ISSN 2229-5518

Effect of Gammatone auditory-filters on speech intelligibility

Boucherit Ismail [1,2] , Guerti Mhania [2]
[1]Electrical engineering Department, Yahia Fares University, Médéa, Algeria
[2]Signal and Communications Laboratory, Ecole Nationale Polythechnique, Algiers, Algeria
boucheritw@yahoo.fr , mhania.guerti@enp.edu.dz

Abstract— Auditory filters are one of the topics that have been most studied in psychoacoustics. Researchers have used different approaches and many of them have been considered as possible basis for this study. The auditory filter bank can be seen as a bank of band-pass filters that divides the very often broadband input signal into multiple narrowband output signals. The analytic description of the shape of the auditory filters has improved through the years. Our study investigates how the shape of the auditory filters affects the intelligibility of speech in noise. We compare the performance of filters across noises aiming to find the filter that performed the best. However, the difference in results was not found to be statistically significant. Overall, the results provide a framework for using of the auditory filters.

Index Terms— Auditory filter, Gammatone filter, noise, speech intelligibility.
.

1 Introduction

Over the past half century, the auditory system has been the center of attention of exhaustive study. Knowledge of physiology, psychology and engineering has provided the opportunity for the formation of models which until a certain extension tries to explain and imitate the hearing mechanisms. Auditory models are used to different purposes, such as help on tracking problems on hearing or speech fields [1].

Frequency analysis is the first and most basic stage of auditory processing that takes place in the cochlea. This stage has often been likened to a bank of overlapping bandpass filters, and a great effort has gone into characterizing the magnitude response of these auditory filters [2].

The Gamma-Tone Filter (GTF), introduced by Johannesma to describe the cochlear nucleus response, has been adopted as the basis of a number of successful auditory modeling efforts. The Gammachirp filter (GC) was derived by Irino and Patterson [3] as a theoretically most favorable auditory filter which includes a chirp parameter in the impulse response of each filter. It is constructed by a cascade of a GTF, with an asymmetric compensation filter. It existe another cascade filter system, namely the Differentiated All-Pole Gammatone Filter (DAPGF) and One-Zero Gammatone Filter (OZGF), filter responses that provide a robust base for modeling cochlea transfer functions [4].

In a sensorineural impaired cochlea, auditory filters are generally broader than the normal and are in many cases abnormally symmetrical. Processing through these abnormal filters may produce a smearing of spectral detail in the internal representation of acoustic stimuli. Differences in amplitudes between peaks and valleys in the input spectrum may be reduced, making it more difficult to locate spectral prominence (i.e., formants) which provides crucial cues to speech intelligibility [5]. In our study in [6] we evaluated the hypothesis that DAPGF filters result in higher performance than conventionally-used Butterworth filters at low frequencies , but only for vowels. So we extend here the study to the consonants and for other Gammatone filters.

The major research question involved in auditory filter is : Are these filters actually useful for speech processing in noise? If so, how can we demonstrate their utility, and make use of them? . So our goal is to focus and look for the effect of each Gammatone filter on speech analysis.

2 Auditory Filter

2.1 Gammatone Filter

The Gammatone Filter, is one such filter. Its name is due to the nature of its impulse response, which is a gamma envelope modulated by a tone carrier centered at f_c Hz [7].

$$g(t) = at^{n-1}\exp(-2\pi bB(f_c)t)\exp(j2\pi f_c t) \qquad (1)$$

This function can be seen as a pure tone with a gamma function as an envelope, hence the name Gammatone. $B(f_c)$ is the Equivalent Rectangular Bandwidth (ERB) of the center frequency f_c

$$B(f) = 01039.f + 24.7 \qquad (2)$$

The GTF is intrinsically nearly symmetric in the pass band, while physiological measurements show a significant asymmetry in the biological cochlea transfer function. In addition, it is not easy to use the parameterization of the GTF to model level-dependent changes in the auditory filter [8].

Three approximations of the gammatone filter, will be described in this section.

