Proceedings of the International Symposium on Room Acoustics, ISRA 2010

29–31 August 2010, Melbourne, Australia

Application of an Auditory Filter for the Evaluation of Sounds and Sound Fields

Yuki Matsumoto, Masahiro Suzuki, Hisako Ogushi, and Akira Omoto

Faculty of Design, Kyushu University, Fukuoka 815-8540, Japan

PACS: 43. 55. Hy, Mc, 43.64.Bt, 43.66.Jh

ABSTRACT

Auditory filters that simulate function of the human auditory organ are introduced for the evaluation of a sound and sound field. Three applications are discussed in this report. The first examines the similarity of sound fields. Music signals convolved with different impulse responses are compared subjectively and quantitatively. The output of the auditory filter indicated high correlations with subjective evaluation. The second example examines the decay process of the sound field, especially in rather dead conditions. Impulse responses measured in recording studios are passed through both an auditory filter and a conventional band-pass filter. Decay processes are then calculated by inverse integration. The responses after passing through the auditory filter show more stable decay curves and the obtained reverberation times are closer to the early decay time, which is closely related to human perceived reverberance. The timbres of musical instruments are used as the third example. The signals from instruments with different means of sound production (such as reed or mouthpiece) are processed by auditory filters and normal band-pass filters. Statistically-obtained characteristics such as kurtosis of the waveform again show higher correlation with subjective evaluation when employing the auditory filter. All the results shown in this report strongly suggest the superiority of an auditory filter in the qualitative evaluation of sounds and sound fields.

INTRODUCTION

Recently, various acoustical indices, such as reverberation time, clarity and inter-aural cross-correlation, are often used to evaluate the sound field[1]. These indices basically evaluate the properties of the sound field base concerning human perceptions, such as reverberance, the intelligibility of the music or speech, listener envelopment, and so on[2, 3, 4]. Moreover, most of the indices are calculated from the impulse responses measured with microphones.

In the present work, we have applied an auditory filter that simulates the functions of the human auditory organ to process the impulse response and also some real music signals. The following three subjects are used as examples.

First, we focused on the perceived similarities between sound fields. In this examination, some music signals that have been convolved with different room impulse responses are analyzed by using the *Gammatone* filter[5]. Comparison of the results with experiment involving subjective evaluation indicated the importance of time-domain similarity, i.e., similarity in waveforms, which is processed by the gammatone filter, rather than similarity in frequency characteristics.

In the second example, the reverberation process is examined. In this case, the *dynamic compressive Gammachirp* filter[6] is applied instead of a conventional band-pass filter in evaluating impulse responses. The focus has been on the shapes of decay curves calculated from the impulse responses, which have been inversely integrated and on the 60 dB decay time obtained from these curves. The relation between the shapes of decay curves and the early decay time (called EDT hereafter) was examined.

The dynamic compressive Gammachirp filter is then used for evaluating the timbres of musical instrument sounds in the third example. The sounds from three kinds of musical instruments with several conditions are processed with a dynamic compressive Gammachirp filter. As the perception of timbre is subjective, we have compared statistically processed values from waveforms with two kinds of subjective evaluation experiments.

AUDITORY FILTER

The basic characteristics and the practical implementation of filters are introduced. The Gammatone filter and its modified forms, the Gammachirp and the dynamic compressive Gammachirp filters, have been used in this work.

Gammatone filter

Patterson introduced the auditory filter named "*Gammatone*" to the human auditory model[5]. This Gammatone filter was originally developed to characterize the impulse response data of the basilar membrane, and the filter shape was derived using "notched-noise" masking data. This filter has an envelope of the gamma distribution function and the carrier of the complex exponential function, and simulates the passive vibration of a basilar membrane. The complex impulse response can be expressed as

$$g_t(t) = at^{n_1-1}\exp(-2\pi b_1 \mathrm{ERB_N}(f_r)t) \times \exp(j2\pi f_r t + j\varphi_1), \quad (1)$$

where a is the amplitude, n_1 and b_1 are parameters defining the envelope of the gamma distribution, φ_1 is the initial phase, f_r is the carrier frequency in kHz and $\mathrm{ERB_N}$ (f_r) is the equivalent rectangular band that can be expressed as

$$\mathrm{ERB_N}(f_r) = 24.7(4.37 f_r + 1). \quad (2)$$

Gammachirp filter

Subsequently, Irino developed the "*Gammachirp*" filter, which was the optimum auditory filter in terms of minimal uncertainty when the human auditory system is expressed in the time domain[7, 8]. He demonstrated that the gammachirp filter showed better fit to the data in psychological experiments than the original gammatone filter. The complex impulse response can be expressed as

$$g_c(t) = at^{n_1-1}\exp(-2\pi b_1 \mathrm{ERB_N}(f_r)t) \times \exp(j2\pi f_r t + jc_1 \ln t + j\varphi_1). \quad (3)$$

