ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

AN ALGORITHM TO RECONSTRUCT WIDEBAND SPEECH FROM NARROWBAND SPEECH BASED ON CODEBOOK MAPPING

Yuki YOSHIDA and Masanobu ABE

NTT Human Interface Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 JAPAN

Abstract

This paper proposes a new algorithm to reconstruct wideband speech from its narrow version. The algorithm has two novel points. The first is spectrum envelope reconstruction based on codebook mapping. The other is speech signal reconstruction using the reconstructed spectrum envelope. Because the algorithm makes it possible to generate high quality speech without the use of any additional transmitted information, it is applicable for any network, such as the existing telephone network, networks supporting analog and ISDN services, and so on. The algorithm is applied to 20 speakers. Evaluation by the acoustic distance measure and by listening tests confirms the good performance of the algorithm.

1 Introduction

In recent years, high quality sound has become familiar through CDs(compact disks) and LDs(laser disks). This has increased the demand for improving the sound quality of existing services. Some AM radio stations, for example, have begun to broadcast in stereo instead of monaural. These trends indicate that improved quality is one of the most important requirements in conventional systems or existing services. In terms of the telephone service, one demand is to offer wideband speech instead of narrowband speech. Because wideband speech is clear, and precisely preserves speaker identity, it makes it possible for users to communicate more realistically through telephone lines[1][2].

This paper proposes an approach to generate wideband speech from telephone speech. Because the bandwidth of the analog telephone is limited to 300Hz to 3.4kHz, the proposed algorithm generates an additional lowband signal(50Hz-300Hz) and a highband signal(3.4kHz-7.3kHz). The generation is based on two assumptions. One is that the narrowband speech correlates closely with the lowband and highband signals. The other is that even if the lowband and highband signals are not completely accurate, they significantly enhance the perceived speech quality. One advantage of the proposed algorithm is that wideband speech can be generated without adding any extra information to the speech signal. This makes it suitable for any network, such as the existing telephone network, networks supporting analog and ISDN and so on. Moreover, it is also effective when the transmission bandwidth is limited as in mobile communication.

2 The reconstruction algorithm

The proposed algorithm consists of two procedures. In the first procedure, in terms of spectrum envelope reconstruction for the highband and the lowband, a mapping function is generated using a wideband speech set and the narrowband version of the same. The mapping function is realized by codebooks that map vectors of the narrowband spectrum onto the vector space of the wideband spectrum[3]. The other procedure synthesizes the lowband and highband signals. The lowband signal is synthesized by linear predictive coding (LPC). In terms of highband signal synthesis, we consider two methods. LPC synthesis and the overlap addition of waveforms. Finally, adding the highband and lowband signals to the telephone speech yields the wideband speech. The following explains details of the two procedures.

2.1 Spectrum envelope generation

To generate a wideband spectrum envelope from a narrowband spectrum envelope, a pair of codebooks is needed. One codebook contains the wideband spectrum envelopes and the other contains the equivalent narrowband spectrum envelopes. One codevector of one codebook has a one-to-one correspondence to a codevector of the other codebook. The procedures for making the codebooks are as follows, numbers in the following explanation correspond to the block numbers in Fig. 1.

(1) Generate narrowband speech by passing wideband speech through a bandpass filter.

(2) Extract the spectrum envelopes of the wideband speech and the narrowband speech.

(3) Generate a wideband codebook using the LBG algorithm[4].

(4) Vector-quantize utterances from the wideband speech using the wideband codebook.

(5) Using the relationship of time, classify the spectrum envelopes of the utterances from the narrowband speech into clusters.

(6) Average the spectrum envelopes in each narrowband cluster, and store them as codevectors of the narrowband codebook.

When the highband signal is generated by the overlap addition of waveforms, other codebooks are necessary. The codebooks are generated as follows.

(7) Select wideband waveforms which have the closest spectrum envelope to each codevector, and store them as representative waveforms after passing them through a highpass filter and a bandpass filter.

2.2 Generation of wideband speech from narrowband speech

Figure 2 is a block diagram of the generation procedure. The procedures are as follows.

(1) Analyze input narrowband speech by LPC and extract pitch, power and spectrum envelopes.

(2) Vector-quantize each spectrum envelope using the narrowband codebook, and decode the vectors using the wideband codebook.

(3) Generate the lowband signal. Details are described later.

(4) Generate the highband signal. Details are described later.

(5) Up-sample the input narrowband speech.

(6) Add the lowband and highband signal to the output of (5), which results in the wideband speech.

The following explains the generation of lowband and highband signals. The lowband signal is synthesized using LPC synthesis. To synthesize the highband signal, two methods are proposed. One is using the LPC synthesis(method 1) and the other is using waveform information(method 2). Figure 3, 4 and 5 are block diagrams of lowband signal generation, highband signal generation(method 1 and method 2), respectively. The procedures are as follows.