2.2 Gammachirp Filter

For The complex impulse response of the GC [9] is given as

$$g_C(t) = at^{n-1}\exp(-2\pi bB(f_r)t).\exp(j2\pi f_r t + jc\ln t + j\phi) \qquad (3)$$

where time t > 0; a is amplitude; n and b are parameters defining the envelope of the gamma distribution; c is the chirp factor; fr is the asymptotic center frequency, Φ is the initial phase; and ln (t) is the natural logarithm of time. When c1=0,

ISSN 2229-5518

equation (4) reduces to the complex impulse response of the Gammatone filter. The Fourier magnitude spectrum of the gammachirp filter is :

$$|G_c(f)| = a_\Gamma . |G_T(f)| . \exp(c_1 \theta(f)) \quad (4)$$

$$\theta(f) = \arctan((f - f_r)/b\ ERB(f_r)) \quad (5)$$

$|G_T(f)|$ is the Fourier magnitude spectrum of the Gammatone filter, and $\exp(c_1\theta(f))$ is an asymmetric function since θ is an antisymmetric function centered at the asymptotic frequency, fr.

Figure 1 illustrates a set of GC filters, $|G_c(f)|$, with varying asymmetry, by multiplying a fixed Gammatone filter, $|G_T(f)|$, together with a set of asymmetric functions, $\exp(c_1\theta(f))$. The fixed, Gammatone auditory filter $|G_T(f)|$ is shown by the solid line in the lower part of the figure. The low-pass asymmetric functions (LP-AF) are shown by the fan of dashed lines that pass through the same origin as the gammatone filter. c1 was varied from 0 to –2.

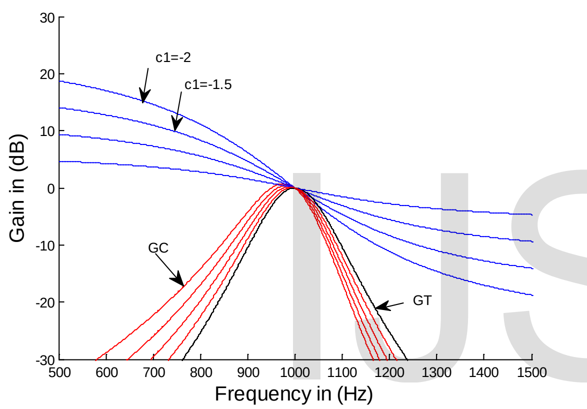


Fig. 1. A set of analytic Gammachirp filters, GC (solid lines); left ordinate,produced by cascading a fixed Gammatone filter, GT, with a low-pass asymmetric function, LP-AF, whose range varies with level (dashed lines); right ordinate.

2.3Differentiated All-Pole Gammatone Filter

The DAPGF response is attractive because it exhibits certain characteristics suitable for modeling a variety of auditory data : level-dependent gain, linear tail for frequencies well below the centre frequency, asymmetry, etc. The DAPGF can be considered as a cascade of (N–1) identical Low Pass biquads (i.e. a (N–1) th-order APGF) and a rightly scaled Band Pass biquad [4]. The DAPGF transfer function is :

$$H_{DAPGF}(s) = \left(w_0^{2N-2}\right)/\left([s^2 + (w_0/Q)s + w_0^2]^{N-1}\right) \times \left(w_0 s\right)/[s^2 + \left(w_0/Q\right)s + w_0^2]$$

$$= \left(w_0^{N-1}s\right)/\left([s^2 + w_0/Qs + w_0^2]^N\right)$$

K1 = ω02(N-1) and K2 = ω0.

