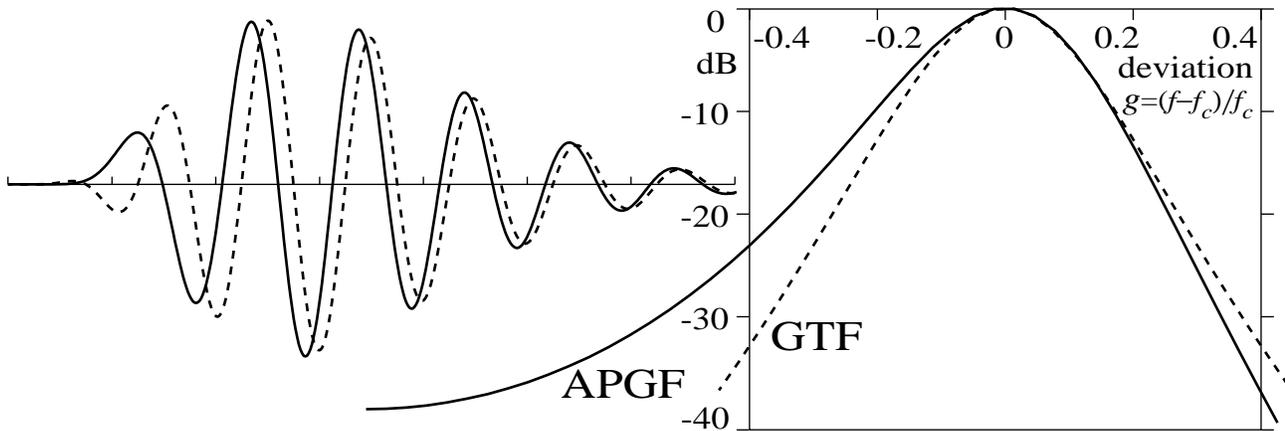


The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
Apple Computer, Inc.
Cupertino, CA 95014 USA

email: lyon@apple.com
fon: (408) 974-4245
fax: (408) 974-8414



ABSTRACT: The All-Pole Gammatone Filter (APGF) is defined by discarding the zeros from a pole-zero decomposition of the Gamma-Tone Filter (GTF) that was popularized in auditory modeling by Johannesma, de Boer, Patterson, and others. Equivalently, the order- N APGF is the N th power of a filter with a complex-conjugate pair of poles; the GTF has this same set of poles, but in addition has “spurious” zeros on the real axis that complicate its description and behavior. The One-Zero Gammatone Filter (OZGF) is also introduced, by differentiating, or adding one zero at DC to, the APGF. Order-3 GTF and OZGF were originally used by Flanagan in 1960 to model Basilar Membrane motion. The APGF does not have the simple time-domain description of the GTF (a gamma distribution times a tone), but is simpler and more well-behaved in other ways, and provides a more robust foundation for modeling and analyzing a variety of auditory data that involve filter asymmetry, frequencies well below CF, level dependence, etc. The GTF magnitude frequency response is approximately symmetric, while the APGF frequency response is inherently asymmetric, being exactly symmetric in frequency-squared space. The low-frequency tail of the APGF is unaffected by the bandwidth parameter, unlike the awkward behavior of the GTF, making the APGF suitable for a range of nonlinear and parametric applications not well addressed by the GTF. Fixing the low-frequency tail gain, rather than the peak gain, as parameters are varied with level, allows better answers to questions about how the auditory system behaves across levels, emphasizing approximate linearity at very low frequencies; varying the damping parameter of the APGF poles can then be interpreted

in terms of automatic gain control (AGC) near CF. The APGF has a simple efficient implementation, a cascade of N identical two-pole filter stages, and is very closely related to the “system of nonlinear differential equations” of Kim et al. and to the even more efficient cascade filterbank auditory models of Lyon and Mead. The APGF provides an improved time-domain match to Basilar Membrane mechanical impulse response measurements, with initial zero-crossing intervals being stretched out. The APGF is offered as a link between physiological, psychological, and mechanical models of auditory filtering.

Introduction

The Gamma-Tone Filter (GTF), introduced by Johannesma to describe cochlear nucleus response in 1972 [J72] and previously used as a basilar membrane model by Flanagan in 1960 [F60], has been adopted as the basis of a number of successful auditory modeling efforts [F62a, S85, P94, MH91, AS89]. Besides its catchy name, due to Aertsen and Johannesma [AJ80], three factors account for the success and popularity of the GTF:

- It provides an appropriately shaped “pseudo-resonant” frequency transfer function with a simple parameterization, making it easy to reasonably well match measured responses
- It has a very simple description in terms of its time-domain impulse response—a gamma-tone: a gamma distribution times a sinusoidal tone
- It provides the possibility of an efficient digital or analog filter implementation.

Two key limitations of the GTF are:

- It is inherently nearly symmetric, while measurements show a significant asymmetry in the auditory filter
- It is not easy to use the parameterization of the GTF to realistically model level-dependent changes in the auditory filter

We introduce a close relative of the GTF, which we call the All-Pole Gammatone Filter (APGF) to highlight its similarity to and distinction from the GTF (and in the choice of acronym, we specifically demote the Tone). The APGF can be defined by discarding the zeros from a pole-zero decomposition of the GTF—all that remains is a complex conjugate pair of N th-order poles. We argue that the zeros of the GTF are “spurious,” and that they limit the usefulness of the GTF in modeling low-frequency tails, impulse response chirping, and level-dependent nonlinearities such as bandwidth, gain, and delay variation.

The APGF was originally introduced by Slaney [S93] as an “All-Pole Gammatone Approximation,” an efficient approximate implementation of the GTF, rather than as an important filter in its own right. In this paper we also introduce the One-Zero Gammatone Filter (OZGF), a better approximation to the GTF which inherits all the advantages of the APGF. The OZGF is defined by differentiating the APGF to introduce one zero at DC (i.e., at $s = 0$ in the Laplace domain). An order-3 OZGF was first used to model Basilar Membrane motion by Flanagan [F60], as an alternative to the order-3 GTF. See the section “History of Gamma-Tone Development” for more details.

