

History and Future of Auditory Filter Models

Richard F. Lyon
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
Email: dicklyon@acm.org

Andreas G. Katsiamis
Toumaz Technology, Ltd.
Bldg. 3, 115 Milton Park
Abingdon OX14 4RZ, UK

Emmanuel M. Drakakis
Department of Bioengineering
The Sir Leon Bagrit Centre
Imperial College London
South Kensington Campus
London SW7 2AZ, UK

Abstract—Auditory filter models have a history of over a hundred years, with explicit bio-mimetic inspiration at many stages along the way. From passive analogue electric delay line models, through digital filter models, active analogue VLSI models, and abstract filter shape models, these filters have both represented and driven the state of progress in auditory research. Today, we are able to represent a wide range of linear and nonlinear aspects of the psychophysics and physiology of hearing with a rather simple and elegant set of circuits or computations that have a clear connection to underlying hydrodynamics and with parameters calibrated to human performance data. A key part of the progress in getting to this stage has been the experimental clarification of the nature of cochlear nonlinearities, and the modelling work to map these experimental results into the domain of circuits and systems. No matter how these models are built into machine-hearing systems, their bio-mimetic roots will remain key to their performance. In this paper we review some of these models, explain their advantages and disadvantages and present possible ways of implementing them. As an example, a continuous-time analogue CMOS implementation of the One Zero Gammatone Filter (OZGF) is presented together with its automatic gain control that models its level-dependent nonlinear behaviour.

I. INTRODUCTION

Over about the last half century, many auditory filter models have been developed, analyzed, and applied to a variety of hearing-related problems. We review several lines of development, and several criteria that filter models might try to satisfy, and show how the one-zero gammatone filter (OZGF) and the pole-zero filter cascade (PZFC) models achieve these desired properties.

Transmission-line models of wave propagation on the basilar membrane go even further back, but the basis for approximating these systems as filter cascades was not made clear until after Zweig, Lipes, and Pierce showed how to apply the WKB approximation in their 1976 “Cochlear Compromise” paper [1]. They ended up with a circuit model similar to the old transmission-line models of Wegel and Lane [2] and Peterson and Bogert [3], but the method that they explained led directly to a wider class of filter-cascade models of the cochlea, “cascade filterbanks” (as opposed to conventional parallel filterbanks) [4]. Many analog VLSI implementations are based on such cascades; others are based on parallel filterbanks, where each channel is some variant of a gammatone filter, the other popular family of auditory filter models [5].

II. WHAT IS AN AUDITORY FILTER?

The auditory filters that we consider here include both those motivated by psychoacoustic experiments, such as detection of tones in noise maskers, as well as those motivated by reproducing the observed mechanical response of the basilar membrane or neural response of the auditory nerve. These are not necessarily going to lead to the same best filter models, but it is one thesis of this work that a single form of model can do a good job for all of these, and thereby provide a good basis for a machine-hearing system. Since there’s a lot of neural processing between the cochlea and our psychoacoustic perceptions, it would not be surprising if the best parameters were different between these types of models, but it seems likely that the linear and nonlinear filter due to the cochlea plays a big enough role in perception that we may find one set of parameters is adequate, at least for a range of machine-hearing applications.

In 1958 already, Green [6] summarized the state of auditory filter models and the concept of the critical band. Measurements at that time were not adequate to determine much more than bandwidth, and his comparisons of different filter shapes (rectangular, simple resonance, and Gaussian) did not yet lead to an understanding of how to determine better-fitting shapes from psychophysical data. He said that “psychophysical data were not used as critical evidence” in the debates among theorists about how auditory filtering was accomplished, and suggested that progress could be made by more psychophysical studies to complement the physiological and anatomic data being used. Since that time, psychophysical experiments, especially on detection of sinusoids in notched-noise maskers, have driven progress in fitting auditory filter shapes for human hearing, giving rise to the roex and gammatone families of filters, and good fits that include their nonlinear level dependence [7].

Besides the level dependence, other nonlinearities generate distortion products, or combination tones, as was discussed with respect to the cochlear resonance theory by Barton over a hundred years ago [8], who said, “there does not exist in the air any clearly sensible pendular vibration corresponding to the combinational tone, and we must conclude that such tones, which are often powerfully audible, are really produced in the ear itself.” The incorporation of such instantaneous nonlinearities, as well as of the level-dependent quasi-linear type

of nonlinearity, has been an ongoing theme in the development of auditory filter models.