2.4 One Zero Gammatone Filter

The OZGF transfer function, described in (4), is derived from the GTF by discarding all but one of its zeros, with that zero lying anywhere on the real axis. From the implementation point of view, an Nth-order OZGF can be considered as the composition of two individual transfer functions; a cascade of (N–1) identical lowpass (LP) biquadratic filters (i.e. a (N–1) th-order APGF) coupled with an appropriately scaled lossy bandpass (BP) biquadratic filter i.e. a 2-pole, 1-zero resonant transfer function [8]:

$$H_{OZGF}(s) = \left(w_0^{2N-2}\right)/[s^2 + \frac{w_0}{Q}s + w_0^2]^{N-1} \times \left(w_0(s + w_z)\right)/\left([s^2 + \frac{w_0}{Q}s + w_0^2]\right)$$

$$= \left(w_0^{N-1}(s + w_z)\right)/\left([s^2 + \frac{w_0}{Q}s + w_0^2]^N\right)$$

3 Acoustic Analysis

In order to measure the effect of noise on different regions of the spectrum, we calculated the critical-band spectral difference between the clean and noisy vowels and consonants in three different bands, the first and the secand correspond to the F1 and F2 regions respectively. The vocalic segment was first filtered through a 21-channel filter bank implemented using the three auditory filters, the center frequencies of the filter bank were chosen according to critical-band spacing. (Table 1) [10].

The spectral distance between the clean and noisy signals was then computed for three different frequency bands spanning the 0–8 kHz bandwidth, using a Normalized Euclidean Distance (NED) metric of the filter bank energies [11].

The three bands considered include a low-frequency (LF) band straddling the 0–1 kHz region, a middle-frequency (MF) band straddling the 1–2.7 kHz region and High-frequency(HF) band straddling the 2.7–8 kHz region. F1 typically resides in the LF band, and F2 resides in the MF band. Spectral difference measurements were made every 10 ms:

TABLE 1
Critical Bands

Critical Bands	Frequency (Hz)		
	Basse	Haute	Centrale
1	1	100	50
2	100	200	150
3	200	300	250
4	300	400	350
5	400	510	450
6	510	630	570
7	630	770	700
8	770	920	840
9	920	1080	1000
10	1080	1270	1175
11	1270	1480	1370
12	1480	1720	1600
13	1720	2000	1850
14	2000	2320	2150
15	2320	2700	2500
16	2700	3150	2900
17	3150	3700	3400
18	3700	4400	4000
19	4400	5300	4800
20	5300	6400	5800
21	6400	7700	7000