The APGF and OZGF have several properties that make them particularly attractive for applications in auditory modeling:

- They exhibit a realistic asymmetry in the frequency domain, providing a potentially better match to psychoacoustic data than the GTF
- They have a simpler parameterization than the GTF and its asymmetric variants
- With a single level-dependent parameter, they exhibit reasonable bandwidth and center frequency variation, while maintaining a linear low-frequency tail
- They are even more efficiently implemented than the GTF
- They provide a logical link to filter cascade models of the underlying cochlear mechanics

These points are supported in the paper by way of examples, illustrations, and mathematical derivations. The OZGF is proposed as the better filter for most auditory filter applications, due to its sloping tail, but the closely related APGF is explored in more depth due to its analytic simplicity.

In the short main body of the paper we define the APGF and OZGF, and detail the history of the GTF, especially in relationship to the properties of auditory filters that have not been well modeled to date. Supporting analysis is all deferred to the appendices, as we did not want to discourage the general reader.

Defining the All-Pole Gammatone Filter

We define the APGF in terms of its Laplace transform $H(s)$. For conciseness, let the s -plane pole positions be given by the complex number p and its conjugate p^* . Then the APGF is given by:

$$H(s) = \frac{K}{[(s-p)(s-p^*)]^N}$$

where K is a constant that will normally be adjusted to give unity gain at DC: $H(0) = 1$. In terms of the Cartesian parameterization of complex pole position, $p = -b + i\omega_r$, and using the denominator form used by Flanagan [F60], the APGF becomes

$$H(s) = \frac{K}{[(s+b)^2 + \omega_r^2]^N} \quad \text{with } K = b^2 + \omega_r^2$$

The OZGF adds “one zero” at DC, a differentiator, or a factor of s (some other rationale for choosing K is needed, since the DC gain is 0):

$$H^{(OZ)}(s) = \frac{sK}{[(s+b)^2 + \omega_r^2]^N}$$

At very low frequencies, the APGF gain is nearly constant (near unity), so we say it has a “flat” low-frequency tail. The gain of the high-frequency tail, by contrast, falls very rapidly with frequency. The OZGF, due to the differentiator, has zero gain at D.C. and a slope of 6 dB per octave at very low frequencies, so is said to have a “sloping” low-frequency tail. Its passband and high-frequency tail are very much like the APGF’s.

The filter shape, gain, bandwidth, delay, dispersion, impulse response, etc. of the APGF and OZGF are analyzed, illustrated, and compared with the GTF in the appendices, starting with and building on the complete analysis of the $N = 1$ degenerate case.

In addition to the Cartesian pole position parameterization mentioned here, we introduce several other parameterizations that are about equally simple; some parameterizations are easier than others for certain analyses. There are several reasonable options for how to couple a single parameter to be level dependent to extend the filter models to include nonlinear behaviors. In the Cartesian parameterization, for example, the b parameter, or real part of the pole position, related to bandwidth, could depend on level. Figure 1 shows a range of shapes of the GTF, APGF, and OZGF in the frequency domain as the b parameter is varied, while Figure 2 shows corresponding impulse responses.

In the appendices we discuss a “theta parameterization” that has several advantages over the Cartesian parameterization, including relating the value of θ conceptually to the degree

of active amplification due to outer hair cells. Dropping θ to zero changes the APGF from a band-pass active filter to a low-pass passive filter.

