GENERATION OF BROADBAND SPEECH FROM NARROWBAND SPEECH USING PIECEWISE LINEAR MAPPING

Y. Nakatoh, M. Tsushima, and T. Norimatsu
Multimedia Development Center
Matsushita Electric Industrial Co., Ltd.
1006 Kadoma, Kadoma-shi, Osaka, 571 JAPAN.
Tel. +81 6 906 4552, FAX: +81 6 908 6802, E-mail: nakatoh@arl.drl.mei.co.jp

ABSTRACT

This paper proposes a recovery method of broadband speech form narrowband speech based on piecewise linear mapping. In this method, narrowband spectrum envelope of input speech is transformed to broadband spectrum envelope using linearly transformed matrices which are associated with several spectrum spaces. These matrices were estimated by speech training data, so as to minimize the mean square error between the transformed and the original spectra. This algorithm is compared the following other methods, (1)the codebook mapping, (2)the neural network. Through the evaluation by the spectral distance measure, it was found that the proposed method achieved a lower spectral distortion than the other methods. Perceptual experiments indicates a good performance for the reconstructed broadband speech.

1. INTRODUCTION

Most telephone networks carry only narrowband speech for economic and historical reasons. For example, on the analog telephone network and the mobile communication system, bandwidth of the speech is limited from 300Hz to 3.4kHz. So it is very important to improve speech quality on telephone communication. One solution is to generate broadband speech from narrowband speech. Because broadband speech is clear and natural, it makes it possible for users to communicate more realistically through telephone lines.

Recently, several studies about band expansion of the band-limited speech were proposed. For example: the statistical method[1] indicates a good performance but needs complicated calculation. The codebook mapping method[2] and the neural network method [3] can be realized comparatively. And other simply methods were proposed.

This paper proposes an new algorithm to generate broadband speech from narrowband one. This algorithm comprises two novel parts of the recovery process. In the first part, the broadband spectrum envelope is reconstructed using the piecewise linear mapping. In the other part, the residual wave with broadband spectrum is generated by the non-linear computation. Finally, the broadband speech is synthesized from the reconstructed spectrum envelope and the generated residual wave. Perceptual experiments indicates a good performance for the reconstructed broadband speech and Almost everybody replied that the reconstructed speech was broader than the original narrowband speech.

2. SPECTRUM CONVERSION ALGORITHM

Given that there is a correlation between the band-limited spectrum (narrowband speech) and original broadband spectrum, we assumed the broadband spectrum envelope can estimate from narrowband spectrum envelope by piecewise linear mapping. In this method, narrowband spectrum envelopes are partitioned into the narrowband spectrum space using clustering algorithm. The partitioned subspace is corresponded VQ code or phoneme like a /a/ or /s/. By the same way, broadband spectrum envelopes are partitioned. There is the relationship of one to one between narrowband subspace and broadband one. So transformed matrices can be estimated by each subspace, and each transformed spectrum envelope is made by a set of mapping matrices which are associated with the fuzzy partitioned spaces.

2.1. Piecewise Linear Mapping method

The point of this algorithm consist in that spectrum envelope is converted by linearly transformed matrix. In proposed method, narrowband spectrum envelope and broadband one is made from the same training data and narrowband codebook is generated from narrowband spectra. First, Input spectrum envelope x_i, which is narrowband spectrum envelope, is Vector-Quantized by narrowband codebook $\{V_k\}$. Transformed spectrum envelope of each subspace $\{\Omega_k\}$ is estimated using matrices $\{A_k\}$. Finally, broadband spectrum envelope $\hat{y}_i$ is mapped onto the target spectral space by a weighted sum of transformed spectrum envelopes using eq. (1).

$$\hat{y}_i = \sum_{k=1}^{M} w_{ik} A_k x_i \qquad (1)$$

$$w_{ik} = \frac{\left\| x_i - V_k \right\|^{-p}}{\sum_{r=1}^{M} \left\| x_i - V_r \right\|^{-p}} \qquad (2)$$

Here, the subspace $\{\Omega_k\}$ is divided by codebook $\{V_k\}$ and the p in eq. (2) is parameter for spectrum smoothing.

2.2. Training of transformed matrices

The transformed matrices $\{A_k\}$ were estimated by the training data, so as to minimize the mean square error between the transformed $\{\ddot{y}_i\}$ and original spectra $\{y_i\}$ using eq. (3).

$$J(\{A_k\}) = \sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2 \qquad (3)$$

Fig. 1 is a flowchart of this training algorithm. First, narrowband spectrums and broadband one are made by many training data. Secondly, narrowband codebook is generated from narrowband spectrums, and a codeword are found that narrowband spectrum is vector-quantized by narrowband codebook. Here, a pair of broadband and narrowband spectrum are grouped using the relationship of one to one in the same frame. Finally, the transformed matrices were estimated by each subspace.

