NOISE SUPPRESSION USING A TIME-VARYING, ANALYSIS/SYNTHESIS GAMMACHIRP FILTERBANK

Toshio Irino

ATR Human Information Processing Research Laboratory
2-2 Hikaridai Seika-cho Soraku-gun Kyoto, 619-0288, JAPAN

ABSTRACT

Spectral subtraction has been cited most often as a noise suppression method for speech signals in steady background noise, because it is basically a non-parametric method and simple enough to implement for various applications using FFT. It has also been well known, however, that spectral subtraction produces so called “musical noise” in synthetic sounds. Since such musical noise, even at low levels, can often bother humans in speech perception, spectral subtraction has not been very successful in signal processing applications for human listeners. To suppress noise without producing musical noise, an alternative method has been developed using a time-varying, analysis/synthesis gammachirp filterbank; this was initially proposed as an auditory filterbank. The present method achieves about the same SNR improvement as spectral subtraction when using the same information on the non-speech interval. Moreover, the synthetic sounds only contain steady white-like noise at reduced levels when the original noise is white. This method is, therefore, advantageous in various applications for human listeners.

1. INTRODUCTION

Spectral subtraction [1] has been cited most often as a noise suppression method for speech signals with static background noise. This may be because it is basically non-parametric in terms of speech parameters and simple enough to implement for various applications using FFT. Non-parametric methods are advantageous to parametric methods, e.g., adaptive comb filter methods [2], when the signal to noise ratio (SNR) is small. This is because the estimation of speech parameters, such as the fundamental frequency, becomes harder as the SNR decreases.

Spectral subtraction, however, has not been very successful in signal processing applications for human listeners, because of the well-known “musical noise” in synthetic sounds. Such musical noise, even at low levels, can often bother humans in speech perception. Various methods have been proposed to reduce musical noise [3], but proper evaluation have not been made at low SNR values nor has musical noise suppression been guaranteed under any conditions. Solution seems distant because musical noise is produced by physical inconsistencies of subtracted amplitude spectra and unmodified phase. To essentially overcome the problem, it is necessary to develop a filtering process to attenuate or enhance spectra without separating the amplitude and phase components. It is also necessary to adopt a non-parametric method as advantageous as spectral subtraction.

This paper proposes such a noise suppression method based on a time-varying, analysis/synthesis gammachirp filterbank [4]. This filterbank was originally developed as an auditory filterbank based on a level-dependent gammachirp filter to explain psychoacoustic masking data [5]. Unlike conventional auditory filterbanks for signal analysis, synthesis filterbanks were established through an IIR implementation of the gammachirp filter. Moreover, this filterbank guaranteed the small and time-invariant resynthesis error. The main interest of this development is, however, the capability to modify output representations of the analysis filterbank to reproduce sounds with desirable characteristics. One of the most important applications is the noise suppression presented here.

Section 2 explains noise suppression using the gammachirp filterbank. Subsection 2.1 shows the definition and the characteristics of the gammachirp. Subsection 2.2 shows a basic structure of the noise suppression filterbank. Subsection 2.3 explains the principle of noise suppression by using this filterbank. Subsection 2.4 shows another filterbank structure to improve the performance, only by using the same condition as spectral subtraction. Section 3 shows an evaluation of noise suppression by the gammachirp filterbank method and spectral subtraction.

2. NOISE SUPPRESSION FILTERBANKS

2.1 Gammachirp

The gammachirp was analytically derived as a function satisfying minimum uncertainty in a time-scale representation and was applied to fit a level-dependent auditory filter shape explaining psychoacoustic masking data [5].

The complex impulse response is given as

$$g_c(t) = a\, t^{n-1} \exp\left(-2\pi b \mathrm{ERB}(f_r)\ t\right) \exp\left(j2\pi f_r t + jc \ln t + j\phi\right) \quad (1)$$

where t is the time, a is the amplitude, n and b are parameters defining the distribution, f_r is the asymptotic frequency, c is the parameter for the frequency modulation, ϕ is the initial phase, $\ln t$ is a natural logarithm of time, and $\mathrm{ERB}(f_r)$ is the equivalent rectangular bandwidth of the filter at f_r [6]. When c=0, this equation represents the complex impulse response of the gammatone [7] having only a sinusoidal carrier in the same envelope.