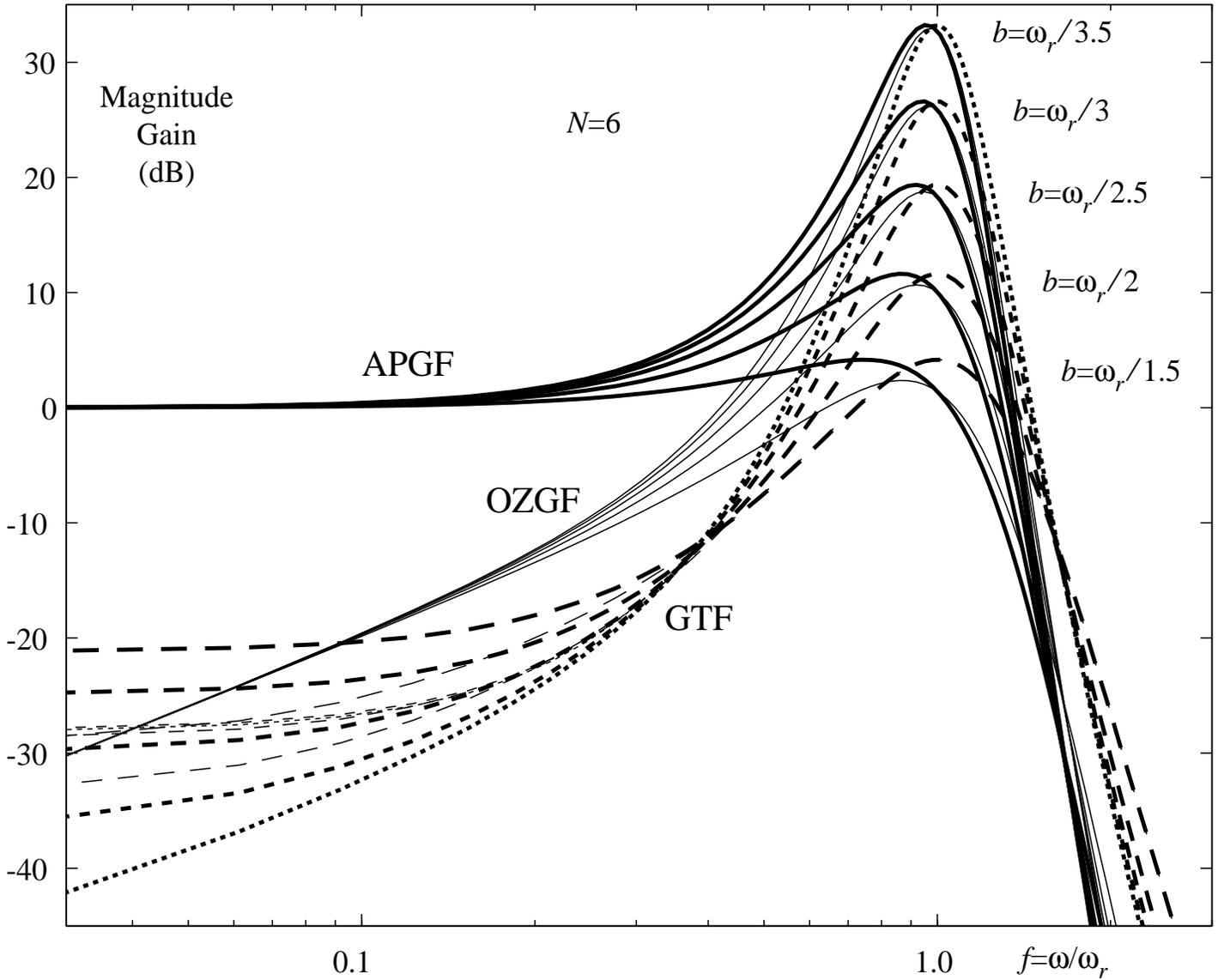


Figure 1. Several APGF, OZGF, and GTF transfer functions. Each set of APGF (solid, heavy), OZGF (solid, light), cosine-phase GTF (dashed, heavy), and sine-phase GTF (dashed, light) share a set of pole locations. The imaginary part of the pole location, the b or bandwidth parameter, is varied from $1/3.5$ to $1/1.5$ (pole Q from 1.68 down to 0.90). The order is $N = 6$ and the OZGF's differentiator gain is taken to be unity at ω_r .