The difference compared with the original gammatone filter (Eq. (1)) is the term of $jc_1 \ln t$ in the second exponential function. The Gammachirp filter uses the product of chirp factor c_1 and the natural logarithm of time. When $c_1 = 0$, the gammachirp filter is reduced to the gammatone filter. The gammachirp filter provides the optimum filter function in terms of calculation theory and is able to approximate the physiological and psychological data.

Dynamic Compressive Gammachirp filter

Irino and Patterson also devised $h_c(t)$ to simulate the nonlinear compression characteristics of the human auditory system[6] and proposed the complex impulse response of the *dynamic compressive gammachirp* filter (dcGC filter hereafter) $g_{cc}(t)$ by convolving $h_c(t)$ with g_c as

$$g_{cc}(t) = a_c \cdot g_c(t) * h_c(t). \quad (4)$$

The $h_c(t)$ is the impulse response of an asymmetric high-pass filter $H_C(f_a)$[8] to control the gain of $g_{cc}(t)$ with an additional term a_c. $H_C(f_a)$ simulates the asymmetric function such that

$$|H_C(f_a)| \cong \exp(c_2 \cdot \theta)$$
$$\theta(f_a) = \arctan\left(\frac{f - f_a}{b_2 \mathrm{ERB_N}(f_a)}\right). \quad (5)$$

The center frequency f_a of $H_C(f_a)$ shifts, depending on the input signal level. In short, the gain of the dcGC filter depends on the input signal level, and the dcGC filter can simulate the auditory compression characteristics with non-linearity. Table 1 shows coefficient values for the dcGC filter used in this report.

Table 1: Coefficient values

a, a_c	n_1	b_1	c_1	φ_1	b_2	c_2
1	4	1.81	−2.96	0	2.17	2.20

These auditory filters are composed of a bank of band-pass filters, and the user can arbitrarily set an individual center frequency for this auditory filter-bank.

EVALUATION OF SOUND FIELD SIMILARITY

First, we applied the Gammatone filter to the evaluation of the similarity of sound fields. The validity of the auditory filter described in this chapter has been studied by comparing the result of the analysis with the result of a subjective evaluation experiment.

Evaluation Method

The number of the filters in the Gammatone filter-bank was set to sixteen. The lowest and highest center frequencies were set to 50 Hz and 16000 Hz, respectively. These frequencies were divided on an"ERB-Rate" scale.

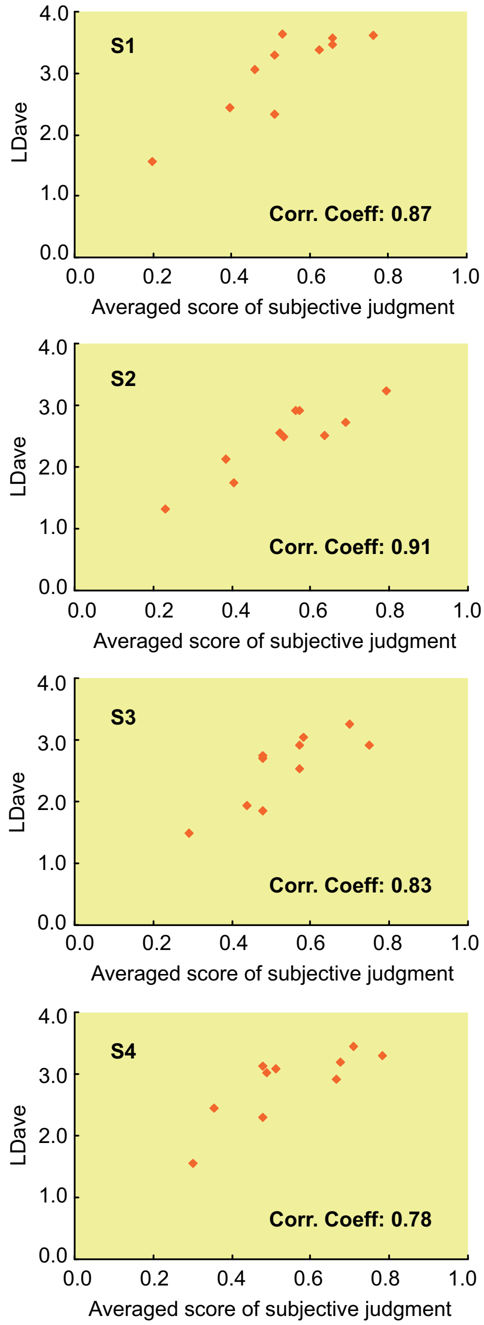


Figure 1: The correspondence of subjective judgment and the Gammatone filter output, LD_{ave}.

