mathrm{B} = -\frac{F_S}{\pi}\log(\gamma) \quad \mathrm{Hz},$$

where $F_S = 16$ kHz is the sampling frequency of wideband speech.

EXPERIMENTAL RESULTS

We have evaluated the proposed wideband speech recovery algorithm. We have designed three lowband to highband LSF mapping vector-quantized codebooks and three corresponding modulation gain mapping codebooks using Generalized Lloyd-Max algorithm on acoustic-phonetic classified training data subsets. The wideband speech corpora taken from a Phonetic-Balanced-Sentence set (Harvard database). Fourteen LP coefficients are extracted from lowpass-filtered signal (< 4 kHz) of wideband speech. Ten LP coefficients are calculated from highpass-filtered signal (4–8 kHz). All the codebooks have the same size of 32. We have, first, measured the mean Log Spectrum Distortion (SD) in the missing highband ($f_l = 3.5$ kHz – $f_h = 7$ kHz). The definition of SD is as follows:

$$SD^2 = \frac{1}{\pi}\int_{\omega_l}^{\omega_h}\left(20\log_{10}\frac{g}{|A_{hb}(\omega)|} - 20\log_{10}\frac{g_{vq}}{|A_{hbvq}(\omega)|}\right)^2 d\omega$$

where ω_l and ω_h are the cut-off frequencies of the missing band; g and g_{vq} are the original modulation gain and the quantized gain; $|A_{hb}(\omega)|$ and $|A_{hbvq}(\omega)|$ are the magnitudes of the inverse filters of the highpass-filtered signals of wideband speech. The transfer functions of the LP synthesis filter for the highband component reconstruction are the reciprocals of the inverse filters. The SD computation excludes the lowband, because there is no distortion for the lowband component in the recovery system.

The bandwidth expansion gives the mean Log spectrum distortion (SD) reduction of 0.28 dB and the 2 dB outlier reduction of 4.9 %. The classified mapping codebooks with bandwidth expansion have brought down the SD reduction of 0.93 dB, the SD standard deviation reduction of 0.17 dB. The 2 dB outliers decrease by 18.7 %. Some segments with unexpected highband energy boosts are removed Spectrograms of the recovered wideband speech, the narrow band speech and the original wideband speech are illustrated in Fig. 6. The utterance shown is “Weave the carpet on the right hand side”. The top part shows the original wideband speech. The middle section is the recovered wideband speech from the narrowband speech (lowpass-filtered wideband speech), shown in the bottom part. The spectrograms demonstrate that most of missing highband components have been reconstructed, although they sometimes are weaker or overestimated. A typical wideband test sentences have been processed and evaluated. Informal listening shows that the proposed wideband recovery algorithm generates significantly better and more natural speech than conventional telephone speech.

CONCLUSIONS

The proposed wideband recovery method using three LSF and three gain classified lowband to highband VQ mapping codebooks provides better performance than our previous work employing a single mapping codebook. Although the new system requires more computation in training and producing the mapping codebooks than before, the system complexity increases only a little since the classification is based on voice periodicity only. After classification, the rest part of the system is the same as before.

REFERENCES

Chan, C. F. & Hui, W. (1997) *Quality enhancement of narrowband CELP coded speech via wideband harmonic resynthesis*, Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1187–1197.

Epps, J. & Holmes, W. (1998) *Speech enhancement using PSTC-based bandwidth expansion*, Proceeding of International conference on Speech Language Processing, 519–522.

Jax, P. & Vary, P. (2002) *An upper bound on the quality of artificial bandwidth extension of narrowband speech signals*, Proc. IEEE Int. Conf. on. Acoustics, Speech and Signal Processing, 237 –240.

Kleijn, W.B. *et al.* (1995) *Speech coding and synthesis, 440-441,* The Netherlands*:* Elsevier science B.V.

McCree, A. *et al.* (2001) *An embedded adaptive multi-rate wideband speech coder,* Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, 761–764.