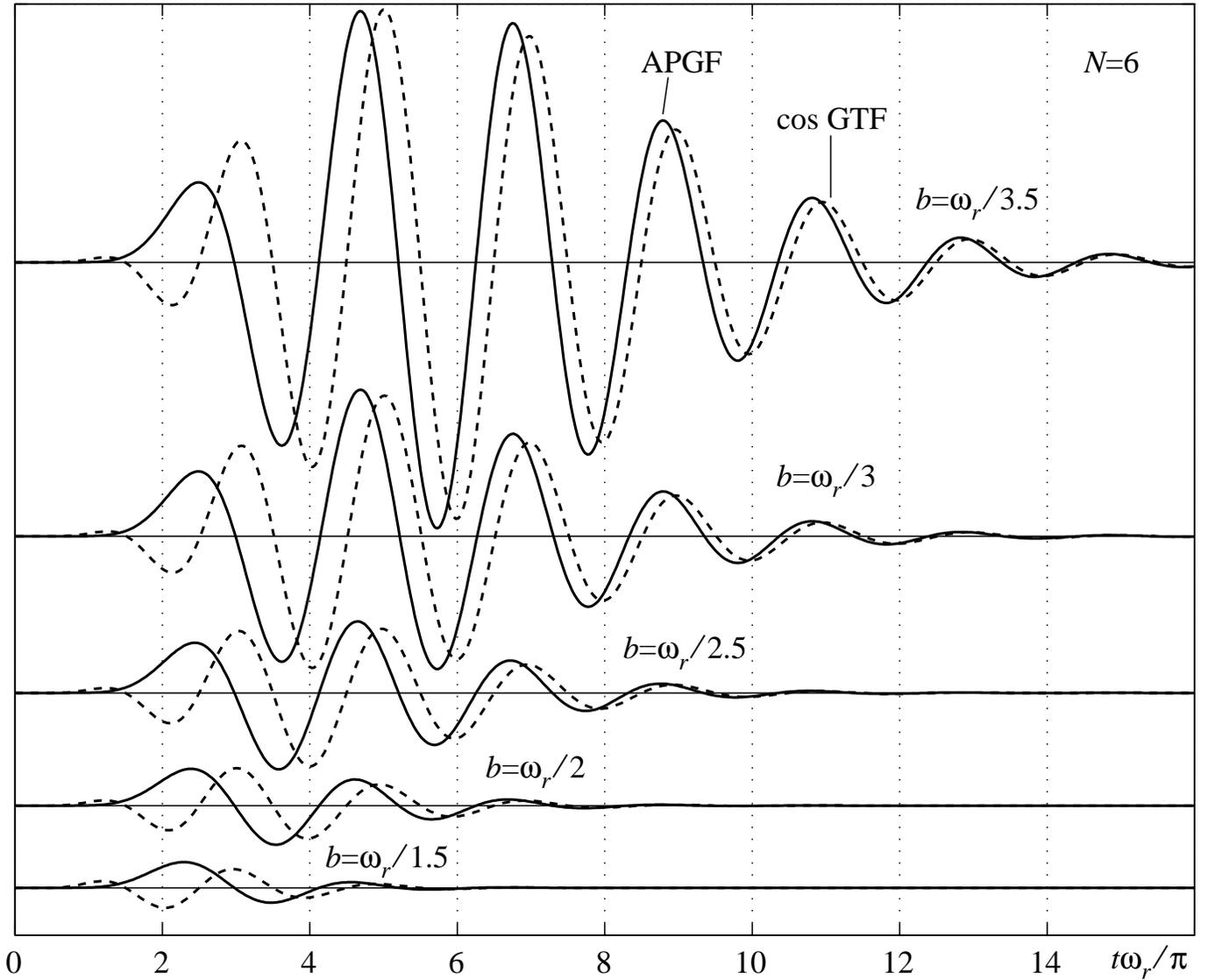
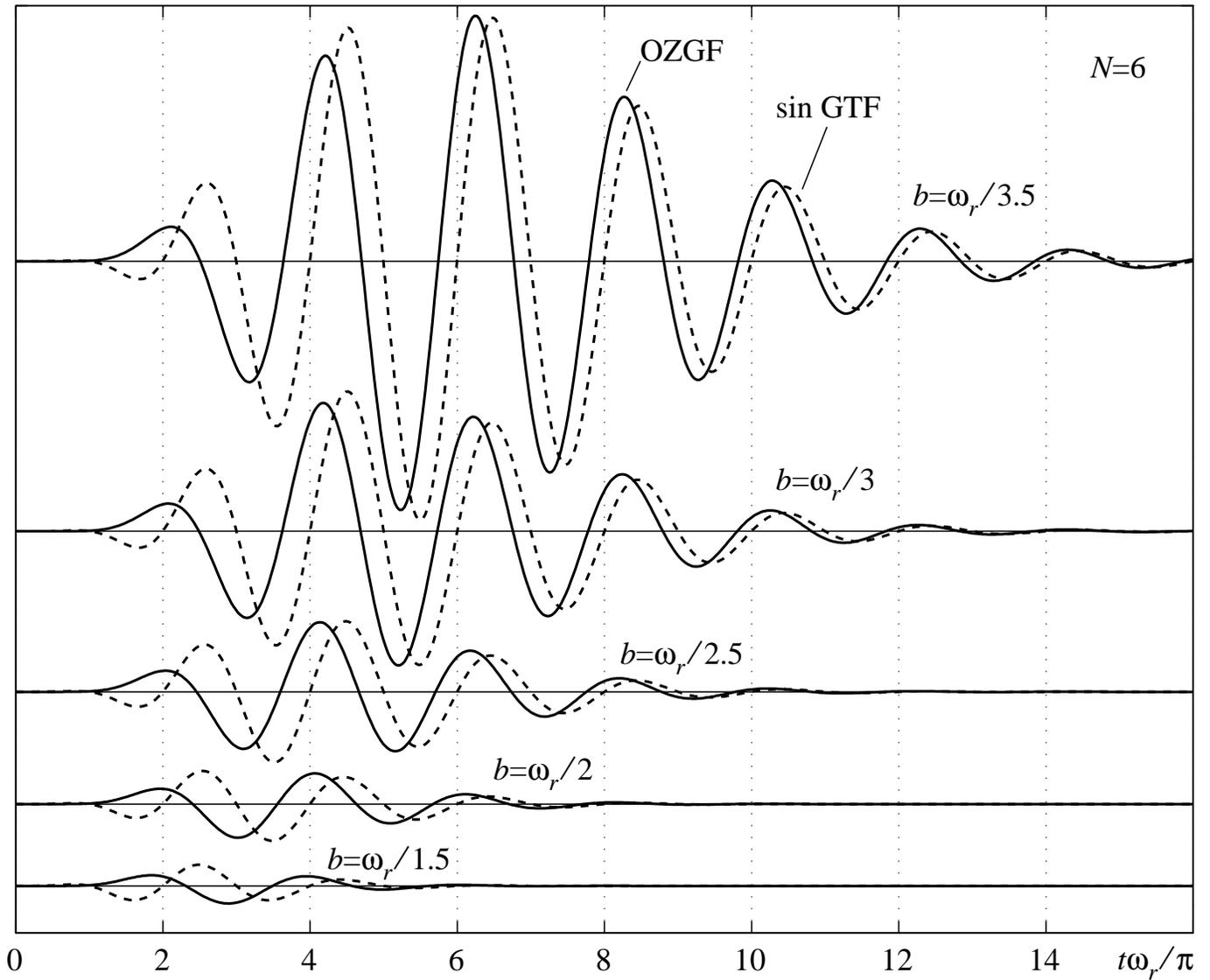


Figure 2a

Figure 2. Impulse responses corresponding to the transfer functions of Figure 1; (2a) APGF (solid) and cosine-phase GTF (dashed), which are asymptotically equal for large t ; (2b) OZGF (solid) and sine-phase GTF (dashed). Grid lines reveal the equally-spaced zero-crossings of the GTF and the relative lead at small t of the APGF and OZGF. Time is labeled in units of π radians of the ringing frequency (half-cycles of the GTF response).

Figure 2b (next page)



A Brief Tour of the Appendices

The detailed development of the mathematical properties of the GTF, APGF, OZGF, and APFC is left for the appendices, which are intended to be interesting reading but may be a bit too dense to include in a first-pass. We describe here what they're about. The body of the paper consists mainly of an extensive historical discussion of the GTF and its relations to other auditory filter model efforts, and should be read first to get an understanding of the context within which the work is evolving.

Notation (Appendix 1)

We explain why we use the acronym CF for center frequency, rather than reserving it for characteristic frequency as some would prefer. We also provide a table of symbols for reference.

Where the Gammatone Filter Gets Its Zeros (Appendix 2)

We analyze the GTF in terms of its Laplace transform, a rational function with both poles and zeros, and give a graphical method for easily finding the locations of the zeros. Since discarding these zeros is the basis of the APGF, understanding where they come from is useful.

Review of Two-Pole Systems (Appendix 3)

The simplest possible GTF or APGF is a filter with just a complex conjugate pair of poles, known as a second-order section (SOS). Although it covers familiar material for many, this section is important to read, as it introduces the parameterizations and many of the formulas used in the coming sections.

Observations on the APGF Response (Appendix 4)

We examine the APGF transfer function, its magnitude, asymmetry, peak gain, tip-to-tail ratio, and shape and tail slope issues.