- The lowband generation
  - **(1)** Synthesize the wideband speech by LPC using the analyzed pitch, power and the spectrum envelope decoded by the wideband codebook.
  - **(2)** Extract the lowband signal by passing (1) through a lowpass filter. (In this case, we use STFT Analysis/Synthesis[5] as the lowpass filter)
  - **(3)** Multiply the output of (2) by a constant, because the power of (2) is insufficient for wideband speech. This generates the lowband signal.
- The highband generation (method 1)
  - **(1)** Synthesize the wideband signal by LPC using the analyzed pitch, power and the spectrum envelope as decoded by the wideband codebook.
  - **(2)** Extract the highband signal by passing (1) through a highpass filter. (In this case, we use STFT Analysis/Synthesis as the highpass filter)
  - **(3)** Multiply the output of (2) by a cosine function to reduce the pulses caused by LPC synthesis, and normalize the power. This generates the highband signal.
- Highband generation (method 2)
  - **(1)** Referring to the codevector number, obtain two waveforms: one is from the narrowband representative codebook and the other is from the highband representative codebook.
  - **(2)** Check if the waveforms are voiced or unvoiced.

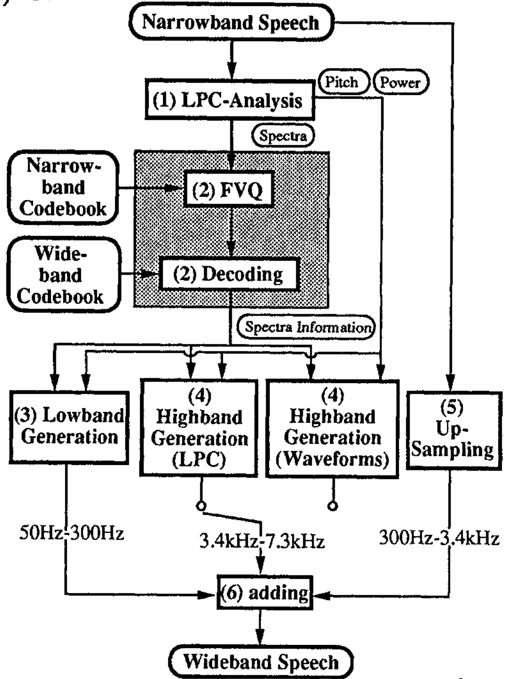

Figure 2: Block diagram of generation procedures

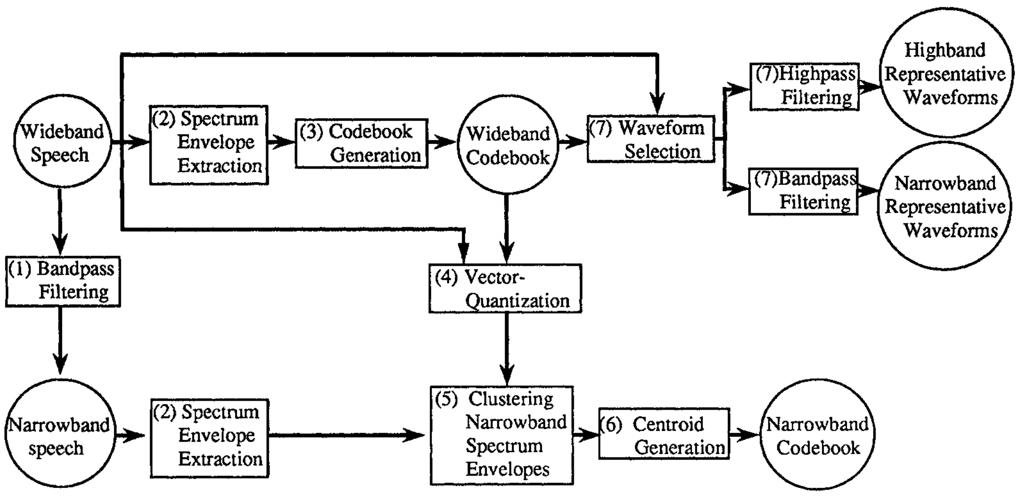

Figure 1: Block diagram of codebook generation

(3) If voiced, synthesize the narrowband speech by pitch-synchronous overlap addition. If unvoiced, synthesize the narrowband speech by frame-wise overlap addition.

(4) Calculate the power ratio between the output of (3) and the input speech.

(5) Synthesize the highband signal in the same way as (3).

(6) Multiply the output of (5) by the power ratio, to yield the highband signal.