The filtered signal by the Gammatone filter was converted to a decibel scale and the comparisons were carried out by the differences between levels. Practically, a simple averaged value of level difference, LD_{ave}, was defined as

$$LD_{\mathrm{ave}} = \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}|L_a(m,n) - L_b(m,n)|, \quad (6)$$

where m is the number of bands, n is the time sample data, M and N are the total number of m and n respectively, and L_a and L_b are the levels of filtered signals obtained by convolving two different impulse responses.

Correlation between LD_{ave} and Subjective Evaluation

The experiment for subjective evaluation was carried out simultaneously. Twelve subjects aged 22 to 25 years were asked to

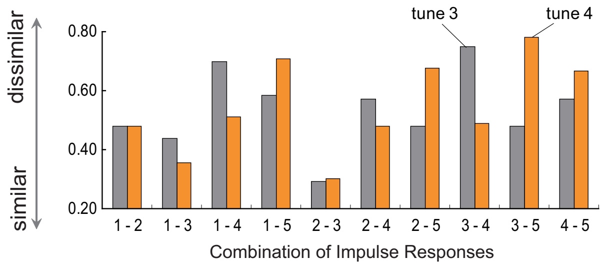


Figure 2: Difference of subjective evaluation between two tunes.

Table 2: Correlation coefficients between LD_{ave} and subjective evaluation.

Music number	S1	S2	S3	S4	ave
Gammatone filter	0.87	0.91	0.83	0.87	0.85
band-pass filter	0.84	0.89	0.66	0.77	0.79

judge the similarity of two acoustic signals that formed the stimulation pair, and the averaged score from a five-step judgment was calculated for all subjects. Five different room impulse responses were convolved with the same music signals to generate the paired acoustic signals. Ten evaluation pairs $({}_5C_2 = 10)$ were obtained in total.

We then calculated the correlation coefficient between the averaged scores of subjective evaluation and the LD_{ave} value. The trial was repeated using four kinds of music. The music indicated as S1 was a kind of Rhythm and Blues, S2 was orchestral music, S3 was Irish music, and S4 was Big Band Jazz.

Figure 1 shows the results. In each figure, the horizontal axis indicates the averaged score of subjective judgment and the vertical axis indicates the distribution of LD_{ave}. The correlation coefficients are indicated on the graphs and the averaged value for all four songs was 0.85.

The genre of music seriously affected the subjective evaluation. Figure 2 shows the result of subjective evaluation when the evaluation objects in this example were S3 and S4. The horizontal axis indicates different combinations of the five impulse responses and the vertical axis indicates the averaged scores.

For example, the two evaluated values for the combination of "3-4" or "3-5" showed quite different values for the tunes. In this particular case, the conventional band-pass filter could not follow the variation in subjective evaluation. Table 2 shows the correlation coefficients between subjective evaluation and LD_{ave} from both Gammatone and conventional band-pass filters. It is apparent that the Gammatone filter scored more highly than the conventional band-pass filter, even in the case of S3 and S4.

EVALUATION OF THE REVERBERATION PROCESS

We also have adopted the dcGC filter in the evaluation of the reverberation process. In the proposed method in this chapter, the dcGC filter and a band-pass filter has been used in the calculation of the reverberation curve, and we have compared EDT (Early Decay Time) obtained from both filters.

Analysis method

The analysis system proposed in this report filters the room impulse responses and calculates reverberation curves. Thus most of the method is the same as in the traditional method of calculating reverberation curves. However, we use here two filters for analysis: the conventional octave-band filter (Oct filter, hereafter) and the dcGC filter outlined above.