Qian, Y. & Kabal, P. (2002) *Pseudo-wideband speech reconstruction from telephony speech,* Proc. 21st Biennial Symposium on Communications, 1187–1197

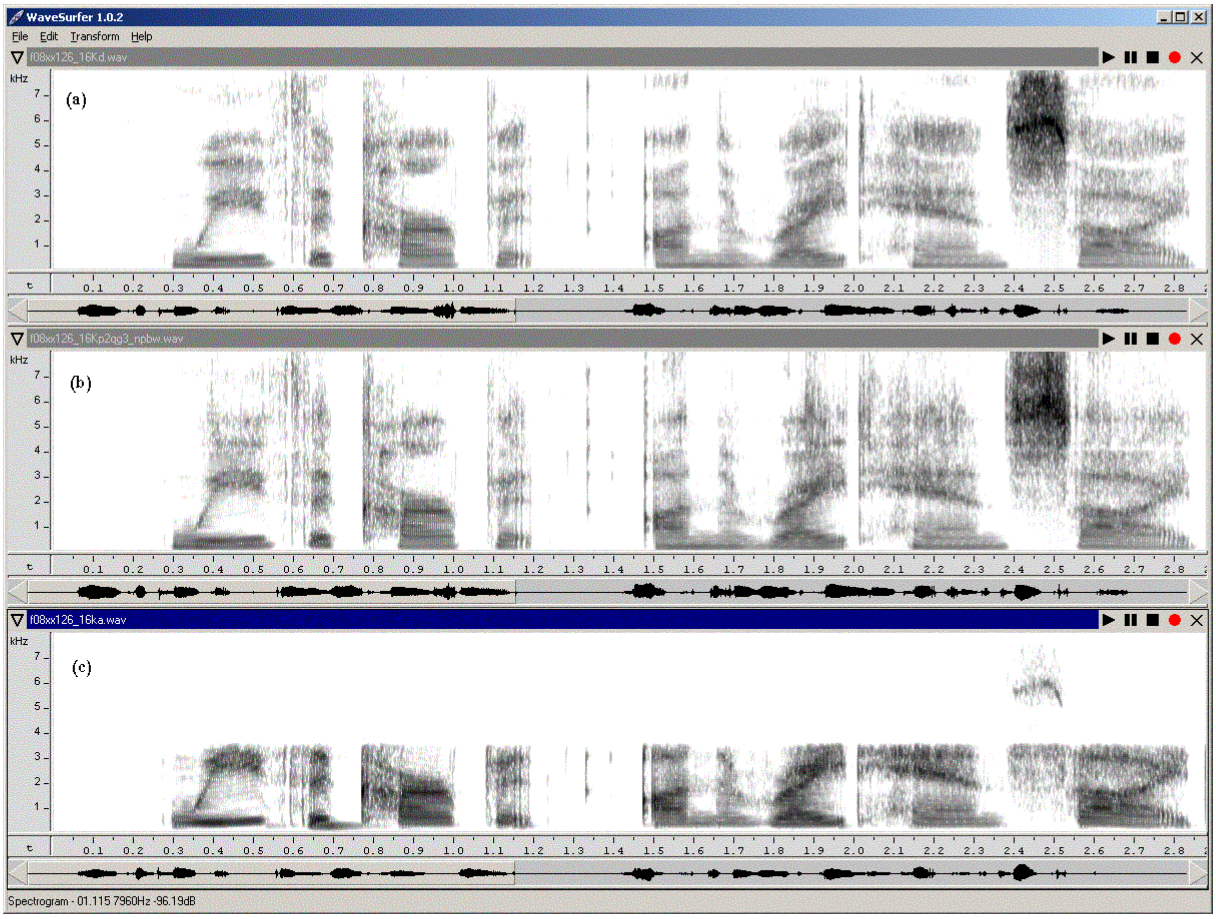

Fig. 6 Spectrograms for (a) an original wideband speech, "Weave the carpet on the right hand side", (b) the wideband recovered speech using 3 acoustic-phonetically classified codebook mapping and (c) the narrowband speech