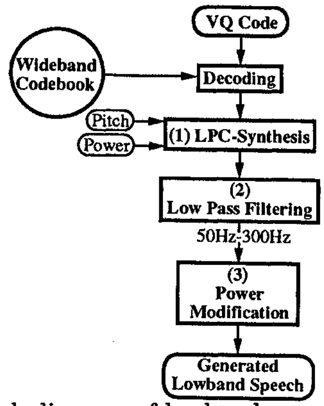

Figure 3: Block diagram of lowband speech generation

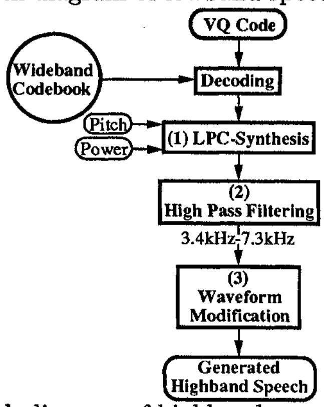

Figure 4: Block diagram of highband speech generation (method 1)

3 Performance evaluation

The performance of the proposed algorithm was evaluated by spectrum distortion and listening tests. The conditions of the experiments are shown in Table 1.

The term "speaker dependent" indicates that the speaker of the input speech also input the training speech. "Speaker independent" is used when the two speakers are different.

3.1 Evaluation by spectrum distortion

The spectrum distortion was measured by using the wideband codebook and the narrowband codebook. We used 10 male speakers and 10 female speakers.

VQ distortion was calculated as follows.

(1) Extract the spectrum envelope from the wideband speech.

(2) Vector-quantize (1) by using the wideband codebook.

(3) Calculate separately the squared error between (1) and (2) for the lowband and highband signals. The error is defined as follows.

$$D = \sum_{t=0}^{T'}\left[\frac{1}{2\pi}\int_a^b\left[10\log_{10}\frac{\hat{Y}_t(\omega)}{Y_t(\omega)}\right]^2 d\omega\right]^{\frac{1}{2}}$$

Reconstruction distortion is calculated as follows.

(4) Obtain the narrowband speech by filtering the speech used in (1) and extract narrowband spectrum envelopes.

(5) Reconstruct the wideband spectrum envelope that corresponds to the output of (4) using the narrowband codebook and the wideband codebook.

(6) Calculate separately the squared error between (4) and (5) for lowband and highband signals. The error is defined as in (3).

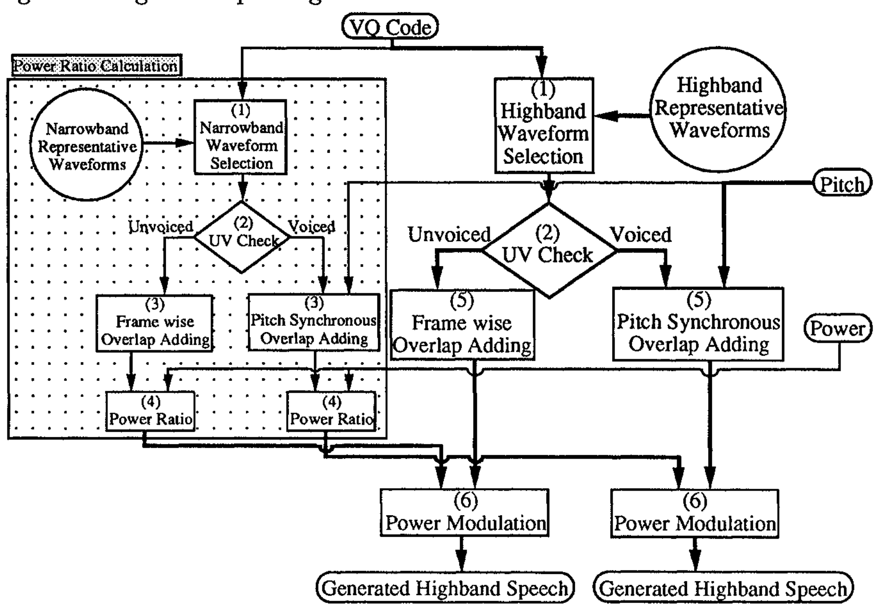

Figure 5: Block diagram of highband speech generation (method 2)

Experimental results are shown in Fig. 6 and 7. Each distortion value is the average for all speaker pairs. Judging from the results, the proposed algorithm can reconstruct the lowband spectrum as accurately as vector quantization and reconstruction distortion is decreased as codebook size increases (an 8bit codebook causes 3.5dB of spectrum distortion). In terms of highband reconstruction, the decrease in reconstruction distortion saturates at 6.5dB with a 4bit codebook. This shows that the correlation between the highband signal and narrowband speech is not as high as that of between the lowband signal and narrowband speech.