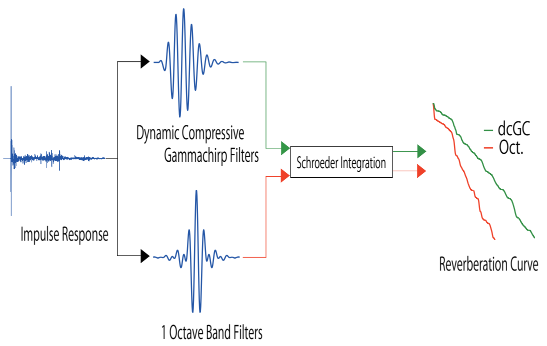


Figure 3: Block diagram of the analysis method

Figure 3 shows the block diagram of the analysis system. First, the filters are applied to the impulse responses. The center frequencies of the two types of filters were set to 500 Hz, 1,000 Hz and 2,000 Hz. Additionally, in the process of dcGC filtering, an equal-loudness contour filter, which was implemented as a FIR filter, was used to simulate the outer- and mid-ear transfer functions for preprocessing[7, 9]. Two kinds of filtered signals were then obtained and reverberation curves were calculated by Schroeder integration[10]. The impulse responses measured at a recording studio were used.

Results and discussion

The reverberation time of the studio measured was around 0.12 to 0.14 sec. As mentioned in the section "AUDITORY FILTER", the impulse response of the dcGC filter can be time-varying, depending on the magnitude of the input signal. Therefore, the magnitude of the impulse response affected the shape of the decay curves and the decay time. To examine this relationship, the magnitude of the impulse response was modified by multiplying by arbitrary constants.

Figure 4 shows the decay curves obtained for impulse response after passing through the Oct and dcGC filters. Since the impulse response of the studio has rather 'sparse' characteristics, the initial part of the decay process in the Oct filter (around 0 to 2 ms) showed a shape integrated with the impulse response of the band-pass filter itself. These parts were therefore omitted from the evaluation process for the decay times.

For convenience, the waveforms used in our examination were assumed to be pressure waves having the dimension of Pa (N/m^2), to demonstrate behavior of the dcGC filter at different intensities of sound. In practice, the magnitude of the response was scaled as the peak value of the waveform corresponding to the levels of 50 dB and 80 dB re 2×10^{-5} Pa, the reference sound pressure.

As shown in the figure, when the signal level was 80 dB, the gradient of the decay curves obtained with the Oct filter was calculated using a rather early decay process, for example, within -5 dB to -12 dB, which is similar to the range for calculating EDT, 0 to -10 dB. It has been said that EDT is closely correlated with the human perception of reverberance. The 80dB level is also close to that used in the psychology experiment. When comparing these two facts with the analysis result using a dcGC filter, it may be said that the proposed method is evaluating the reverberance.

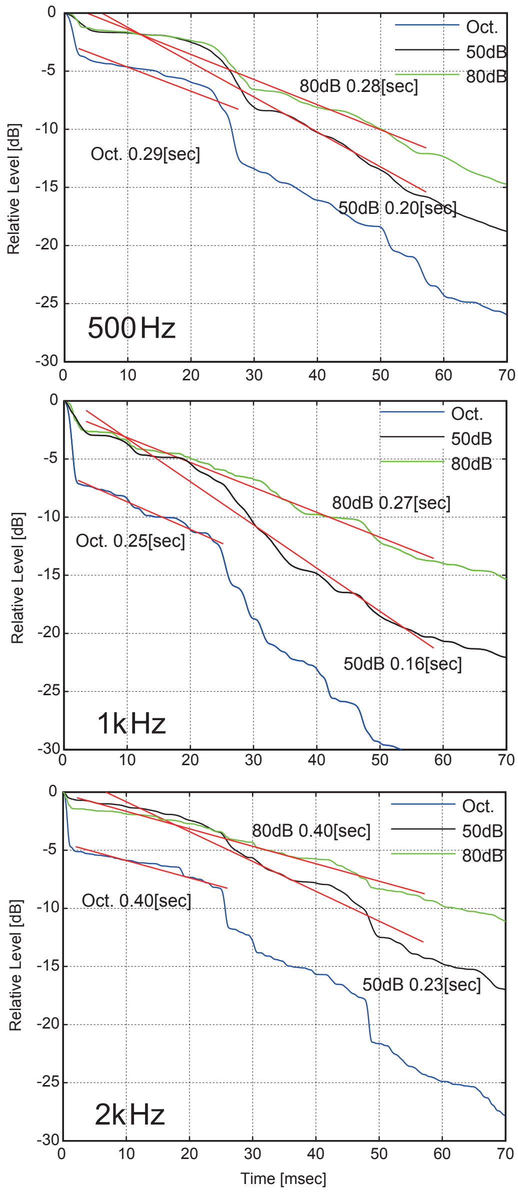


Figure 4: Reverberation Curves by using Oct and dcGC, at the recording studio.

