Frequency Mapping for Multi-band Frequency Compression for Improving Speech Intelligibility

P. N. Kulkarni and P. C. Pandey
SPI Lab, Department of Electrical Engineering
Indian Institute of Technology Bombay
Powai, Mumbai 400 076, India
E-mail:{pnkulkarni, pcpandey}@ee.iitb.ac.in

***Abstract*- As a result of increased spectral masking, elderly persons with sensorineural hearing impairment and persons using mobile phones under adverse listening conditions have difficulty in speech perception. Multi-band frequency compression can be used for reducing the effect of spectral masking. The speech spectrum is divided into a number of analysis bands and the frequency components in each of these bands are compressed towards its center, by a constant compression factor. We have investigated three frequency mappings for this processing: (i) sample-to-sample mapping, (ii) superimposition of spectral samples, and (iii) spectral segment mapping. Evaluation of these schemes through listening tests showed that spectral segment mapping achieves desired compression by retaining the spectral distribution of energy in the signal.**

I. INTRODUCTION

Sensorineural hearing loss occurs when the functioning of the cochlea is affected or when there is dysfunction of the auditory nerve or higher centers in the auditory pathway. In cases of sensorineural hearing loss, the auditory filter bandwidth generally increases and frequency selectivity gets reduced due to increased spectral masking [1] - [5]. The perceptual consequences of loss of frequency selectivity due to widened auditory filters are greater susceptibility to masking, and difficulty in separating two or more simultaneously presented sounds. Hence persons with sensorineural hearing loss have difficulty in speech perception. Normal hearing persons face similar difficulty under adverse listening conditions, e.g. using a mobile phone in a noisy environment.

Earlier studies [6],[7] have shown that binaural dichotic presentation, using critical bandwidth based spectral splitting with perceptually balanced comb filters, helps in reducing the effect of spectral masking for persons with moderate bilateral sensorineural hearing impairment. However, the scheme is useful only for binaural hearing. For monaural hearing, multi-band frequency compression can be used for reducing the effect of spectral masking. In this technique, the speech spectrum is divided into a number of analysis bands, and frequency components in each of these bands are compressed towards its center by a constant compression factor. Thus, the speech energy is presented to the impaired ear over relatively narrow bands [8].

The signal processing technique employed by Reed *et al* [9], was aimed at compressing the speech spectrum to fit into the reduced hearing range of the impaired ear. The technique involved four steps: segmentation, warping, dilation and time aliasing, and resynthesis. The speech signal was segmented into small intervals, equal to pitch period for voiced sounds and of a fixed length for unvoiced sounds. Each segment was subjected to time-variant warping operation, transforming the frequency axis nonuniformally. The resulting segments were dilated in time to lower the warped spectrum, and time aliased to compensate for the prolongation associated with warping and dilation. Since the segments had precisely the same duration as the unprocessed segments, resynthesis was done by simply adding successive processed segments.

The frequency compression technique, as reported in [8], [10], is based on auditory critical bands. In this technique, speech signal was compressed towards the center of each critical band along the frequency axis. First the input speech was divided into frames by a Hamming window. Next, FFT was computed on each frame to obtain amplitude and phase spectra. Amplitude spectrum was then compressed towards the center of each critical band along the frequency axis. Compression in the range of 10% to 90% was used. Amplitude spectrum, after piece-wise frequency compression, was then combined with original phase spectrum. Finally, resynthesized signal is obtained using overlap-add method. Listening tests were conducted on impaired subjects, with fifty VCV syllables uttered by a male speaker were taken as test material. A modest improvement in the recognition score (38.3% for the processed set as against 35.4% for the unprocessed set) has been reported.

Multi-band frequency compression concentrates spectral energy towards the band centers in order to partly compensate for the increased frequency masking. The quality and intelligibility of speech signal obtained after multi-band frequency compression depends on the frequency mapping employed. The objective of present investigation is to select the most appropriate scheme of frequency mapping. Three different frequency mapping schemes are investigated: sample-to-sample mapping, spectral sample superimposition, and spectral segment mapping.