2.3. Other Methods of Spectrum Conversion

In the experiment, we compare between our method and the following other methods: (1)the codebook mapping method, (2)the neural network method. In the codebook mapping method, narrowband spectrum is converted to the broadband spectrum using the relationship between the narrowband codebook and the broadband codebook. First, input narrowband spectrum is Fuzzy-Vector-Quantized by narrowband codebook and next, the transformed spectrum is Fuzzy-Decoded using both menbership function and codeword which are found by Fuzzy-VQ. Each codebooks are made by training data beforehand. In the neural network method, the narrowband spectrum is converted to the broadband spectrum by 4 layers neural network. The network structure is estimated using back-propagation algorithm by the training data. Neural network is nonliner conversion, so it show a good performance.

3. BAND EXPANSION SYSTEM

3.1. Whole Block Diagram

Fig.2 is the whole block diagram of our system. In this figure, an input speech is divided to the spectrum envelope and the residual wave using the LPC analysis method. This algorithm comprises two parts for the reconstruction. In the first part, broadband spectrum envelope is reconstructed. And in the other part, residual wave with broadband spectrum is generated. And the synthesized speech is made by this envelope and this residual wave using the LPC synthesis method. The synthesized speech is broadband one, but has a little distortion due to synthesis. So the synthesized speech is smoothed by smoothing algorithm on wave domain, and separated into low frequency part, this is 50-300Hz, and high frequency part, this is 3.4-7.4kHz. Finally, the broadband speech is made by the original narrowband speech and separated two parts.

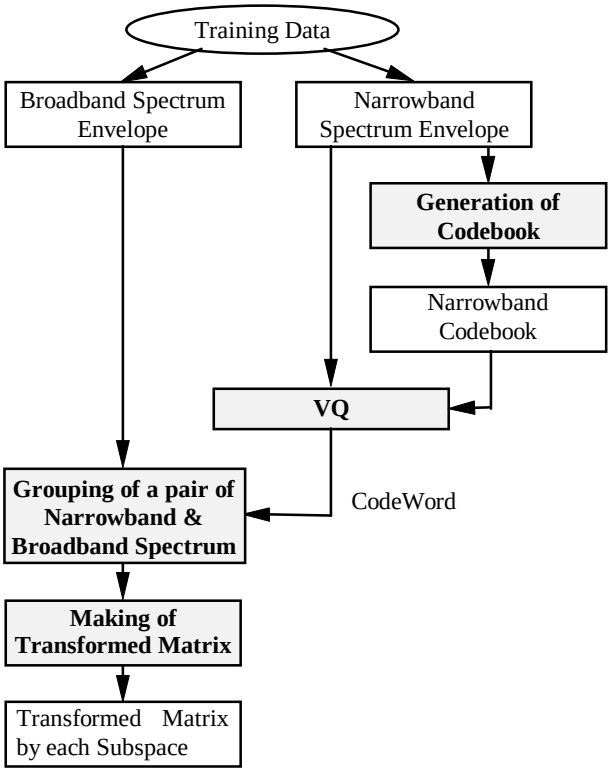


Fig. 1 Flowchart of training algorithm

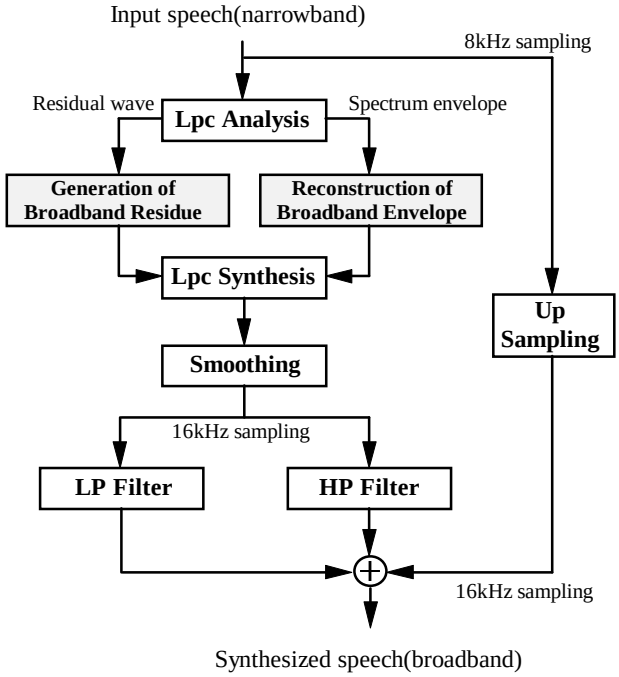


Fig. 2 Whole block diagram of our system

3.2. Generation of Broadband Residue

The residue with broadband spectrum is generated from the residue of input speech. First, the residue of input speech is up-sampled and full-rectificated. We use the full-rectification in our system, because it is possible that the lowband speech(under 300Hz) and the highband speech(over 3.4kHz) are generated with frequency the fine structure keeping. Next, rectificated residue is smoothed on frequency domain by low order LPC analysis . Finally, the smoothed residue is gain-controled for synthesis.