Stepwise changes in the decay process and changes in the slope (the individual decay process is referred as decay mode, hereafter) are apparent in the case of the Oct filter, before and after around 25 ms. The gradients obtained using the dcGC filter, particularly in the case of 80 dB, attenuate linearly and are almost parallel with that of the first decay mode in the Oct filter and therefore all decay times tend to be similar. These results indicate the possibility that the dcGC filter can enhance the important initial decay mode (which is of a similar duration) for evaluating EDT. Moreover, the decay time obtained from the dcGC filter varied around 40 % to 70 % with changes in the magnitude of waveforms. These results also showed the capability of this method to indicate the perceived reverberance at various sound pressure levels.

To verify the function of dcGC filter more generally, an extremely 'sparse' impulse response was introduced. Figure 5 shows the original and the artificially created sparse impulse response, which was made by following procedures; (1) picking up the extreme peak positions within the initial 30 ms; and (2) adding the reflection sound (pulse) element at these positions

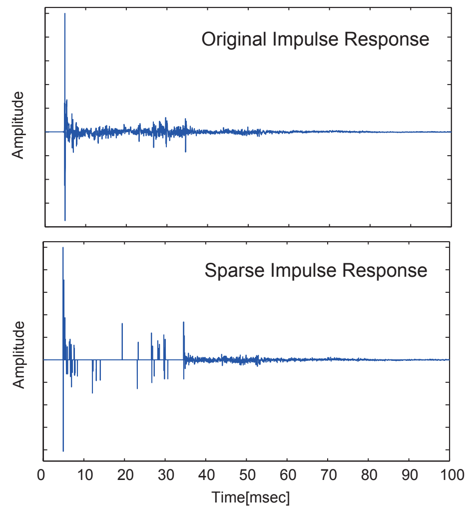


Figure 5: The original and sparse impulse responses.

after removing all other reflections. The overall energy of response was kept constant and the reverberation time observed from each impulse response was almost identical.

Figure 6 shows the decay curves obtained by dcGC (the level was set to 80 dB) and Oct filters. The Oct filter generated several decay modes corresponding to the sparse reflections. The dcGC filter, on the other hand, resulted in smoother decay curves at all the calculated frequencies. These 'dull' characteristics of filters might be a convenient way to discuss perceived reverberance in dead conditions such as a recording studio or car cabin.

EVALUATION OF THE TIMBRES OF MUSICAL INSTRUMENTS

The dcGC filter is then applied to the evaluation the timbres of musical instruments. Similar methodology of analysis is used to that in the previous section, i.e., the outputs of an auditory filter are compared with the output of conventional band-pass filters.

Analysis by using dcGC filter

The sounds of musical instruments (clarinets, trumpets, and violins) are recorded in rather dead acoustics where the reverberation time is about 0.3 sec. Different instruments are used in the recording process, and ten waveforms for C major scale are prepared for each instrument. "Different" in the above sentence means the differences of instrument's maker, parts (reeds, mouthpieces, and the rest), and nuance that comes from how to perform. There are ten kinds of combinations of these differences for each instrument. Outputs of both dcGC and band-pass filters are compared by using statistical methods. In practice, a frequency-distribution (here, a mathematical function showing the number of instances in which a variable takes each of its possible values) is calculated and the kurtosis of this distribution, which indicates the degree of sharpness, is examined. Larger kurtosis corresponds to smaller fluctuation in the waveform, and vice versa.

The results are shown in Fig. 7 in which the horizontal axis indicates the filter number (which collectively cover the range 50 Hz to 16,000 Hz). Systematic changes of the kurtosis can be observed in the case of dcGC outputs. Especially at lower(the filter number from 1 to 3) and higher(14 or above) frequency ranges, the kurtosis value does not randomly interchange, and these frequency ranges include fundamentals and higher harmonics, respectively. On the other hand, the results of using