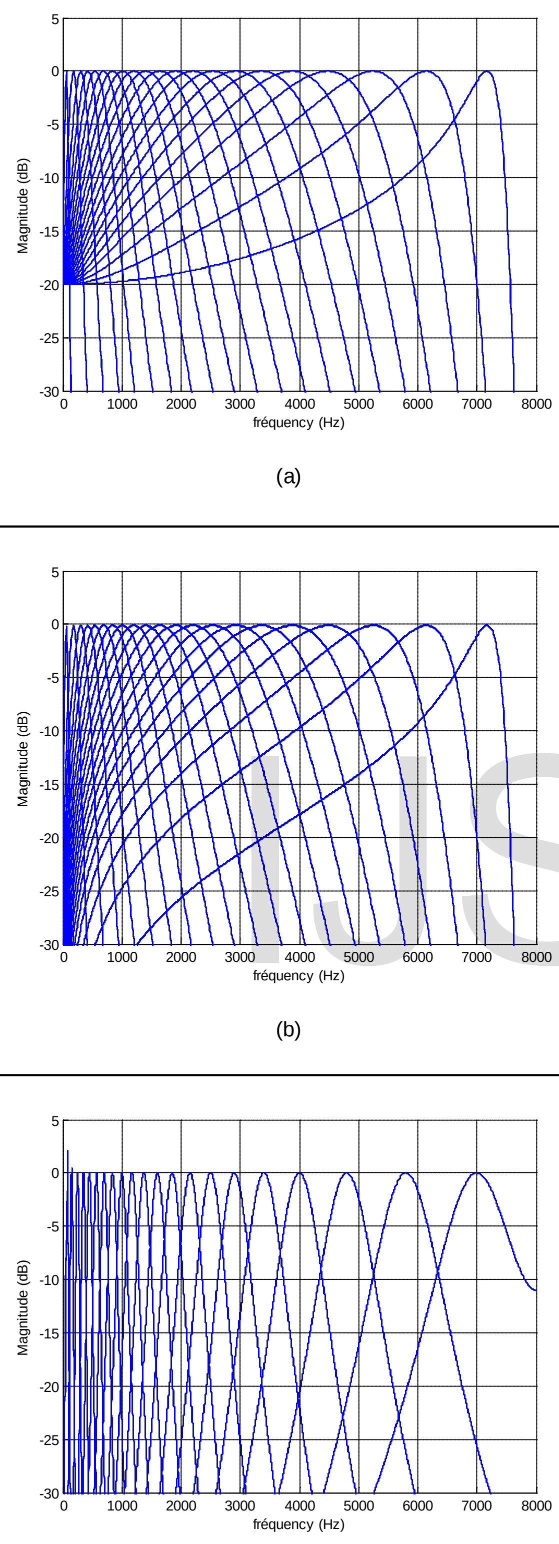


Fig. 2. A series of bandpass GTF derives covering a range of (0-8 KHz) for specific signal parameters (a) DAPGF (b) OZGF (c) GC.

$$LF = \sqrt{\sum_{i=1}^{9}\left(F_i^c - F_i^n\right)^2} \Big/ \sqrt{\sum_{i=1}^{9}\left(F_i^c\right)^2} \tag{5}$$

$$MF = \sqrt{\sum_{i=10}^{15}\left(F_i^c - F_i^n\right)^2} \Big/ \sqrt{\sum_{i=10}^{15}\left(F_i^c\right)^2} \tag{6}$$

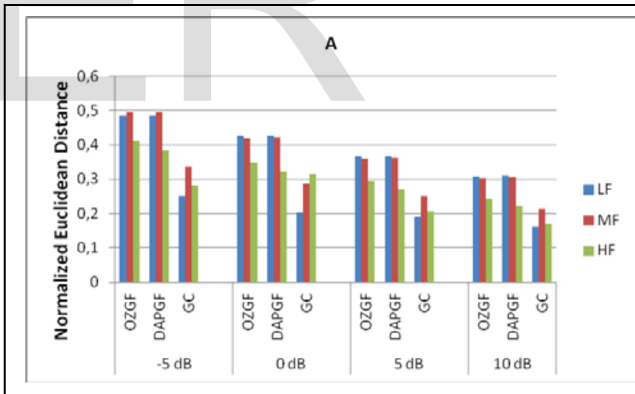
$$HF = \sqrt{\sum_{i=16}^{21}\left(F_i^c - F_i^n\right)^2} \Big/ \sqrt{\sum_{i=16}^{21}\left(F_i^c\right)^2} \tag{7}$$

Where : 1) using a number F_i^c denotes the i-th filter-bank energy of the clean vowel, 2) F_i^c denotes the i-th filter-bank energy of the noisy vowel.

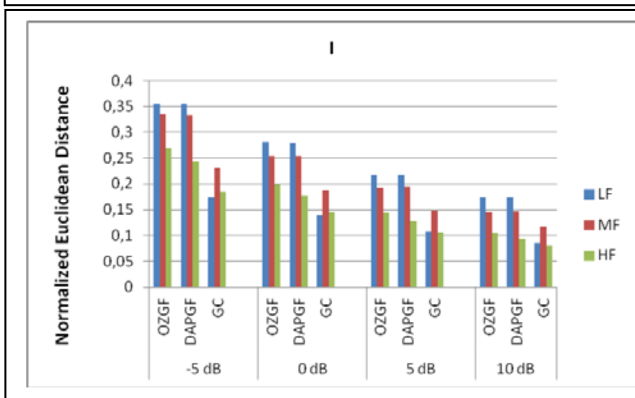
Speech Corpora Used

Depending on the nature of input excitation, speech sounds are formed which are called phonemes. If the input excitation is periodic, then voiced sounds are produced and if the input excitation is masked with noise, then unvoiced sounds are produced.