II. Signal Processing

Multi-band frequency compression involves three steps: (1) segmentation and spectral analysis, (2) spectral modification, and (3) resynthesis. A set of frequency components (complex values) falling in each of the predefined frequency bands, are compressed towards the center of the band, by a compression factor. In our implementation, the input speech signal, sampled at 10 k samples/s, is divided into segments of length 200 samples (20 ms) with 50% overlap. Each windowed speech segment is zero padded to the length of N and then N-point FFT is computed on it. Frequency scale of the spectrum is then divided into eighteen analysis bands based on auditory critical bands [11] as shown in Table 1. The frequency components falling in each of these bands are compressed by a constant compression factor towards the center of the corresponding critical band. The modified complex spectrum is converted back to time domain by N-point IFFT, and modified speech is resynthesized by overlap-add method [12], [13]. Investigations involving different FFT sizes showed that N = 1024 was adequate for various compression factors and hence 1024-point FFT was used. Since the frequency samples in each band are compressed towards the center of the band by a constant factor, by using same compression factor for all the bands, the processing preserves the harmonic structure in case of voiced speech and randomness in case of unvoiced speech.

TABLE 1

LIST OF CRITICAL BANDS ALONG WITH THEIR CENTER FREQUENCIES

Critical band	Center frequency in kHz	Frequency range in kHz
1	0.13	0.01 – 0.20
2	0.25	0.20 – 0.30
3	0.35	0.30 – 0.40
4	0.45	0.40 – 0.51
5	0.57	0.51 – 0.63
6	0.70	0.63 – 0.77
7	0.84	0.77 – 0.92
8	1.00	0.92 – 1.08
9	1.17	1.08 – 1.27
10	1.37	1.27 – 1.48
11	1.60	1.48 – 1.72
12	1.86	1.72 – 2.00
13	2.16	2.00 – 2.32
14	2.51	2.32 – 2.70
15	2.92	2.70 – 3.15
16	3.42	3.15 – 3.70
17	4.05	3.70 – 4.40
18	4.70	4.40 – 5.00

A. Sample-to-Sample Mapping

The relationship between compressed spectrum Y and the original spectrum X for frequency samples k falling in the i^{th} analysis band is given as $Y(k') = X(k)$, with

$$k' = k_{ic} + \mathrm{round}\left(\alpha\left(k - k_{ic}\right)\right) \tag{1}$$

where, α is the compression factor, and k_{ic} is the center frequency of the i^{th} analysis band, and is given by

$$k_{ic} = 0.5\left(k_{is} + k_{ie}\right) \tag{2}$$

where k_{is} and k_{ie} are the starting and ending indices for the i^{th} band. The frequency mapping is illustrated in Fig. 1 for two of the analysis bands i.e. 0.2 kHz – 0.3 kHz and 0.3 kHz-0.4 kHz, with the center frequency of 0.25 kHz and 0.35 kHz respectively. For f_s of 10 k samples/s, and for 1024-point FFT, frequency samples from 21 to 31 with 26 as a center point will fall in the first band and samples from 32 to 41 with 36 as a center point will fall in the second band.

In this mapping, if two or more frequency components are mapped on to the same point, then only the one with the largest index amongst them will be retained. As a result, some of the frequency components in the unprocessed spectrum may not contribute to the compressed spectrum, resulting in irregular variation in the spectrum and signal energy.

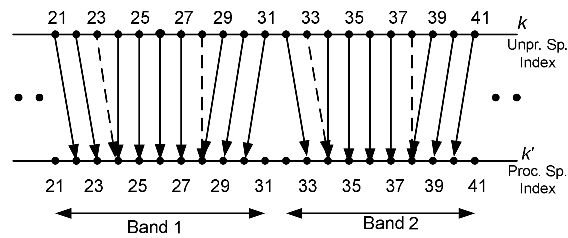


Fig.1 Sample-to-sample mapping

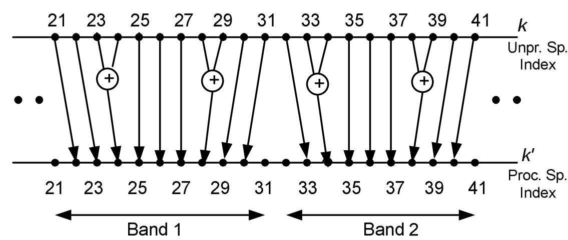


Fig.2 Superimposition of spectral samples

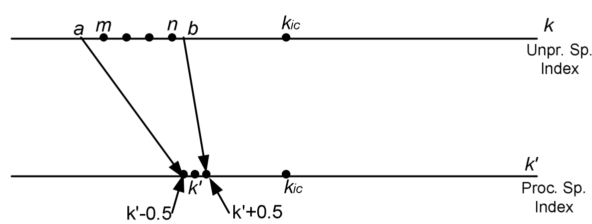


Fig. 3 Spectral segment mapping

B. Spectral Sample Superimposition

This mapping addresses the problem of missing components in the earlier mapping, by adding the spectral samples which map to the same index.