The amplitude spectrum of Eq. (1) is derived as

$$|G_c(f)| = \frac{|a\Gamma(n+jc)|}{|2\pi b\mathrm{ERB}(f_r) + j2\pi(f-f_r)|^n} \cdot e^{c\theta} \quad (2)$$

$$\theta = \arctan\{(f-f_r)/b\mathrm{ERB}(f_r)\} \quad (3)$$

The peak frequency of Eq. (2) is obtained analytically as

$$f_p = f_r + c \cdot b\mathrm{ERB}(f_r)/n. \quad (4)$$

The first term in Eq. (2) represents the amplitude spectrum of the gammatone; it is almost symmetric in the linear frequency axis. The second term represents an asymmetric function that introduces asymmetry into the amplitude spectrum of the gammachirp. The asymmetry becomes greater as the absolute c value increases. This function can be simulated by an asymmetric compensation filter designed as a minimum-phase IIR filter [4].

Consequently, a gammachirp filter can be simulated with the combination of a gammatone filter and an asymmetric compensation filter.

It has been demonstrated [5] that the gammachirp filter fits human psychoacoustic masking data [8] well when the parameter c is associated with the sound pressure level typically as

$$c = 3.38 - 0.107\, P_s \quad (5)$$

where P_s is the threshold level (in dB) of a probe sinusoid in notched noise. Since P_s is usually greater than about 30 dB, the parameter c is typically negative and produces a shallow slope below the peak frequency and a steep slope above the peak frequency in the amplitude spectrum. When the sound level increases, the absolute c value becomes larger, and then, the shallowness and steepness are enhanced.

In this paper, input sounds are properly normalized in advance to utilize the effective range of the c value, and as a result, the value of P_s does not correspond to the absolute sound pressure level since the main purpose is to demonstrate effective noise suppression by using the gammachirp filterbank.

2.2 Basic structure

Figure 1 shows the block diagram of a basic noise suppression filterbank. The left box labeled “Gammachirp Analysis” is the gammachirp filterbank proposed in [4]. The gammachirp filterbank consists of a gammatone filterbank and an asymmetric compensation filterbank controlled by a parameter controller with sound level estimation. The parameter controller changes the value of the parameter c in the asymmetric compensation filter in accordance with the estimated sound pressure level at the output of the filterbank on a sample-by-sample basis.

The right box labeled “Noise Suppression Synthesis” is a synthesis mechanism that reproduces sounds with suppressed noise. The first stage is a channel shift block, which shifts the output of the gammachirp analysis filters uniformly for a few channels below to be fed into gammatone synthesis filters of lower channels. The next block consists of a time-reversal linear gammatone filterbank and performs weighted summation to synthesize signals from the shifted gammachirp output. Noise suppressed sounds are, then, resynthesized from the shifted gammachirp output. The principle of noise suppression with this mechanism is described in subsection 2.3.

By directly connecting the gammatone analysis filterbank in the first block of Fig. 1 and the time-reversal gammatone synthesis filterbank with summation in the last block, the structure is almost the same as the wavelet transform [9] commonly used for signal resynthesis. Thus, this system performs noise suppression by using a level-dependent, time-varying asymmetric compensation filterbank and a channel shift operator between these two linear filterbanks.

2.3 Principle of noise suppression

2.3.1 Problem formulation

Let’s consider the k_1-th and k_2-th channels of the gammachirp analysis filterbank and the gammatone synthesis filterbanks in Fig. 1, where the k_2-th filter does not largely overlap the k_1-th filter. In addition, let’s assume the dominant formant frequencies of a steady speech signal are located close to the k_1-th channel and far from the k_2-th channel. We further assume that white noise at a low level exists as the background noise. Then, the estimated sound pressure level is greater in the k_1-th channel than in the k_2-th channel. The power of the noise is much smaller than the power of the speech in the k_1-th channel and little smaller in the k_2-th channel.

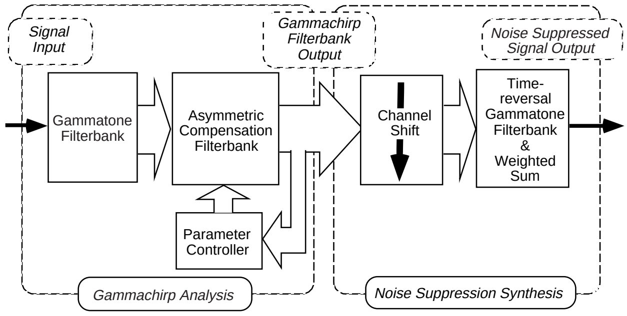

Figure 1. Block diagram of a basic noise suppression system based on a time-varying analysis/synthesis gammachirp filterbank.

When the idea of spectral subtraction is applied in this situation, the estimated noise level is subtracted from the total level of each channel. Then, the output of the k_1-th channel is reduced relatively little, whereas the reduction in the output of the k_2-th channel is relatively large. This is equivalent to the state that the gain of the analysis/synthesis filter is greater in the k_1-th channel than in the k_2-th channel. This is one of the conditions to perform the noise suppression.