Gain/Bandwidth/Delay/Order Relations (Appendix 5)

We continue with the analyses of the previous section, examining particularly how various parameters trade off. Delay dispersion is also considered from the point of view of its impact on frequency chirping in the impulse response.

Impulse Response of the APGF (Appendix 6)

The impulse response of the APGF is derived and expressed both in terms of a gamma envelope times a Bessel function and as a sum of simple gamma-tones.

All-Pole Filter Cascade Models (Appendix 7)

In the section we make the link to filter cascade models of cochlear wave propagation. The filter cascade is a unidirectional abstraction of the bi-directional wave propagation

capabilities of the underlying hydrodynamic system, analogous to the WKB approximation. An all-pole filter cascade (APFC) version is the basis of our analog VLSI cochlea chips, and is an efficient and flexible alternative to the APGF.

Parametric Nonlinearity and Gain Control (Appendix 8)

We discuss how to tie a parameter of the APGF or APFC to the detected power level, to model certain kinds of cochlear nonlinearity.

Comparing GTF and RoEx Filter Shapes (Appendix 9)

Finally, we provide more detail on the relationship between GTF and RoEx filter shapes, supporting the discussion in the history section below.

History of Gamma-Tone Development

Cooke [C91] and Patterson [P94]) have recounted the history of the use of the gamma-tone, but some interesting early relationships have been missed. A common reference for the gamma-tone is a paper on revcor functions by de Boer and de Jongh [BJ78], which actually doesn't use or refer to the gamma-tone at all. It seems that everyone knew that the gamma-tone came from de Boer's work on revcor functions, but they had trouble finding the right reference. Carney and Yin [CY88] and Carney [C93] got it right, referencing a paper on synthetic whole-nerve action potentials by de Boer [B75], in which the gamma-tone function was indeed used as a model; they also reference Grashuis [G74], where another gammatone-like formula is used. Patterson [P94] referenced Johannesma's [J72] work on revcor functions in cochlear nucleus as the earliest use of the gamma-tone. de Boer [B73] also credits Johannesma [J72], though it is an unofficial or internal publication. But later de Boer [B75] did not reference Johannesma, having recently learned of an earlier precedent; instead he said that his filter "resembles one introduced by Flanagan [F65], but the parameters are completely different." Aertsen and Johannesma [AJ80] subsequently also reference Flanagan [F65]. These references suggest that the gamma-tone idea was original with Johannesma, but in partial duplication of the earlier work by Flanagan.

Flanagan [F65] shows, on the page referenced by de Boer, a basilar membrane model filter with order-2 pole pairs, an additional real pole, a zero near DC, no parameterization of order or bandwidth, and no simple time-domain description—not really very gammatone-like. Further on, he did mention the actual order-3 gamma-tone, and plotted its "surge"

function or envelope, a gamma distribution, to motivate the choice of a speech analysis window length. At that time, data on cochlear mechanical tuning were quite broad, so a broad low-order filter was a good fit. Flanagan also referenced part II of his earlier work on Basilar Membrane modeling [F62a], which had the order-3 GTF, and which referenced the earlier part [F60], which had the order-3 GTF and also a filter with order-3 pole pairs and a zero at DC—an OZGF. All of Flanagan’s filters were specified in terms of their Laplace transforms as rational functions, but the GTF one was constructed “with a thought toward inverse transformations”—he explicitly worked out the gamma-envelope times tone formulation, as shown in his appendix [F60].

Flanagan understood that the gamma-tone had zeros: he showed both the quadratic numerator in the Laplace transform [F60, F62a], as well as an s -plane pole-zero plot [F62a]. It is unclear however, whether he ever considered generalizing to orders higher than 3; nor did he ever use anything like a ϕ parameter. Flanagan [F60] also understood that “The function obviously becomes non-minimum phase...” Yet Patterson et al. [P88] “... thank Egbert de Boer for ... pointing out that it is a minimum-phase filter” (they probably intended to refer to the fact that the lowpass filter in their frequency-shifting implementation was minimum-phase, or that the one-sided approximation was minimum-phase, rather than the final gammatone filter).

Aertsen and Johannesma [AJ80] appear to have coined the catchy name; referring to the envelope, they said:

The form $m(t)$ appears both as the integrand in the definition of the Gamma function $\Gamma(\gamma)$ and as the density function of the Gamma distribution, therefore we propose to use ... the term “Gamma-tone” or “ γ -tone.”

In that paper, they use gamma-tones as stimuli, not as auditory models, so they do not have a “filter.” The non-hyphenated “gammatone,” as an adjective modifying “filter,” appears to be due to Patterson et al. [P88]. This one-word form was also used by van Stokkum and Gielen [SG89] for the impulse waveform, crediting Aertsen and Johannesma [AJ80]. De Boer and Kruidenier [BK90] use the two-word “gamma tone” for the shape of the filter impulse response waveform, again crediting Aertsen and Johannesma [AJ80]. In this paper, we try to be consistent in using the hyphenated compound as the noun for the waveform, and the one-word compound for the adjective (with the exception of the basis of the GTF acronym). Capitalization is usually not preferred.

The gamma distribution with integer order is also known as the Erlang distribution [P91], but “Erlang-tone” is somehow less exciting as a name.

Key recent sources for popularizing the GTF within the auditory community include: the work on speech visualization by Schofield [S85]; the in-depth “gammatone filterbank” work of Patterson et al. [P88] in collaboration with a number of other researchers seeking a standard auditory filterbank; the tutorial and implementation follow-up of Darling [D91]; the work of Carney and Yin [CY88] and Carney [C93] on level-dependent filterbank models of revcor functions; Cooke’s and Brown’s [C91, BC95] work on auditory processing models and scene analysis; and Assman and Summerfield’s [AS89] and Meddis and Hewitt’s [MH91, MH92, SMH92] applications to perceptual modeling.