Table 1: Test Conditions

The number of training data	186 words balanced all phonemes
Analysis window	hamming
Window length	21msec
Shift length	3msec
LPC order	14
The number of FFT point	512
Distance measure	Euclid distance of LPC cepstrum

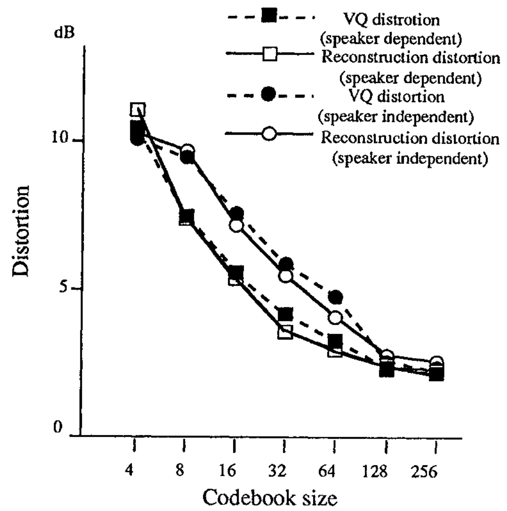

Figure 6: Spectrum distortion (lowband)

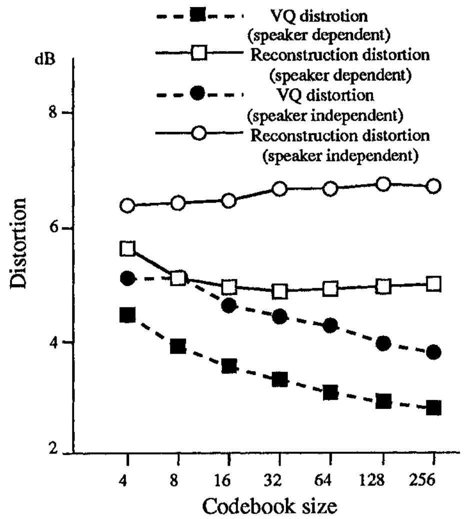

Figure 7: Spectrum distortion (highband)

3.2 Evaluation by listening tests

Pair comparison listening tests were carried out. Based on the above results, wideband speech was generated using 8bit and 4bit codebooks for lowband and highband signals, respectively. Two male and two female speakers were employed in a speaker independent manner. At random, two speech sets were picked from three speech sets: telephone speech, speech1 generated using method 1 and speech2 generated using method 2. Six listeners were asked to select one speech as being wider than the other. Seventy seven pairs were used.

Experimental results are shown in Fig. 8. Judging from the results, we conclude that the proposed algorithm can effectively reconstruct wideband speech from telephone speech. In terms of the best method for highband generation, there is no significant difference between them.

4 Conclusion

We proposed an algorithm to generate wideband speech from narrowband speech using codebook mapping, and confirmed the performance of the algorithm. Judging from spectrum distortion, no performance difference exists between speaker dependent and independent reconstruction. Listening tests confirmed that the quality of generated wideband speech was better than the original telephone speech. We have a plan to improve the highband signal generation process to improve the correlation to the narrowband signal.

Acknowledgment

We are grateful to the members of the Speech Processing Department for their helpful discussions. We also thank Dr. Kitawaki, director of the Speech and Acoustic Labs., and Dr. Sugamura, leader of the Speech Processing Group, for their continuous support of this work.

References

[1] Y. Cheng, D. O'Shaughnessy, P. Mermelstein, *"Statistical Recovery of Wideband Speech From Narrowband Speech,"* Proceedings of ICSLP92, pp.1577-1480, 1992

[2] N. Jayant, *"High-Quality Coding of Telephone Speech and Wideband Audio,"* Advances in Speech Signal Processing, pp.85-108, 1992

[3] M. Abe, S. Nakamura, K. Shikano, H. Kuwabara, *"Voice conversion through vector quantization,"* ICASSP'88, pp.655-658, 1988

[4] Y. Linde, A. Buzo, and R. M. Gray, *"An algorithm for vector quantizer design,"* IEEE Trans. Commun., COM-28,1,pp.84-95 (Jan. 1980).

[5] Lawrence R. Rabiner, Ronald W. Schafer : "Digital Processing of Speech Signals".

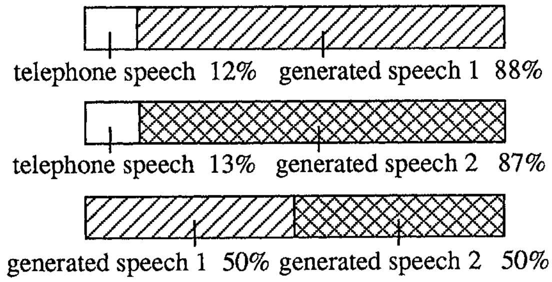

Figure 8: Preference score