Figure 2 illustrates this frequency mapping. Reduction in the energy of the processed signal, as observed earlier, got partly compensated. However, variation in the number of spectral samples contributing to the mapping cause some irregular variation in the spectrum of resynthesized speech.

C. Spectral Segment Mapping

The objective of this mapping is to achieve the frequency compression without irregular variation in the spectrum. Let *k'* be an FFT index on the compressed spectrum $Y(k')$. We need to find the total contribution at *k'*, on the compressed frequency scale, from the unprocessed spectrum $X(k)$. As shown in Fig. 3, the spectral segment from *a* to *b* in the unprocessed spectrum, contributes for *k'* on the compressed scale. The values of *a* and *b* are given as

$$a = k_{ic} - \left[\left(k_{ic} - \left(k' - 0.5 \right) \right) / \alpha \right] \quad (3)$$

$$b = a + 1/\alpha \quad (4)$$

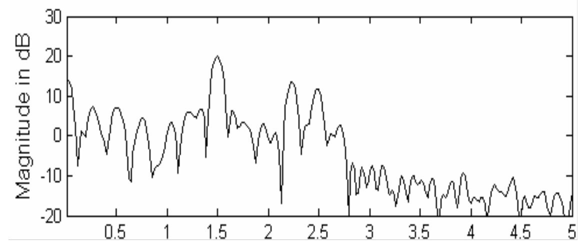
where, α is the compression factor, and k_{ic} is the center frequency of the i^{th} critical band. Let *m* and *n* be the FFT indices of the first and last spectral samples, respectively, falling in the segment from *a* to *b*. Index *m* is lower integer higher than *a* and *n* is highest integer lower than *b*. The processed spectrum is then given by,

$$Y(k') = (m-a)\, X(m) + \sum_{j=m+1}^{n-1} X(j) + (b-n)\, X(n) \quad (5)$$

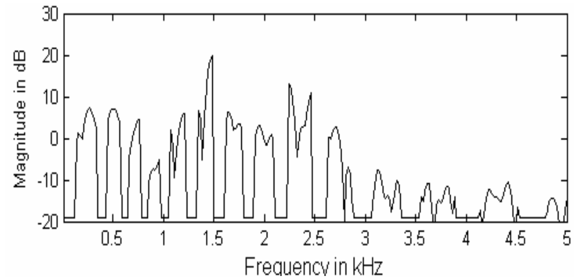
Unlike in previous two mapping schemes, entire spectral segment in the original spectrum contributes to the compressed spectrum. As a result, this scheme achieves desired frequency compression retaining the broad spectral distribution and without introducing irregular variation.

III. Tests and Results

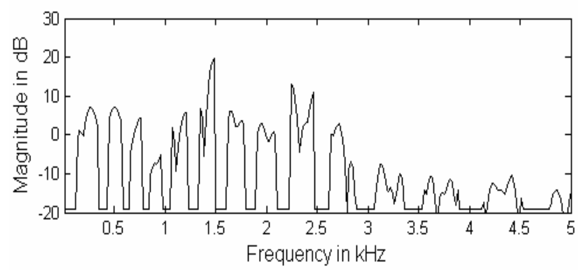
Figure 4 shows the spectra of the vowel /a/, processed using the three frequency mapping schemes, for compression factor of 0.6. It can be observed from the figure that the reduction in the energy over the speech spectrum is more in case of sample-to-sample mapping as compared to other two mappings. Output from segment mapping has least spectral distortion without much reduction in the signal energy. Figure 5 shows the narrow band spectrograms of vowel sequence /*a i u*/, processed using the three mappings. In all the three mappings, harmonic structure is retained with a moderate shift in the formants. Figure 6 shows the wide band spectrogram for the syllable /aka/, processed using segment mapping. It can be observed that formant transitions are retained, with moderate shifts (less than 1/3 octave) in the formant location. Harmonic structure is preserved in the form of vertical striations. Figure 7 shows the spectrogram for processed and unprocessed signal for "*we were away a year ago*".