By auditory filter we mean to include the whole range, from simple linear symmetric critical band concepts through models of nonlinear wave propagation in the cochlea, but especially those models that can be efficiently implemented and applied to problems in human and machine hearing, or to other signal analysis problems.

III. RESONANCE AS A PRIMITIVE AUDITORY FILTER

The resonance theory of cochlear function, developed by Helmholtz but widely discussed for two hundred years before that [9], [10], uses a damped harmonic oscillator, or simple resonance—a “single-tuned filter” that can be represented with one complex-conjugate pair of poles, and optionally a zero at DC. While neither (all-pole nor pole-zero) simple resonance is exactly symmetric, they are often approximated by Terman’s symmetric “universal resonance curve” [11], [12]:

$$|H(f)| = \frac{1}{\sqrt{1 + (\Delta f/\alpha)^2}}$$

where α is the frequency deviation to the half-power point.

The simple resonance has been frequently tried, and often rejected, as a model of auditory filtering. For example, the “single-tuned filter” or “universal resonance curve” has been explicitly applied to critical bands and auditory models by Schafer et al. [13], by Tanner, Swets, and Green [14], and by Patterson [15]; Patterson found the skirts of this filter shape to be not steep enough, and the peak too sharp.

A cascade of several such resonances is a good approximation—at least, not too far from the center frequency—for many of the gammatone-like and filter-cascade-family models. Patterson had earlier used the square of that symmetric function [16], effectively approximating an order-2 gammatone. With multiple resonances in cascade, the skirts get steeper faster than the peak gets sharper, so this structure addresses the basic limitation of the simple resonance shape.

A cascade of very many simple resonances makes a filter that approaches a Gaussian transfer function shape—that is, the Gaussian is the limit of gammatone filters of high order. Tanner et al. had introduced that shape as a potential auditory filter [14], and Patterson [15] observed that the skirts of the Gaussian fall much too fast to be realistic; he thereby defined the extremes between which a good auditory filter shape was to be sought. All of the filter models we address in this paper fall into this middle ground, between a resonance and a Gaussian.

IV. THREE LINES OF AUDITORY FILTER DEVELOPMENT

Three lines of auditory filter development have led to three widely-used families of filter models: (1) the rounded exponential (roex) family; (2) the gammatone family, including gammachirp and all-pole variants; and (3) the filter cascades, both all-pole and pole-zero variants. Each family has good properties, and applications for which it provides a useful solution.

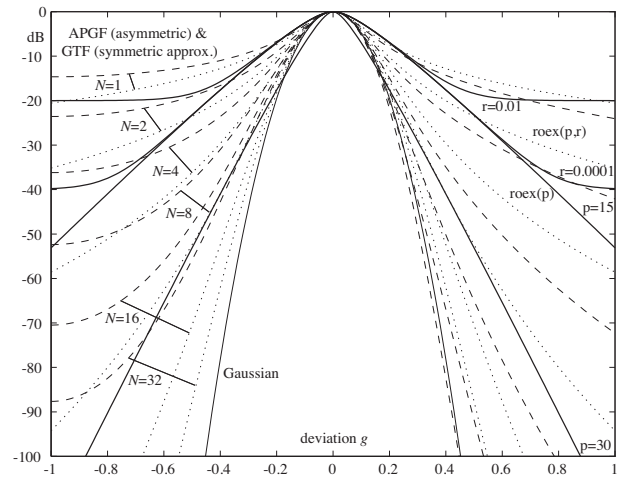


Fig. 1. A range of auditory filter model shapes, from simple resonators ($N=1$) to Gaussian, including symmetric (dotted) and asymmetric (dashed) gammatone-family, and symmetric roex filters; all illustrated filter shapes except the $p = 30$ roex are matched for curvature at the peak, for nearly-equal 3 dB bandwidth. In the semi-log plot, the Gaussian is parabolic and the roex(p) has nearly straight-line skirts.

An important use of auditory filter models is to support applications in which full filterbanks efficiently process real sounds; these applications motivate and benefit from the filter-cascade family of models, since these models minimize the total computational complexity of a filterbank, as opposed to the complexity of a single auditory filter channel. Their structural efficiency is the reason that the filter cascades have been the basis for most work in analog VLSI hearing models.

The roex family is useful mostly as a descriptive model, a way to parameterize and describe the shape of an auditory filter’s magnitude transfer function; it has no corresponding phase, no time-domain equivalent, and no “runnable” implementation.