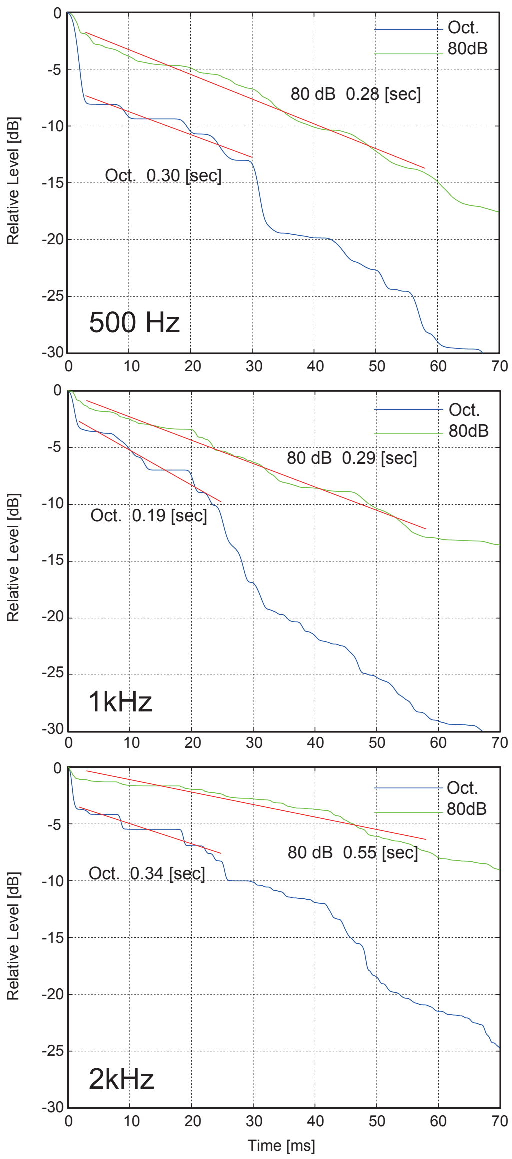


Figure 6: Reverberation curves of sparse impulse response.

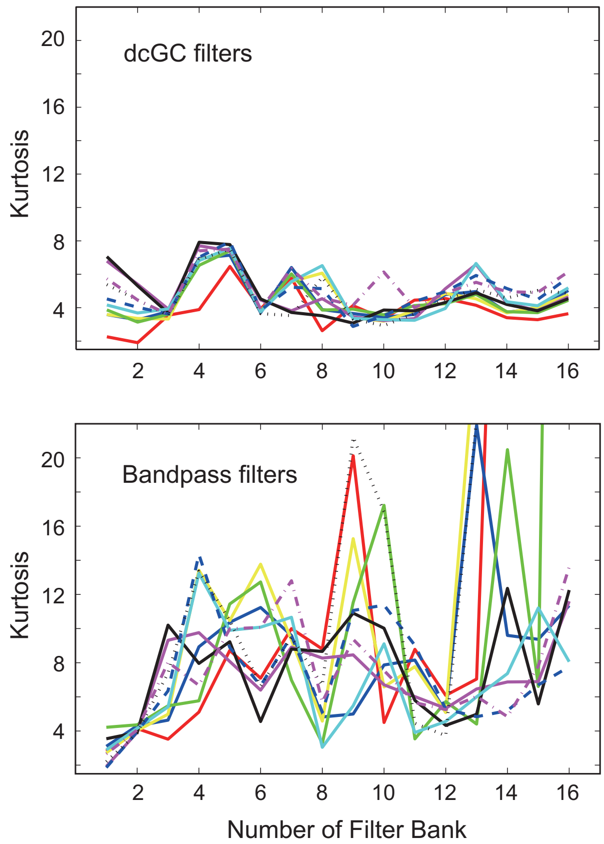


Figure 7: Kurtosis of histogram obtained by waveforms after passing through dcGC and conventional band-pass filters.

Table 3: Adjectives used in the semantic difference experiment.

adj. 1	:	smooth	↔	rough
adj. 2	:	clear	↔	cloudy
adj. 3	:	shiny	↔	non-shiny
adj. 4	:	soft	↔	hard
adj. 5	:	clear	↔	muffled
adj. 6	:	deep	↔	metallic
adj. 7	:	powerful	↔	poor
adj. 8	:	heavy	↔	light
adj. 9	:	loud	↔	calm

band-pass filters gave more random kurtosis value.

Subjective Evaluation

Subjective evaluation was carried out employing Multi-Dimensional Scaling (MDS) and Semantic Difference (SD) methods. Both methods were used to examine the similarity of stimuli and the characteristics of timbre. Fifteen subjects who had normal hearing ability were involved.

In the MDS experiment, degree of similarity was judged for randomly presented paired stimuli with seven-step scaling. In the SD method, nine pairs of adjective were introduced. These adjective pairs are shown in the Table 3.

The median of the data from all subjects was adopted as the representative value.