4. EXPERIMENTS

4.1. Fundamental Performance

In experiments, we have investigated the fundamental performance of our proposed method, through the evaluation by the spectrum distortion based on eq. (4).

$$D = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{W}\int_{0}^{W}\left\{S(f)-\hat{S}(f)\right\}^{2}df} \qquad (4)$$

In eq. (4), $S(f)$ is target broadband envelope and $\hat{S}(f)$ is transformed one. N is number of frame for distance calculation. We used 212 Japanese words which 10 male speakers and 10 female speakers spoken. No.1-100 words on each speaker are used for training and No.101-130 words on each speaker are used for evaluation. Table 1 is analysis conditions. Narrowband speech is obtained with broadband speech filtered through the telephone network.

Fig.3 shows the relation between the number of subspaces and the spectrum distortion. Spectrum distortion decrease as the number of subspaces increase. Fig.4 shows the effect of interpolation parameter. These result were that we obtained 3.57dB as the minimum spectrum distortion, so we determined that the optimal number of subspaces is 16 and the optimal value of interpolation parameter is 0.5. Fig.5 shows a example of the spectrum sequence which one male spoken, this is Japanese /ISHIGUMI/, before and after conversion. In this figure, left figure(a) is narrowband, medium one(b) is converted, and right one(c) is broadband equence. At example of converted speech, we can find the spectrum less than 300 Hz and more than 3400 Hz reconstruct well.

4.2. Comparison with Other methods

The next experiment is comparison between the proposed method(we call LM) and the following other methods; (1)the codebook mapping method(CM), (2)the neural network method(NN). Table 2 is experimental condition by each method and we decided it so as to the same condition. The following results denotes spectrum distortion at low frequency band (under 300Hz) and high frequency band (over 3.4kHz). Table 3 is the result on speaker dependent experiment and Table 4 is the result on speaker independent experiment. We have found that

Table 1 Analysis conditions

Sampling frequency	16[kHz] at broadband 8[kHz] at narrowband
Window length	20[ms] Hamming
Window shift	10[ms]
Preemphasis	$1-0.98z^{-1}$
LPC analysis order	16
Parameter	16 (Cepstrum)

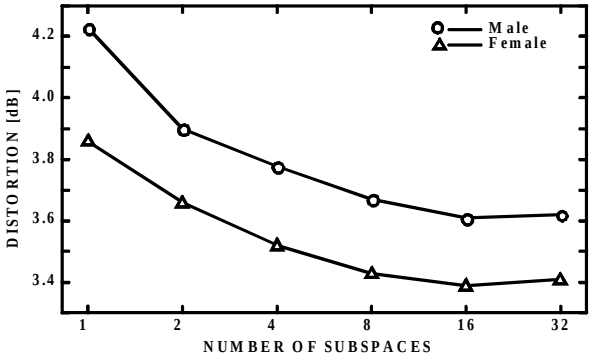


Fig. 3 Effect of number of subspaces

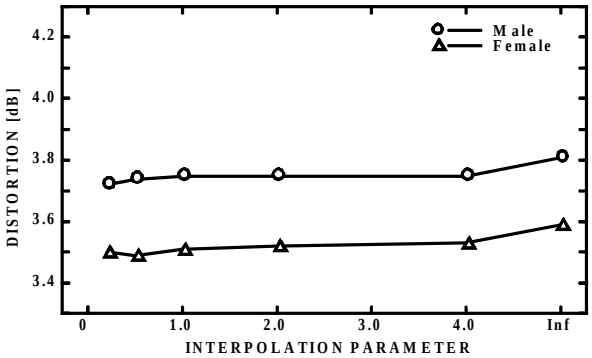
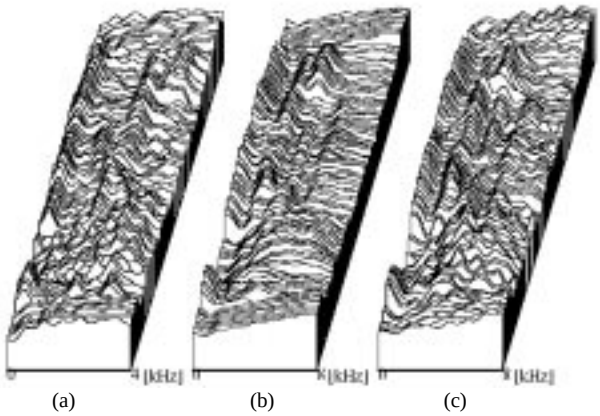