The gammatone-family filters bridge the other two families, being useful as filter shape descriptions, but also implementable as real analog or digital filters to process sounds. The basic gammatone, while very popular, is not particularly accurate or controllable in the ways we would want, but its variations, the gammachirp filter (GCF) [17] and the all-pole and one-zero gammatone filters (APGF and OZGF) [18], [19], are much better in these respects.

Nonlinear extensions of these families of linear filter models are also very important.

V. A FEW REPRESENTATIVE AUDITORY FILTER MODELS

Within the three families, we delineate nine specific filter models to discuss further, to see how they rate on various measures, and to position the OZGF and the PZFC as the most effective auditory models for machine-hearing applications.

A. Three roex Filters

The rounded exponential or roex filters can be seen as an effort to turn the “cartoon” of a triangular filter shape into something quantitatively reasonable, by rounding the peak and

specifying the skirt shape in terms of frequency deviation from the peak frequency.

The roex(p) filter has just one shape parameter, essentially a bandwidth. The roex(p,r) adds a parameter to control the skirt shape (so does the roex(p,t), but it's not significantly different nor widely used, so we skip it). Finally, to get asymmetry, different p parameters for the low side and high side allow control of asymmetry in the roex(pl, pu, r) filter.

There are further roex variations that we won't discuss specifically, such as the roex(p, w, t) and a two-sided version of it with six parameters. These variations can provide more shape control, but are not qualitatively different from the others in their properties. Rosen et al. [20] have expressed the opinion that these parameterizations may be *too* flexible.

B. Four Gammatone-Family Filters

The gammatone filter (GTF) has been hugely popular, mostly due to its simple description in the time domain as a gamma-distribution envelope times a tone. It has been implemented (made runnable) through a number of methods and approximations, but usually not in the most straightforward way based on its Laplace-domain pole-zero decomposition, since that decomposition was not widely understood until at least the mid 1990s. The GTF has an inherently very nearly symmetric frequency response, which is not a great match to auditory data, but has better peak/skirt shape than the roex filters.

The gammachirp filter (GCF) [17] is a generalization of the GTF that allows a realistic and controllable frequency-domain asymmetry, and a corresponding realistic time-domain "chirping", or a "glide" in the instantaneous frequency of its impulse response. But it does not have a pole-zero decomposition, and needs other approximations in its implementation.

The all-pole gammatone filter (APGF) is another approach to providing a realistic asymmetry, while at the same time simplifying the Laplace-domain description and implementation of the GTF. The APGF has been used as an approximation for implementing the GTF [21], [22], but has advantages of its own [19].

The one-zero gammatone filter (OZGF) is a slight generalization of the APGF that by adding a single real zero in the Laplace domain achieves good control of the low-frequency tail shape [19], [23].

C. Two Filter Cascades

The all-pole filter cascade (APFC) [19], [24] is the popular basis for silicon cochlea work; it is closely related to the all-pole gammatone filter. This type of cascade typically has not enough high-side sharpness, and has an unrealistically long delay if tuned for reasonable frequency-domain shape.

The pole-zero filter cascade (PZFC, the "two-pole, two-zero sharper" filter stage in [4]) has a much more realistic response in both time and frequency domains, due to its closer similarity to the underlying wave mechanics, and is not much more complicated. The additional degrees of freedom from the placement of a zero pair near the pole pair allows the

tailoring of response properties, such as the delay and high-side steepness of the filter, while retaining the other desirable features of the all-pole filters.

VI. TEN GOOD PROPERTIES FOR AUDITORY FILTER MODELS

Auditory filter models will ideally fulfill a variety of roles, requiring a variety of good properties, some of which include:

1. Simplicity of description: different types of filter models are simple in different domains.
2. Bandwidth control: the zero-order feature of an auditory filter is its bandwidth, which should be modeled as a function of the characteristic frequency of the cochlear place that it represents, and also as a function of sound level.
3. Realistic and controllable relationship between peak shape and skirts: after bandwidth, the first-order shape feature of an auditory filter is how rapidly the response falls off near the band edges, ideally with level dependence.
4. Filter shape asymmetry: data show that the filter skirt on the high-frequency side of CF is usually steeper, at least for high CFs, than the skirt on the low-frequency side of CF, though some models are inherently symmetric.
5. Peak gain variation: physiological data on cochlear mechanics show the cochlear mechanical filter's peak gain varying with signal level, providing a form of automatic gain control; some models vary shape but not gain.
6. Stable low-frequency tail: when the parameters of an auditory filter are varied with signal level, the response of the low-frequency tail of the filter will ideally not change much, again, to match the input-output relations seen in physiological data.
7. Ease of implementation as digital filters: in order to make a good digital filter, the model either needs to be described in terms of poles and zeros, or convertible to such a description, or approximated by such a description.
8. Connection to underlying traveling-wave hydrodynamics: other than the cascades, most filter models are just phenomenological, or descriptive of abstract filters. But the filter-cascade family was developed to connect with the mathematics of filtering by wave propagation, via the WKB method [4].
9. Good impulse-response timing and phase characteristics: for comparison with physiological measurements, across a