Figure 2(d) shows the amplitude spectra of the k_1-th and k_2-th analysis/synthesis filters, as an example of the above situation. The next subsection explains the principle of noise suppression in Fig. 2.

2.3.2 Explanation of a solution

1) Figure 2(a) shows an example corresponding to the situation described above. According to a negative correlation between the signal level and the parameter c as shown in Eq. (5), the values of the parameter c are assumed to be, say, -3 and –1, in the k_1-th and k_2-th channels, respectively. The peak frequencies of the k_1-th and k_2-th gammachirp filters (solid lines) become lower than those of the corresponding gammatone filters, but the degree of the peak frequency shift is greater in the k_1-th channel than in the k_2-th channel as indicated by Eq. (4).
2) By considering that the output of these gammachirp filters is fed into the gammatone synthesis filters shown in Fig. 2(c), the total filter gain becomes smaller in the k_1-th channel than in the k_2-th channel. This is because the total gain is calculated from the spectral multiplication of the analysis and synthesis filters, and the disparity in the peak frequencies between the analysis and synthesis filters in the k_1-th channel is much larger than in the k_2-th channel. This operation results in a noise enhancement that is against our purpose. Accordingly, some additional operations are necessary in the system.
3) Consider the analysis filters of the α channel above those channels where α is a small positive integer. The estimated sound pressure levels in the $(k_1+\alpha)$-th and $(k_2+\alpha)$-th channels are about the same as those in the k_1-th and k_2-th channels, respectively, because the sound pressure level is estimated with the average of adjacent channels. Therefore, as illustrated in Fig. 5(b), the c values and the degrees of the peak frequency shift in the $(k_1+\alpha)$-th and $(k_2+\alpha)$-th channels are almost the same as those in the k_1-th and k_2-th channels.
4) Select the α value properly so that the peak frequency of the $(k_1+\alpha)$-th analysis filter is close to the peak frequency of the k_1-

th synthesis filter. Then, the output of the $(k_1+\alpha)$-th and $(k_2+\alpha)$-th analysis filters are fed into the k_1-th and k_2-th synthesis filters, respectively. This is the function of the shift block in Fig. 1.

5) The total amplitude spectra of the analysis and synthesis filters are shown in Fig. 2(d). The filter gain is greater in the k_1-th channel than in the k_2-th channel. This result satisfies the condition of noise suppression described in subsection 2.3.1.

2.4 Improvement of noise suppression

In the previous section, it was demonstrated that noise suppression is possible only by using the relative powers of speech and noise. This procedure, however, seems effective only for high SNR cases with white noise. To improve the performance, we introduce the same assumption as used in spectral subtraction, i.e., the non-speech interval can be detected prior to the noise suppression procedure[1].

Figure 3 shows a filterbank structure for using this assumption. Asymmetric compensation filterbank B (ACFB-B) at the bottom of the analysis part is of the same structure and gets the same control parameters as the asymmetric compensation filterbank shown in Fig. 1. The only difference is that this filterbank does not produce the output for the synthesis filterbank.

Asymmetric compensation filterbank A (ACFB-A) does not have a feedback path to the parameter controller and only receives parameter values to produce the output of the gammachirp filterbank for the synthesis filterbank. The c value to control the k-th channel, $c_{Ak}(t)$, is formulated as

$$c_{Ak}(t) = \{c_{Bk}(t) - \eta \cdot c_{BNk}\} \cdot \max \left| \frac{c_l}{c_l - \eta \cdot c_{BNk}} \right| \qquad (6)$$

where $c_{Bk}(t)$ is the c parameter for the k-th channel in ACFB-B, c_{BNk} is calculated by Eq. (5) from the average noise level at the output of ACFB-B during the non-speech interval, c_l is the lower limit of the c value (–3.5 in the simulation), and η is a constant to amplify the estimated average for the noise. The first term is the most important part for noise suppression. The second term is a normalization factor to effectively use the full range of c values allowed in the asymmetric compensation filters.

Since the subtraction is performed in the domain of parameter c and the c value may take either a positive or negative value, this procedure does not require half-wave rectification essentially used in spectral subtraction. The value of $c_{Ak}(t)$ is positive when the noise component $|\eta\, c_{BNk}|$ is larger than $|c_{Bk}(t)|$. The positive c value moves the peak frequency higher in accordance with Eq. (4) and enlarges the disparity between the analysis and synthesis filters. Consequently, it suppresses the filter gain further. In Fig. 2(b), it is likely the case in the k_2-th channel because the noise component is relatively large in this channel.