Fig. 4 Effect of interpolation parameter



(a) (b) (c)

Fig. 5 Spectrum comparison before and after conversion (a)input narrowband spectrum, (b)converted spectrum and (c) target broadband spectrum

Table 2 Experimental condition

LM	Number of subspaces	16
	Interpolation parameter	0.5
CM	Codebook size	256
	Fuzzyness	1.25
NN	Number of Layers	4
	Number of neurons	15-45-45-15

Table 3 Spectrum distortion on speaker dependent

	Training data		Evaluation data	
	lowband	highband	lowband	highband
LM	3.62	4.21	3.79	4.41
CM	3.73	4.24	3.94	4.63
NN	3.36	3.82	3.80	4.22

Table 4 Spectrum distortion on speaker independent

	lowband	highband
LM	5.08	4.62
CM	5.16	4.68
NN	5.31	4.71

Table 5 Spectrum distortion on each S/N

	S/N10dB		S/N20dB	
	lowband	highband	lowband	highband
LM	8.27	6.44	6.02	5.18
CM	8.39	6.58	6.12	5.12
NN	8.94	6.92	6.38	5.28

the spectrum distortion of NN method is lower than the other methods on speaker dependent experiment. But on speaker independent and noisy environment experiment the proposed method achieves a lower spectrum distortion than the other one, especially, at low frequency band.

4.3. Subjective Experiments

Next, we present results of subjective experiment. In our tests, test data are 10 Japanese sentences spoken by one speaker. We made one-pair comparison tests to 12 listeners(including acoustic specialists). We presented a couple of sources at random to all listeners. Listeners decide 7 levels comparison scores after listing a couple of sources, which are narrowband speech with telephone bandwidth and its transformed speech. Test items are "Which source do you feel broadband ?" and "Which source do you feel high quality ?". Fig 6(a) is result for the first question and Fig 6(b) is result for the second question. In Fig 6(a), almost all the listeners feel our transformed source was broader than the other. This result shows the probability of our method reconstructing broadband speech from narrowband one. In Fig 6(b), experimental result presents separate one. We infer that the listeners who checked superior to transformed speech feel more natural about its bandwidth. On the other hand, we infer that the listeners who checked superior to

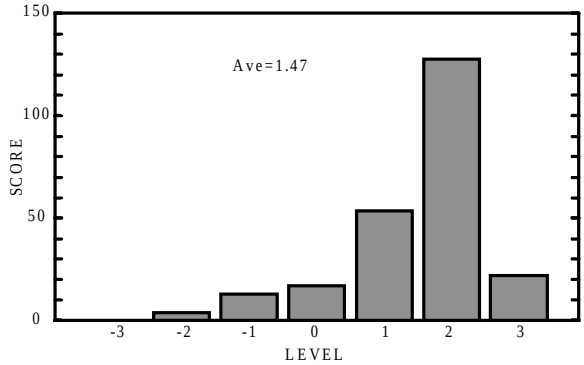


Fig. 6(a) Preference score about bandwidth

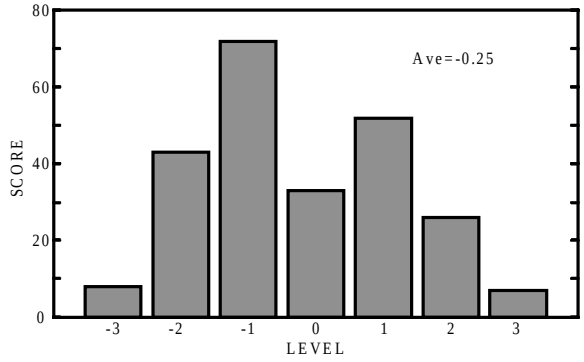


Fig. 6(b) Preference score about speech quality

original narrowband speech feel transformed speech is noisy by synthesis.

5. CONCLUSION

We proposed an algorithm to reconstruct broadband speech using piecewise linear mapping. We have found that our proposed method achieve a lower spectrum distortion than the other one. And the subjective experiment have indicated a good performance for the reconstructed broadband speech such that the reconstructed speech is broader than the original narrowband one. In the future, I will study improvement of our spectrum conversion algorithm and comparison of other residual generation methods.

REFERENCES

[1] Y. C. Cheng, D. O'Shaugnessy and P. Mermelstein, "*Statistical recovery of wideband speech from narrowband speech"*, ICSLP 92, pp. 1577-1580, 1992.
[2] Y. Yoshida and M. Abe, "*An algorithm to reconstruct wideband speech form narrowband speech based on codebook mapping"*, ICSLP 94, pp. 1591-1594, 1994.
[3] Y. Tanaka and N. Hatazoe, "*Reconstruction wideband speech form telephone-band speech by multi-layer neural networks"*, Spring Meeting of ASJ, 1-4-19, pp. 255-256, 1995.