A number of authors, in particular Patterson and his colleagues using the “Auditory Image Model,” have applied gammatone filters to speech signal representation and recognition [PH89, R90, PAH92, PHA92, PAA92, P92]. Patterson credits Cooke and Schofield [S85] for originally bringing gamma-tones to his attention.

A recent novel application to sound localization has been presented by Chau and Duda [CD96].

As a further important step in the popularization of the gamma-tone, a number of workers in the field have made available efficient implementations of gammatone filters in their distributions of computer code for auditory modeling and speech processing: O’Mard’s LUTEar [O94], Slaney’s Auditory Toolbox [S93, S94], Patterson, Allerhand, and Giguere’s AIM [PAG95], Culling’s |WAVE [C96], and Wöhrmann and Solbach’s AnnaLisa [WS96].

Hence we see gamma-tones developing essentially concurrently as stimuli, as speech filterbanks, as neurophysiological models, as psychoacoustic models, as cochlear mechanics models, and as efficient widely available code. We hope that the variations presented in this paper will address these needs even better, as a more solid yet flexible foundation for further work in auditory modeling and speech processing.

Relation to Rounded Exponential Filters

In this section we comment on some relationships between GTF and APGF models and Patterson’s rounded exponential or “roex” family of filter power transfer functions.

Patterson [P70, P74] found that experiments on tone detection threshold near a band of noise implied a nearly symmetric auditory filter, which was unexpected; this observation led to a number of symmetric filter parameterizations before later more precise and discriminative experiments revealed a significant asymmetry. Patterson used what he called the “symmetric filter” described in the frequency domain, which corresponds to the usual approximation for a GTF of order 2 (in the 1970 ASA meeting abstract the parentheses are missing in the denominator, making it a little difficult to interpret, but we believe this is the earliest gammatone-like filter besides Flanagan’s):

$$|G(f)|^2 = \frac{1}{[(\Delta f/\alpha)^2 + 1]^2}$$

Patterson also considered a “universal resonance curve” or “single-tuned filter” (equivalent to an order-1 OZGF, not quite symmetric) but found it to be too sharp in the passband or too shallow in the skirts and tails; the simple resonance had previously been used as an auditory filter by Schafer, Gales, Shewmaker, and Thompson [SGST50]. Patterson remarked [P76], “The skirts of the universal resonance curve do not fall fast enough to provide a satisfactory approximation, and the skirts of the filter suggested in Patterson (1974) [the order-2 GTF-like symmetric filter] fall too rapidly.” He then went on to fit the Gaussian filter as an alternative shape for an auditory filter; the Gaussian had been previously introduced as an auditory filter by Swets, Green, and Tanner [SGT62]. In fact, the Gaussian filter falls much too rapidly, and the symmetric filter, like the universal resonance, falls too slowly (but can fall too rapidly in the “skirts” if the peak is narrowed to force a fit to the “tails” over a dynamic range greater than 20 dB).

Since the Gaussian is the limit of the GTF for high order, Patterson was effectively exploring the ends of the GTF order tradeoff; but since the parameterization had not been recognized explicitly, orders near 4 were not tried. Patterson could not know that his symmetric filter was related to the impulse that would later be come to be known as the gamma-tone. Instead, as a compromise on skirt steepness for a given passband width, Patterson and Nimmo-Smith [PN80] introduced the rounded exponential (“roex”) filter.

The rounded exponential filter is a rounding polynomial times an exponential function of frequency deviation. It has gone through several versions with different parameterizations and different tail behaviors. The original version [PN80] used a cubic polynomial rounding factor to control the shape, including the transition from the passband to the tails, but it gave an awkward nonmonotonic shape. The later versions [PNWM82] provided several

alternative treatments of the tail, and accepted the simple passband shape provided by the linear rounding factor:

$$\begin{aligned} \text{roex} &= P(g)\exp(-pg) \quad \text{with } P(g) = 1 + pg + a_2(pg)^2 + a_3(pg)^3 \\ \text{roex}(p) &= (1 + pg)\exp(-pg) \\ \text{roex}(p, r) &= (1 - r)(1 + pg)\exp(-pg) + r \\ \text{roex}(p, w, t) &= (1 - w)(1 + pg)\exp(-pg) + w(1 + tg)\exp(-tg) \end{aligned}$$

The $\text{roex}(p, r)$ introduces a dynamic range limit or flat tail, and the $\text{roex}(p, w, t)$ introduces a tail with less slope than the skirts. Both of these changes can be thought of as warping the roex shape more into agreement with the GTF shape, though the GTF was not being considered at the time.

The roex filters provided a basis for lots of good auditory filter work. For example, Glasberg et al. [GMPN84] studied the dynamic range and asymmetry of the auditory filter, and found a good fit to the $\text{roex}(p)$ form even at 60 dB below the peak for some subjects. They also discussed the fact that the roex fits are in general too peaky very near the center, and needed different parameters above and below center, according to the experimental data.

Completing the cycle, Patterson [P88, P94] compared the $\text{roex}(p)$ shape with the GTF of orders 3 to 5, and found excellent agreement, justifying the use of the GTF as a model for the psychoacoustically determined auditory filter, instead of the magnitude-only characterization provided by the $\text{roex}(p)$ formula.

Moore and Glasberg [MG87] and Rosen and Baker [RB94], using the roex models, addressed the question of whether the level dependent parameters were affected by the level at the input or at the output of the auditory filter; Rosen and Baker extended their analysis to a filter using the GTF shape separately parameterized above and below the peak. They came to conflicting (and in the case of Moore and Glasberg admittedly self-contradictory) results, probably because the roex and GTF models only exhibit a bandwidth and asymmetry increase, not a peak gain reduction, at high levels. I.e., the parameterization of the roex and GTF families of functions does not allow fixing the gain of the low frequency tail, so they fix the gain at the peak instead; as a result, the largest single nonlinear effect in cochlear filtering, input/output amplitude compression, is missing, and parameterizing by output level has the wrong effect. We expect that using a filter parameterization, like the APGF or OZGF or a modification of the $\text{roex}(p, w, t)$, that keeps the low-frequency tail behaving linearly, will resolve their conflict. Rosen and Baker [RB94] have commented on this suggestion but have not explored it.