TABLE I

SCORING VARIOUS AUDITORY FILTER MODELS ON THE TEN CRITERIA. THE DOMAINS OF SIMPLE DESCRIPTION ARE FREQUENCY DOMAIN (FD), TIME DOMAIN (TD), LAPLACE (POLE-ZERO) DOMAIN (LD), AND LAPLACE PER STAGE (LD/S). THE * REPRESENTS PARTIAL CREDIT.

	roex			gammatones				cascades	
	(p)	(p,r)	(pl,pu,r)	GTF	GCF	APGF	OZGF	APFC	PZFC
1. Simple	fd	fd	fd	td	td	ld	ld	ld/s	ld/s
2. BW control	+	+	+	+	+	+	+	+	+
3. Peak/skirts	-	*	*	+	+	-	+	*	+
4. Asymmetry	-	-	+	-	+	+	+	+	+
5. Gain variation	-	-	-	-	*	+	+	+	+
6. Stable tail	-	+	+	-	*	+	+	+	+
7. Runnable	-	-	-	+	*	+	+	+	+
8. Waves	-	-	-	-	-	-	-	+	+
9. Impulse resp.	-	-	-	-	+	+	+	-	+
10. Dynamic	-	-	-	-	*	+	+	+	+

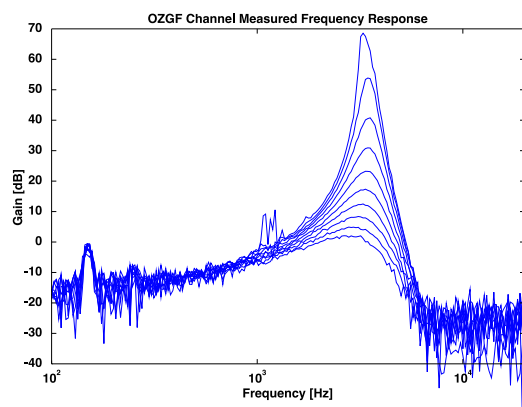


Fig. 2. OZGF circuit tunability across Q or gain values

range of levels, details such as zero-crossing times can be diagnostic of whether the model is faithful to the mechanics.

10. Dynamic: in addition to being parameterized by level, the filter will ideally be dynamically variable, so that it can be used for processing sounds that vary in level dynamically.

VII. AN EXAMPLE IMPLEMENTATION: OZGF WITH AGC

As an example, we have recently published an analog VLSI implementation of an auditory filter which shares nearly all the properties outlined above [23]. The design was implemented in the commercially available AMS 0.35 μ m CMOS process using cross-coupled, pseudo-differential, log-domain class-AB biquadratic sections, together with a compressive AGC scheme for incorporating peak gain and selectivity regulation with level. The OZGF, apart from being asymmetric about its peak, has a very simple Laplace-domain description (a cascade of three lowpass biquads together with a scaled bandpass biquad) and thus it can be efficiently implemented in either analog or digital domains. Our design, as evident from the figure, can be tuned at any frequency within the audio range, maintains a constant low-frequency tail while the gain is automatically adjusted with level, and can provide gain at the CF of up to 70dB for small-signal inputs. The design, while not quite as computationally efficient as the PZFC, still attains a 9 out of 10 score in the criteria of Table I and it can thus be regarded as a good candidate for modelling a variety of auditory data.

Other approaches to analog modeling of cochlea function have been recently reviewed by Hamilton [25].

VIII. CONCLUSION

The field of auditory modeling and processing has a variety of good modern filter models to draw on, with efficient implementation as quasi-linear circuits, including useful nonlinear features such as automatic gain control. Both cascade and parallel filterbank structures will continue to converge on being more accurate and useful models of auditory filtering. The parallel filterbanks evolved from the gammatone are simpler filter shapes to describe, while the cascade structures provide for more efficient implementations and the possibility of being tied more closely to underlying wave mechanics. Both forms

can accommodate level-dependence via feedback control of pole positions, and can exhibit realistic nonlinear distortion products if mild nonlinearities are interspersed between the filter sections.