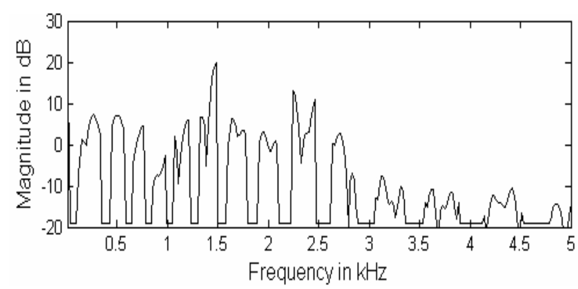
(a) Unprocessed



(b) Sample-to-sample mapping

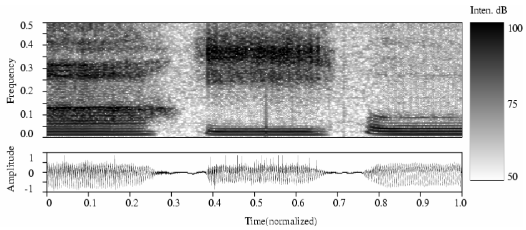


(c) Superimposition of samples

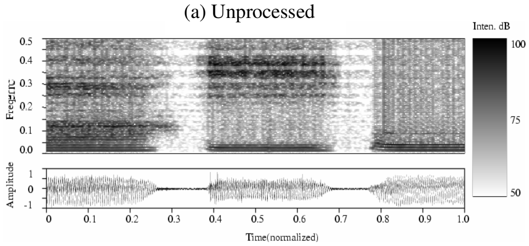


(d) Segment mapping

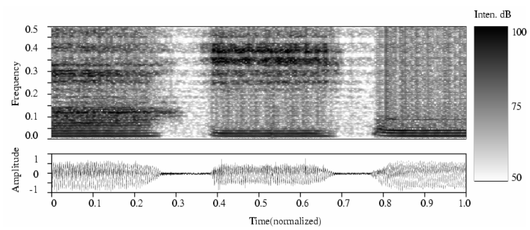
Fig. 4 Spectra of vowel /*a*/: unprocessed and processed using the three mappings for multi-band compression. Compression factor = 0.6, segment length = 20 ms, S.R. = 10 k sa/s.



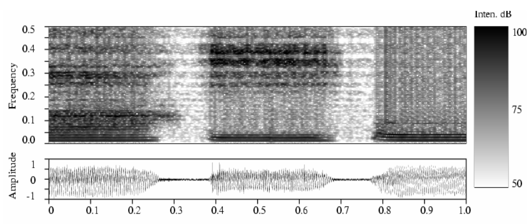
(a) Unprocessed



(b) Sample-to-sample mapping



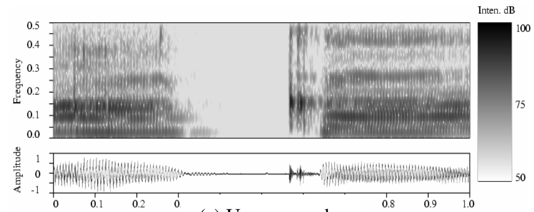
(c) Superimposition of samples



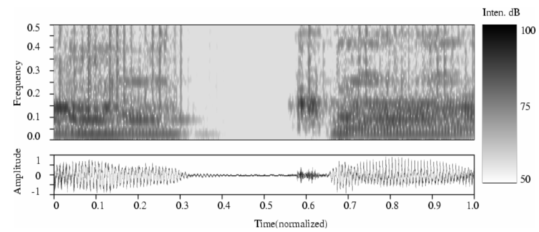
(d) Segment mapping

Fig. 5 Narrow band spectrograms ($\Delta f = 45$ Hz) of /a i u/: unprocessed and frequency compressed with the three frequency mapping schemes. Compression factor = 0.6

Frequency compression scheme was evaluated with the three frequency mappings through listening tests. These tests were conducted on six normal-hearing subjects, at various compression factors, to assess the perceived distortion. Test material included sustained vowels /*a i u*/, a sentence, and a long passage, recorded at 10 k samples/s with 16-bit resolution. The tests indicated that, perceived distortion was not significant in all the three mappings. A reduction in the perceived loudness was observed for sample-to-sample mapping while this reduction was partly restored in mapping by superimposition of samples. There was no change in the perceived loudness in segment mapping. Subjects rated the quality of the processed speech using segment mapping better than other two mappings.

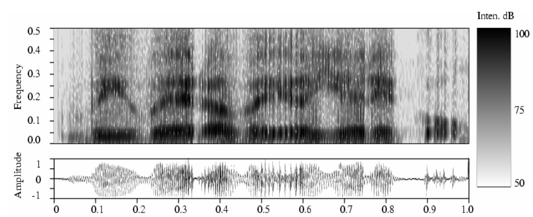


(a) Unprocessed

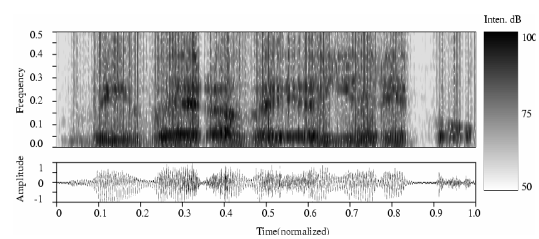


(b) Segment mapping

Fig. 6 Wide band spectrograms ($\Delta f = 300$ Hz) of syllable /aka/: (a) unprocessed and (b) frequency compressed with segment mapping scheme and compression factor = 0.6.