Relations to Other Work

In this section we comment on some relationships between GTF and APGF models and parameter fitting and the work of others.

Kim, Molnar, and Pfeiffer [KMP73] describe “a system of nonlinear differential equations” which in the low-level linear limit is an all-pole filter with 10 pole pairs. They used natural frequencies staggered by a few percent per stage, so it is not exactly an APGF, but it comes very close—it is also very close to the Lyon and Mead [LM88, LM89a, LM89b] all-pole cascade filterbank model. Kim et al. use a nonlinear damping term as a way to instantaneously adjust damping as a function of level, in very much the same way as the nonlinear circuits of Watts et al. [W92]. As a result of the all-pole structure with unity gain at DC, Kim et al. managed to reproduce a wide range of nonlinear effects, while maintaining linearity for low frequencies.

Schofield [S85] used a GTF bank as the basis of an auditory model for speech visualization, and commented on de Boer’s order fits of 4 to 10 that “it is not easy to establish the order from experimental results...”

Van Stokkum and Gielen [SG89] use a sine-phase GTF with an order of only 2 (4 poles, 1 zero) in their models of grassfrog hearing.

Carney [C93] used an order-4 GTF with variable bandwidth parameter as part of a model of auditory nerve response. The model is elegant and comprehensive, and provides excellent fits to data over the range of levels for which it was designed. Although it uses a GTF that does not have an inherent peak gain vs. bandwidth relation, its implementation uses $\tau = 1/b$ in a sample-rate-dependent way that varies the gain in the right direction, effecting an AGC-like compression with nearly 30-dB of potential gain change (this interpretation of the numbers is based on the implementation description, and seems somewhat at odds with her formulae in which it appears that τ merely dilates the envelope in time, which would yield only a factor of 1.5, or only 3.5 dB, gain range).

Carney’s model is the premier representative of the class of models for which we hope a high-order APGF will provide a superior base. To extend the model to a wider range of levels, with more overall gain variation, would require a wider variation of b and/or a higher order N ; for operation at very high levels, the spurious zeros of the GTF would almost certainly start to be a problem. If the order were kept at 4, there would probably be

too much bandwidth variation as the gain changed over about a 50 dB range, and if the parameterization continued to fix the ringing frequency instead of the natural frequency, there would probably be too little shift in the center frequency at high levels. Hence we suggest an order-8 APGF as the next likely candidate for this class of models.

Carney [C93] uses the symbol τ as the parameter to describe the decay time constant $1/b$; it is the time it takes the exponential to decay to $1/e$ of its value, which corresponds to the envelope decay time constant of the GTF in the extreme late tail where the exponential decay dominates the polynomial growth with time. Johannesma [J72] and de Boer [B75] use β with the same meaning as Carney's τ , including using it within both the exponential and polynomial factors to effect a time-dilation of the gamma-envelope without changing its amplitude. Lyon and Mead [LM88] use the symbol τ as the characteristic time of the oscillation, or $1/\omega_n$, derived from the decay time constants of the follower-integrator circuits that were combined with positive feedback in their filter stage. Because of these conflicting uses of the variable τ , we avoid using it in this paper. Flanagan uses β as the ringing frequency [F60, F62a, F62b, F65], so we avoid that ambiguous symbol, too.

Irino's gammachirp filter (GCF) [I95a, I95b] is a recently developed modification of the GTF based on optimization of time-scale uncertainty using operator methods. The gamma-chirp impulse response consists of a gamma-envelope times a frequency-modulated sine wave, a chirp. The chirp is defined using a log-time phase term, corresponding to a $1/t$ convergence to a final ringing frequency. Like the GTF, the GCF has frequency, phase and bandwidth parameters, plus an additional parameter c to scale the added log-time phase term:

$$g_c(t) = At^{N-1} \exp(-bt) \cos(\omega_r t + c \log t + \phi)$$

This added parameter can be regarded as an asymmetry term; it should be negative to give the usual direction of asymmetry, which corresponds to an upward glide in instantaneous frequency. See Appendix 6 for the corresponding effects in the APGF impulse response.

de Vries and Principe [DP91, DP92, PDO93] have recently independently introduced a "gamma filter" (also gamma delay, gamma memory, gamma model, gamma neural model, gamma kernel) corresponding to the gamma envelope without a tone—i.e. a cascade of identical single-real-pole filters. They discuss a discrete-time version corresponding to Cooke's [C91] pole-mapping version (like the impulse-invariance version but with an extra sample delay per stage). A variant on the gamma filter with complex pole pairs has been introduced by Oliveira e Silva et al. [OS92]; its relation to the APGF and OZGF is made more distant by the inclusion of a zero per pole pair. Cascade forms of the gamma filter and its

complex-poles variant are used as flexible delay lines in adaptive filters and temporal neural networks.

The “complex” (or “analytic”) gammatone filter (CGTF) has been used interchangeably with the GTF by some authors [SWK96]. The output of this filter is complex-valued, and its impulse response is a gamma distribution times a rotating complex phasor. The CGTF is an all-pole filter, and its properties are the basis of the usual excellent approximations to the GTF. Taking the real part of the output introduces spurious zeros on the real axis in the Laplace domain, and converts it back exactly to the GTF.