Besides the auditory filter approach as described here, models of bidirectional, multi-modal, and multi-dimensional cochlea response will also play continuing roles in developing our understanding and application of how hearing works.

REFERENCES

- [1] G. Zweig, R. Lipes, and J. R. Pierce, "The cochlear compromise," *The Journal of the Acoustical Society of America*, vol. 59, p. 975, 1976.
- [2] R. L. Wegel and C. E. Lane, "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," *Physical Review*, vol. 23, pp. 266–285, 1924.
- [3] L. Peterson and B. Bogert, "A dynamical theory of the cochlea," *The Journal of the Acoustical Society of America*, vol. 22, p. 369, 1950.
- [4] R. F. Lyon, "Filter cascades as analogs of the cochlea," *Neuromorphic Systems Engineering: Neural Networks in Silicon*, pp. 3–18, 1998.
- [5] A. Katsiamis, E. Drakakis, and R. Lyon, "Practical Gammatone-Like Filters for Auditory Processing," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, 2007.
- [6] D. M. Green, "Detection of complex auditory signals in noise, and the critical band concept," Electronic Defense Group, Department of Electrical Engineering, University of Michigan, Tech. Rep. No. 82, 1958.
- [7] S. Rosen and R. J. Baker, "Characterising auditory filter nonlinearity," *Hear. Res.*, vol. 73, pp. 231–243, 1994.
- [8] E. H. Barton, *A Text-Book on Sound*. Macmillan and Co., 1908.
- [9] J. G. Du Verney, *Traité de l'organe de l'ouïe, contenant la structure, les usages & les maladies de toutes les parties de l'oreille*. Paris: chez Estienne Michallet, 1683.
- [10] E. G. Wever, *A Theory of Hearing*. Wiley, 1949.
- [11] F. E. Terman, *Radio Engineering*. McGraw-Hill, 1932.
- [12] W. M. Siebert, *Circuits, Signals, and Systems*. MIT Press, 1986.
- [13] T. Schafer, R. Gales, C. Shewmaker, and P. Thompson, "The frequency selectivity of the ear as determined by masking experiments," *The Journal of the Acoustical Society of America*, vol. 22, p. 490, 1950.
- [14] W. P. Tanner, J. A. Swets, and D. M. Green, "Some general properties of the hearing mechanism," Electronic Defense Group, Department of Electrical Engineering, University of Michigan, Tech. Rep. No. 30, 1956.
- [15] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.*, vol. 59, no. 3, pp. 640–654, Mar 1976.
- [16] —, "Auditory filter shape," *J. Acoust. Soc. Am.*, vol. 55, pp. 802–809, 1974.
- [17] T. Irino and R. D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp," *J. Acoust. Soc. Am.*, vol. 101, no. 1, pp. 412–419, 1997.
- [18] R. F. Lyon, "The all-pole gammatone filter and auditory models," in *Forum Acusticum '96*. European Acoustics Association, 1996.
- [19] —, "All-pole models of auditory filtering," in *Diversity in Auditory Mechanics*, S. C. E. R. Lewis and R. F. Lyon, Eds. World Scientific Publishing, Singapore, 1997, pp. 205–211.
- [20] S. Rosen, R. J. Baker, and A. Darling, "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," *J. Acoust. Soc. Am.*, vol. 103, pp. 2539–2550, 1998.
- [21] D. Van Compernelle, "Development of a computational auditory model," Institute for Perception Research, Eindhoven, Tech. Rep. IPO Report no. 784, 1991.
- [22] M. Slaney and R. Lyon, "On the importance of time—a temporal representation of time," in *Visual Representations of Speech Signals*, M. Cooke, S. Beet, and M. Crawford, Eds. John Wiley and Sons, 1993, pp. 95–116.
- [23] A. G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "A biomimetic, 4.5 μ W, 120+dB, log-domain cochlea channel with AGC," *IEEE Journal of Solid-State Circuits*, vol. 44, pp. 1006–1022, 2009.
- [24] R. F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 36, no. 7, pp. 1119–1134, 1988.
- [25] T. J. Hamilton, "Analogue vlsi implementations of two dimensional, nonlinear, active cochlea models," Ph.D., University of Sydney, 2008.