(a) Unprocessed



(b) Segment mapping

Fig. 7 Wide band spectrograms ($\Delta f = 300$ Hz) of the sentence "we were away a year ago": (a) unprocessed and (b) frequency compressed with segment mapping scheme and compression factor = 0.6.

To asses the usefulness of the scheme in improving speech perception under adverse listening conditions and under simulated hearing loss, tests were conducted by adding broadband noise (maintaining constant SNR on short-time, i.e. 10 ms, basis) to the processed speech signal. Noise was added at SNR values of 6, 3, 0, -3, -6, -9, -12, and -15 dB. Compression factors 0.8 and 0.6 were found to be beneficial for SNR values of 6, 3, 0, and -3 dB. At lower SNR (-6, -9, -12 and –15 dB), the intelligibility of the processed speech increased with decrease in the compression factor (0.4 and 0.2).

Intelligibility of the processed speech using segment mapping was found better than other two mappings for the same SNR.

IV. CONCLUSION

Multi-band frequency compression is a speech processing technique for improving speech intelligibility under adverse listening conditions. For use in this processing, three frequency-mapping schemes, i.e. sample-to-sample mapping, mapping by superimposition of spectral samples, and segment mapping schemes were investigated. Segment-mapping scheme achieved desired compression retaining the spectral distribution of energy, and without introducing irregular variations. The scheme needs to be further evaluated using different test materials, larger number of subjects, and different listening conditions.

REFERENCES

[1] B. C. J. Moore, *An Introduction to Psychology of Hearing*. 4th ed. London: Academic, 1997.

[2] D. O'Shaughnessy, *Speech Communication: Human and Machine*. Hyderabad: University Press, 2001.

[3] B. R. Glasberg and B. C. J. Moore, “Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,” *J. Acoust . Soc. Am.*, vol. 79, pp. 1020-1033, 1986.

[4] A. E. Carney, D. A. Nelson, “An analysis of psychophysical tuning curves in normal hearing and pathological ears,” *J. Acoust . Soc. Am.*, vol. 73, pp. 268-278, 1983

[5] J.R. Dubno and D. D. Dirks, "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," *J. Acoust . Soc. Am.*, vol. 85, no.4, pp. 1666-1675, 1989.

[6] A. N. Cheeran and P. C. Pandey, “Evaluation of speech processing schemes using binaural dichotic presentation to reduce the effect of masking in hearing-impaired listeners,” *Proc. 18th International Congress on Acoustics, ICA 2004,* Kyoto, Japan, pp. 1523 - 1526, Apr. 4–9, 2004.

[7] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, “Perceptually balanced filter response for binaural dichotic presentation to reduce the effect of spectral masking,” *J. Acoust . Soc. Am.*, vol. 120, no.5, pp. 3253, 2006.

[8] K. Yasu, K. Kobayashi, K. Shinohara, M. Hishitani, T. Arai, and Y. Murahara, "Frequency compression of critical band for digital hearing aids," in *China-Japan Joint Conf. on Acoustics*, pp. 159-162, 2002.

[9] C. M. Reed, B. L. Hicks, L. D. Braida, and N. I. Duriach, "Discrimination of speech processed by low pass filtering and pitch invariant frequency lowering," *J. Acoust. Soc. Am.*, vol. 74, no. 2, pp. 409- 419, 1983.

[10] T. Arai, K. Yasu, and N. Hodoshima, “Effective speech processing for various impaired listeners,” in *Proc. 18th International Congress on Acoustics* (*ICA 2004,* Kyoto, Japan), pp. 1389 - 1392, Apr. 4–9, 2004.

[11] E. W. Zwicker, “Subdivision of audible frequency range into critical bands (Freqenzgruppen),” *J. Acoust . Soc. Am.*, vol. 33, pp. 248, 1961

[12] J. G. Proakis and D.G.Manolakis, *Digital Signal Processing Principles, Algorithms, and Applications*. New Delhi: Prentice Hall, 1997

[13] S. K. Mitra, *Digital Signal Processing, a Computer- Based Approach*. Singapore: McGraw-Hill, 1998.