In the sense of noise component subtraction, however, this procedure resembles spectral subtraction since Eq. (6) can also be interpreted as a function of the signal level for each channel when using Eq. (5). Therefore, it is advantageous for this procedure to easily incorporate many of the procedures developed for spectral subtraction to improve the performance.

3. EXPERIMENTAL RESULTS

The solid lines in Fig. 4 denote experimental results on noise suppression when using the gammachirp filterbank method

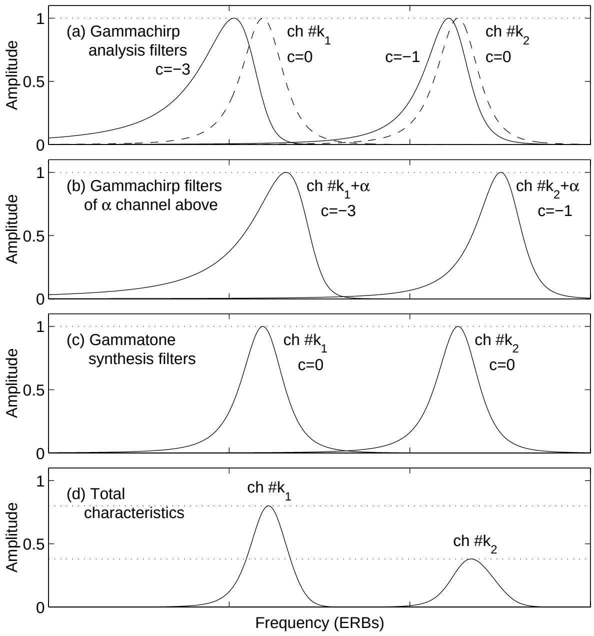

Figure 2. Schematic graphs for explaining the principle of noise suppression. Amplitude spectra of gammachirp and gammatone filters are plotted as a function of frequency (arbitrary ERB rates). See text for more detail.

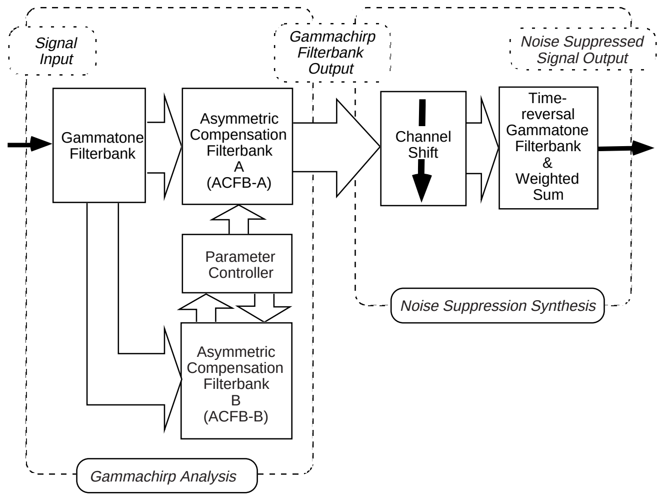

Figure 3. Block diagram of a noise suppression system usable for information on non-speech intervals.

(GCFB_NS) and spectral subtraction. The abscissa is the SNR of the input speech signal and the ordinate is the SNR of the processed speech signal. In the present experiment, two spoken words (from one male speaker and one female speaker) are used as speech signals and white noise is used as the background noise. The average SNR values are plotted with solid lines since the results are not so different from each other. GCFB_NS1 (triangle and solid line) shows the result when using the basic noise suppression filterbank in Fig. 1 explained in subsection 2.2. This filterbank improves the SNR for about 5 dB when the input SNR is high, but does not improve the SNR so much when the input SNR is low. GCFB_NS2 (circle and solid line) shows the result when using the filterbank in Fig.3 and the procedure explained in subsection 2.4. The improvement of the SNR is slightly better than with spectral subtraction (asterisk and solid line) when both methods are assumed to have information on the non-speech interval in advance. Since the results may depend on the coefficient for the subtraction, both methods can achieve comparable SNR improvement.

The important difference is that GCFB_NS2 does not produce any musical noise even at the lowest input SNR, whereas spectral subtraction clearly produces musical noise even at the highest input SNR. The most advantageous feature of GCFB_NS is that white noise is simply resynthesized as steady white-like noise at a reduced level. It is impressive that the noise is efficiently suppressed to release speech sounds from loud noise when the SNR is -2 dB. The speech sounds are slightly distorted and cut off at a high frequency, but are still clear and intelligible [10].