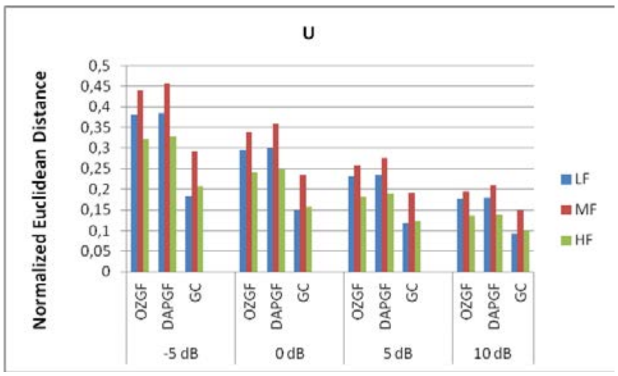
In this analysis we used TIMIT database which consists of phonemes sampled at 16KHz. Vowels can be continued for a relatively longer duration than consonants and for this reason both vowels and consonants are taken for the spectral study. amongst all the vowels, only the vowels /a/, /i/ and /u/ are considered to represent the three categories of vowels.



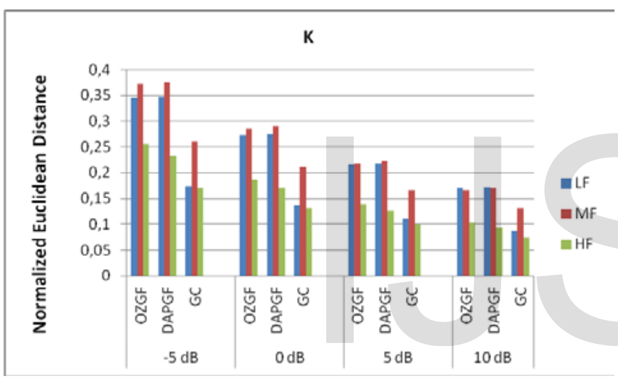
(a)



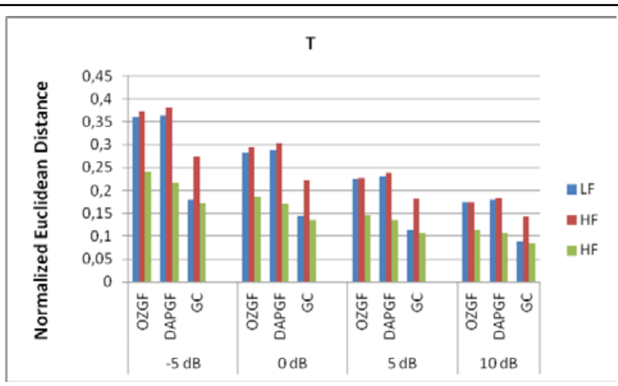
(b)



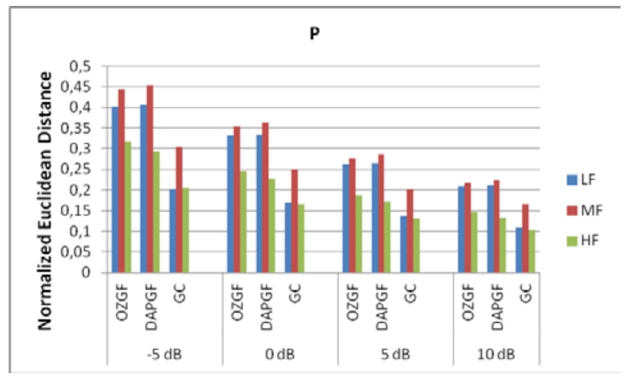
(c)



(d)



(e)



(f)

Fig. 3. Spectral distance results for (a) vowel [a], (b) vowel [I], (c)vowel [U], (d) consonnant [k], (e) consonnant [T],(f) consonnant [P].

Signal processing

Only Speech-Shaped Noise (SSN) was considered because in [11] results indicate that this type of noise is alike to the multi-talker babble real noise and both are perceptually and acoustically equivalent. It means that the two types of noise multi-talker babble and SSN affected the vowel and consonant spectra in a similar way. The SSN here was generated by applying a second-order Butterworth lowpass filter (cutoff frequency 1100 Hz) to white Gaussian noise (WGN), so that it's spectral shape is similar to that of speech waveforms.