Comparison with Filter Output

Correlations between kurtosis of filter output signals and subjective evaluated values were examined. The results indicated that higher correlations were obtained with the dcGC filter for both MDS and SD methods. Table 4 shows an example of clarinet sound, and the results of using the SD method. The upper and lower tables correspond to the dcGC and conventional band-pass filters, respectively. The colored cells indicate that the correlation was higher than 0.7. For convenience, the values of correlation are drawn as graphs for each adjective.

In the low frequency bands of dcGC filters, such as 1 and 2, and in the high frequency bands (14 or above), systematic differences in the kurtosis can be observed. High correlations were observed in such frequency bands in Table 4. In contrast, low correlations were obtained in the middle range at where the kurtosis expressed rather random characteristics for different sounds.

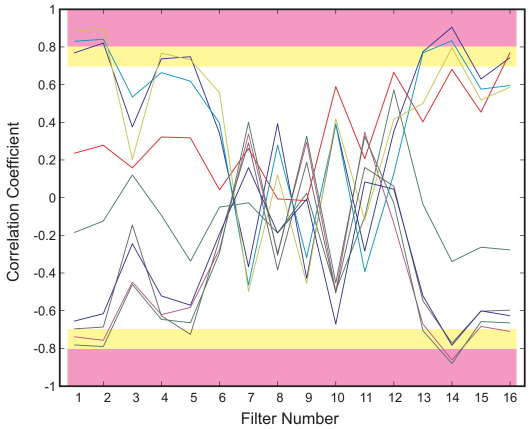
The results for adj. #1 indicated good correspondence with dcGC filter output for all three instruments. Especially in the case of clarinets(Table 4), the adj. #4 or #6 resulted in the higher correlations in many frequency bands.

The results obtained here indicated that it is possible to evaluate timbre effectively by using auditory filters such as dcGC. More detailed analysis of the results obtained and also examinations of other sounds are currently being carried out.

Table 4: Correlation coefficients of subjective evaluation (SD method) and Kurtosis of histogram of dcGC filtered waveform (Cl), in which the yellow and red cells indicate the coefficient is greater than 0.7 and 0.8, respectively. Nine lines in the left graph correspond to nine adjectives.

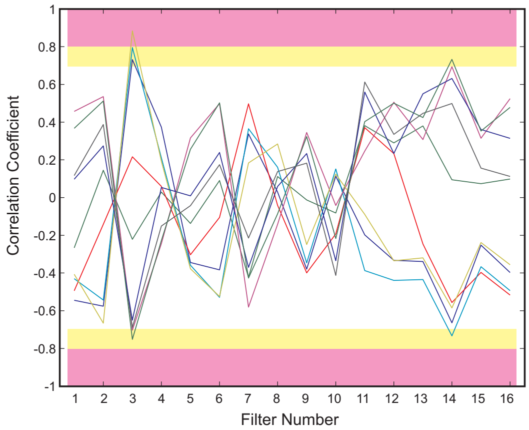
dcGC

Filter No.	adj. 1	adj. 2	adj. 3	adj. 4	adj. 5	adj. 6	adj. 7	adj. 8	adj. 9
1	0.768	-0.185	0.236	0.830	-0.738	0.885	-0.697	-0.656	-0.782
2	0.821	-0.124	0.278	0.841	-0.757	0.897	-0.687	-0.616	-0.790
3	0.376	0.121	0.158	0.534	-0.448	0.202	-0.145	-0.245	-0.461
4	0.737	-0.093	0.322	0.663	-0.621	0.767	-0.627	-0.522	-0.646
5	0.748	-0.337	0.317	0.619	-0.583	0.732	-0.726	-0.571	-0.664
6	0.342	-0.050	0.041	0.396	-0.275	0.557	-0.218	-0.192	-0.293
7	-0.366	-0.027	0.261	-0.464	0.338	-0.498	0.291	0.160	0.400
8	0.393	-0.188	-0.007	0.279	-0.295	0.120	-0.384	-0.187	-0.305
9	-0.428	0.023	-0.015	-0.318	0.296	-0.454	0.188	-0.008	0.326
10	0.390	-0.494	0.589	0.390	-0.498	0.417	-0.506	-0.671	-0.453
11	-0.283	-0.103	0.207	-0.392	0.347	-0.120	0.160	0.083	0.325
12	0.352	0.572	0.664	0.128	-0.137	0.416	0.060	0.045	-0.030
13	0.774	-0.035	0.403	0.769	-0.671	0.501	-0.547	-0.521	-0.706
14	0.905	-0.339	0.681	0.833	-0.862	0.796	-0.772	-0.784	-0.881
15	0.630	-0.263	0.454	0.576	-0.683	0.518	-0.603	-0.601	-0.657
16	0.743	-0.278	0.771	0.595	-0.710	0.587	-0.597	-0.627	-0.665