The dashed lines in Fig. 4 indicate results obtained using a recent method incorporating adaptive orthogonal comb filters (square and dashed line, labeled “Adaptive KLT”) and spectral subtraction (cross and dashed line) in [11]. It has been reported that the adaptive KLT does not produce musical noise. However, it does not show a better SNR improvement than GCFB_NS2, even though it uses precise information on the fundamental frequency of the speech at every moment. GCFB_NS2 is more advantageous than comb filter methods because it only uses information on the non-speech interval that is estimated more easily than any speech parameters at low SNR values.

4. SUMMARY

A new method has been proposed to achieve effective noise suppression without producing any musical noise. The method is based on a time-varying, analysis/synthesis gammachirp filterbank that enables spectral enhancement without separating the amplitude and phase spectra. It is basically non-parametric in terms of speech parameters and accomplishes almost the same SNR improvement as spectral subtraction when using the same information on the non-speech interval. Moreover, it can easily incorporate many of the techniques developed for spectral subtraction to improve the performance, because it has a procedure including subtraction that is similar to the procedure in the spectral subtraction. It can also combine other speech signal processing methods since no uncontrollable musical noise is produced. Even when the SNR is low, say, -2 dB, noise is efficiently suppressed to release speech sounds from loud noise. The synthetic noise still sounds like white noise at a reduced level when the input noise is white.

It has been demonstrated that the gammachirp is an excellent function for regulating the signal flow, as demonstrated in ACFB-A in Fig. 3, as well as for characterizing human auditory filter shapes [5]. Although the absolute sound pressure level is ignored in the present simulation, more precise control can be introduced into ACFB-B based on human peripheral auditory functions. Consequently, this method is advantageous in various speech processing applications for human listeners.

ACKNOWLEDGMENTS

This work was partially supported by CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (JST).

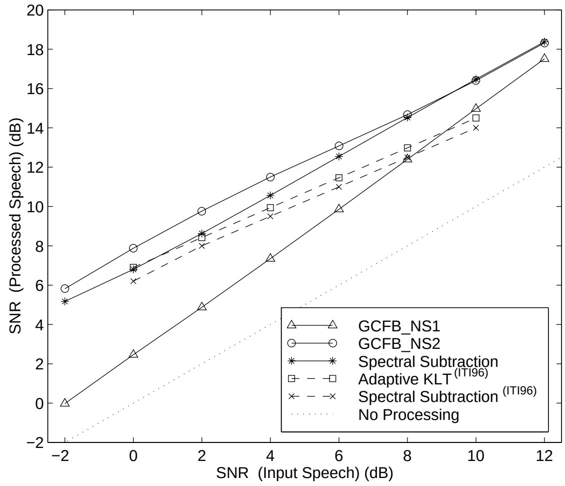

Figure 4. Performance comparison between Gammachirp Filterbank Noise Suppression (GCFB_NS) and Spectral Subtraction (solid lines). The results represented by the dashed lines (ITI96) are adapted from [11].

REFERENCES

[1] Boll, S.F., “ Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Processing, **ASSP-27**, pp. 113-129, 1979.

[2] Frazier, R.H., Samsam, S., Braida, L.D. and Oppenheim, A.V., ”Enhancement of speech by adaptive filtering,” IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-76), 251-253, 1979.

[3] Arslan, L., McCree, A. and Viswanathan, V., ”New methods for adaptive noise suppression,” IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-95), 812-815, 1995.

[4] Irino, T. and Unoki, M., ”A time-varying, analysis/ synthesis auditory filterbank using the gammachirp,” IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-98), 3653-3656, 1998.

[5] Irino, T. and Patterson, R.D., "A time-domain, level-dependent auditory filter: The gammachirp", J. Acoust. Soc. Am. **101**, 412-419, 1997.

[6] Glasberg, B. R. and Moore, B. C. J. , ”Derivation of auditory filter shapes from notched-noise data,” Hear. Res., **47**, 103-138, 1990 and reference herein.

[7] Patterson, R. D., Allerhand, M. and Giguere, C., "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," J. Acoust. Soc. Am., **98**, 1890-1894, 1995 and reference herein.

[8] Rosen, S. and Baker, R.J., ”Characterising auditory filter nonlinearity,” Hear. Res., **73**, 231-243, 1994.

[9] Combes, J.M., Grossmann, A. and Tchamitchian, Ph. Eds., ”Wavelets,” Springer-Verlag, Berlin, 1989.

[10] For demonstration of noise suppression, see the web page, URL: http://www.hip.atr.co.jp/~irino

[11] Ikeda, M., Takeda, K., and Itakura, T., “Speech enhancement by quadrature comb-filter. “ Tech. Report IEICE, DSP96-70, SP96-45, 1996.