During the current simulation for a specific signal the filter bank parameters are fixed. From psychoacoustics point of view, it is known that this is not correct. The filters instantaneously follow input level changes and change shape and bandwidth accordingly. The implementation of such an algorithm for the compressive GC ,DAPGF and OZGF filter is currently not available, but should become available in the future. These parameters give a response of the filter corresponding to high level input of speech and are shown in table 3. (see figure 2).

Results

According to the results illustrated in Figure 3, we note that the third region (HF) has the least value of NED in the three filters across all SNR because the SSN affects especially the first and the second region hence the WGN was filtered by lowpass filter. The SSN affected these two frequency bands (LF and MF) of the vowel spectra identically as the frequency bands of the consonant.

Comparing the performance of filters in reference to the noisy vowels and consonants, we observe that both DAPGF and OZGF filter have approximately the same values of NED at the first and the second region, but change in the third one, as in 0 dB for vowel /A/ and for consonant /K/. For vowel /I/ the first region has the greatest value of NED unlike the other phonemes. So the noise affects more the F1 formant in the case of /I/.

We notes that the GC filter have the less spectral values in all regions for all vowels and consonants. We obtained the same results for the other vowels and consonants.

5 Conclusion

Further research aimed, at improving the recognition of consonants and vowels because of their importance for speech intelligibility.

The results of the spectral difference measurements are not only important for understanding vowel and consonant perception in noise, but are also important for the development of a new filter bank in which individual frequency bands are treated differently.

Comparing results of our simulation, we conclude that the lower slopes of auditory filters can affect independently the intelligibility of speech because each slope can contribute separately to filter widening. This is the reason why it is so important to carefully choose a category and type of a filter during a design process.

The motivation for studying auditory filter is to improve the front end speech processing. Finally, of the three auditory filter types examined, the GC performed the best in noise.

References

[1] L. C. Araujo and T. N. Magalhaes, "A Brief History of Auditory Models,"

[2] A. J. Oxenham and T. Dauand "The effect of parametric variations of cochlear implant processors on speech understanding," *J. Acoust. Soc. Am.*,vol. 108, no. 2, pp. 790- 802, August 2000.

[3] T. Irino and R. D. Patterson, "Towards a measure of auditory-filter phase response," *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3169-3178, December 2001.

[4] A.G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "Practical Gammatone-like Filters for Auditory Processing," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, 15 Pages, 2007.

[5] I. J. Yang, F. Luo, and A. Nehorai, "Spectral contrast enhancement : Algorithms and comparisons," *Elsevier, Speech Communication*, vol. 39, pp 33–46, 2003.

[6] I. Boucherit, and M. Guerti, "Using the DAPGF Filter for Speech Analysis : Application for cochlear implants ," *Proc. 24th IEEE International Conference on Microelectronics*, Decempber 17-20, Algers, Algeria, 2012.

[7] R. D. Patterson, K. Robinson, J. W. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex sounds and auditory images," in Auditory Physiology and Perception, Y. Cazals, L. Demany, and K. Horner, Eds., pp. 429-446. Pergamon, Oxford, 1992.

[8] A. G. Katsiamis , E. M. Drakakis, and F. Richard Lyon, "Introducing the Differentiated All-Pole and One-Zero Gammatone Filter Responses and their Analog VLSI Log-domain Implementation," In *Proc. 49th IEEE International Midwest Symposium on Circuits and Systems*, vol 1, pp. 561–565, Augus 6-9, San Juan, Puerto Rico, 2006 .

[9] R. D Patterson, M. Unoki and T. Irino, "Extending the domain of center frequencies for the compressive gammachirp auditory filter," *J. Acoust. Soc. Am.*, vol. 114, no. 3, pp. 1529–1542, September 2003.

[10] E. Zwicker and H. Fastl, "Psychoacoustics: Facts and Models," Springer Verlag: Berling, 1999.

[11] G. K. Parikh, and P C. Loizou, "The effect of noise on the spectrum of speech," *J. Acoust. Soc. Am.*, vol. 118, no. 6, pp. 3874–3888, December 2005.