bandpass

Filter No.	adj. 1	adj. 2	adj. 3	adj. 4	adj. 5	adj. 6	adj. 7	adj. 8	adj. 9
1	-0.544	-0.267	-0.494	-0.430	0.457	-0.407	0.117	0.097	0.367
2	-0.576	0.145	-0.142	-0.543	0.536	-0.666	0.387	0.274	0.512
3	0.731	-0.221	0.216	0.796	-0.703	0.883	-0.688	-0.650	-0.751
4	0.373	0.029	0.060	0.211	-0.245	0.192	-0.151	0.054	-0.228
5	-0.344	-0.137	-0.303	-0.359	0.316	-0.378	-0.043	0.009	0.256
6	-0.383	0.090	-0.105	-0.529	0.498	-0.524	0.175	0.239	0.502
7	0.337	-0.423	0.497	0.365	-0.581	0.183	-0.214	-0.370	-0.427
8	0.029	0.112	-0.041	0.161	-0.142	0.284	0.138	0.059	-0.086
9	-0.379	-0.013	-0.399	-0.346	0.345	-0.249	0.183	0.233	0.323
10	0.113	-0.082	-0.193	0.151	-0.042	0.124	-0.412	-0.335	-0.217
11	-0.198	0.382	0.371	-0.387	0.247	-0.089	0.612	0.559	0.403
12	-0.333	0.290	0.233	-0.439	0.506	-0.335	0.335	0.234	0.501
13	-0.339	0.380	-0.246	-0.435	0.308	-0.320	0.447	0.550	0.425
14	-0.664	0.094	-0.556	-0.733	0.694	-0.585	0.499	0.631	0.732
15	-0.252	0.074	-0.397	-0.367	0.315	-0.238	0.156	0.361	0.353
16	-0.397	0.098	-0.517	-0.494	0.524	-0.357	0.112	0.314	0.478



CONCLUDING REMARKS

In this report, we have applied an auditory filter for the evaluation of similarity of sound fields, the reverberation process, and timbre of several musical instruments. The Gammatone and Gammachirp filters with dynamic compressive characteristics are used instead of conventional band-pass filters.

Throughout the examination, it was demonstrated that the proposed methods using auditory filters showed higher correlation with subjective evaluation than did conventional methods. In the examination of reverberation process using the room impulse response, the decay times obtained by the integrated response after passing through an auditory filter were similar to the early decay times, which correspond to the perceived reverberance of the sound field.

The results shown in this report strongly suggest the superiority of auditory filters in the qualitative evaluation of sounds and sound fields. Systematic examinations for other conditions are currently under discussion.

ACKNOWLEDGMENT

The authors would like to thank Dr. Irino of Wakayama University for providing the "*dynamic compressive Gamma-Chirp*" filter software.

REFERENCES

1 ISO 3382 *Acoustics - Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, (International Standard Organization, Geneva, 1997)
2 M. Barron and A. H. Marshall, "Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *J. Sound Vib*. **77**, 211–232 (1981).
3 J. S. Bradley, "Predictors of Speech Intelligibility in Rooms," *J. Acoust. Soc. Am*. **80**, 837–845 (1986).
4 A. Abdou and R. W. Guy, "Spatial Information of Sound Fields for Room-Acoustics Evaluation and Diagnosis," *J. Acoust.Soc. Am*. **100**, 3215–3226 (1996).
5 R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.*, **59**, 640-655, 1976.
6 T. Irino and R. D. Patterson, "A dynamic compressive gammachirp auditory filter-bank," *IEEE Trans. Audio, Sp. Lang. Process*., **14**, 2222–2232, 2006.
7 T. Irino, "A computational theory of the peripheral auditory system," *Tech. Rep. IEICE*, **95**, 23–30, 1996.
8 T. Irino and M. Unoki, "An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp," *J. Acoust .Soc. Japan*(E), **20**, 397-406, 1999.
9 B. C. J. Moore, "An Introduction to the Psychology of Hearing, 5th ed.," *Academic Press, 2003*.
10 M. R. Schroeder, "New Method of Measuring Reverberation Time" , *J. Acoust. Soc. Am.*, **37**, 409-412, 1965.