Møller and Nilsson [MN79] observed chirping of instantaneous frequency in revcor-derived impulse responses, and commented that “this modulation results from the fact that the impulse response of a transmission line contains Bessel functions, whereas the impulse response of ordinary band-pass filters are made up of damped exponentials.” In that sense the APGF is not “ordinary,” while the GTF is; in Appendix 6 we show how the APGF impulse response contains Bessel functions and exhibits chirping.

Ruggero’s [R92, RRR92] laser Doppler velocimeter measurements of Basilar Membrane motion in response to a range of levels of clicks and tones have shown more clearly than most previous data how cochlear mechanical gain, tuning, and delay dispersion change with level. While we have not yet made an effort to quantitatively fit the data, the APGF with constant natural frequency provides a reasonable qualitative fit.

Implementation Considerations

Even before the birth of the digital signal processing (DSP) field, Flanagan [F62b] implemented a discrete-time version of his basilar membrane filter models on a digital computer, using a block diagram compiler [K61] with a pole-mapping method. The particular filter he used was his third alternative (with a real pole in addition to order-two pole pairs and one zero at DC), not the GTF nor the OZGF. His blocks were cascaded one-pole and two-pole filter stages, so his method would have been equally applicable and efficient for these alternatives.

In the rebirth of the gamma-tone in the 1970’s, the fact that it had a rational s -plane transfer function, and could therefore be easily approximated as a z -plane (discrete-time) function and corresponding digital implementation, seems to have been lost. The first digital realizations of the GTF treated the gamma-tone as an arbitrary infinite-duration function, truncated it and sampled it, and implemented the filter via direct convolution; i.e., as a

finite-impulse-response (FIR) filter. This approach is very expensive, as many filter taps (samples of the impulse response) are needed to achieve reasonable accuracy.

The breakthrough in making the GTF acceptable for wider use was Holdsworth's [H88] recognition that the gamma-envelope was the impulse response of a filter with N coincident real poles. He applied this insight via a frequency-shifting filter technique: modulate with a complex oscillation at CF, apply an all-pole infinite-impulse-response (IIR) lowpass filter, and frequency shift back to CF. This approach is drastically cheaper than the FIR filter approach, and gives a more exact result.

Darling [D91] found a useful simplification of Holdsworth's method, using a complex-coefficient IIR filter to place N poles directly at the complex location corresponding to CF, avoiding the need to frequency shift down and back up. This all-pole "complex gamma-tone" was then converted back to a gamma-tone by taking the real part.

Both Holdsworth and Darling therefore implemented minimum-phase all-pole filters as part of their GTF filters, and then converted a single order- N pole to a complex-conjugate pair of order- N poles, effectively by addition of frequency-shifted terms (Holdsworth by shifting DC to $+CF$ and $-CF$, Darling by shifting $+CF$ to $-CF$ via the real-part operation).

Cooke [C91] has provided detailed comparisons of methods to approximate the s -plane ideal continuous-time result by discrete-time z -plane versions, as applied to the all-pole lowpass filter in Holdsworth's method.

In all of these efforts, it appears that nobody ever asked whether the GTF had a simple Laplace transform, i. e., the form that Flanagan started with. Fourier transforms were studied extensively, as they completely describe the magnitude and phase vs. frequency. But Fourier transforms are only a special case of Laplace transforms, evaluated on the complex axis only, rather than on the entire s plane, so the existence of zeros was never noticed.

Finally Slaney [S93] did the Laplace transform, noticed the zeros, and provided the most direct, simple, and efficient IIR filter implementation with real-valued coefficients. At that point, we still did not realize that Flanagan had already done essentially the same thing three decades earlier. Like Flanagan, Slaney factored the Laplace transform into N stages, each having an identical complex-conjugate pair of poles, and then added the zeros. He further investigated the simplification of omitting the zeros completely, as an "all-pole gammatone approximation," primarily as a computational simplification.

The simplicity of implementation of the APGF as a cascade of N identical two-pole filters is one of the advantages of the APGF, but the other good properties of the APGF provide even more important reasons for considering it.

With one more level of approximation, the APGF can be converted to an all-pole filter whose poles are not coincident. The all-pole filter cascade (APFC) configuration allows a further reduction in computation, as discussed in Appendix 7. The APFC is also closely related to the WKB approximation for the solution of an underlying wave equation model of cochlear mechanics.

Conclusions

The APGF is essentially a “missing link” between the GTF and the filter-cascade cochlear model, which is in turn a step away from hydrodynamic wave models of cochlear function. We therefore hope that by introducing the APGF we catalyze a better understanding among modelers from psychophysical, physiological, and mechanical modeling backgrounds.

Many applications in which the GTF has been successful will be unaffected by changing to the APGF or the OZGF. But the APGF or OZGF will provide a significant benefit in applications that need a better model of level dependence or a better low-frequency tail behavior. The ability of the APGF and OZGF to model filter gain, not just shape, will unify the modeling of compressive gain control and filter shape as a function of signal level.

Acknowledgments

My thanks go out to the many people who helped make this work possible:

- Roy Patterson and Egbert de Boer for their extensive help and encouragement in getting the history right, developing the ideas, and improving the presentation
- Malcolm Slaney for the first round of work in APGF analysis and for his help in numerous other respects
- Rahul Sarpeshkar, Sanjoy Mahajan, and David Feinstein for helping to develop and analyze the theta parameterization
- Jim Flanagan, Ray Meddis, Lowell O’Mard, Alfred Nuttall, Angela Darling, Birger Kollmeier, Bill Woods, Jozef Zwislocki, Toshio Irino, Jose Principe, Laurel Carney, Ludger Solbach, Martin Cooke, Stuart Rosen (and maybe some others I’ve failed to note) for filling me in on the details of their gammatone-related activities

- Carver Mead for convincing me years ago that we didn't need zeros to make a good cochlea chip
- and my boss Don Norman and Apple Computer for supporting this diversion

All this help probably didn't stop me from inserting a few errors of my own, though.

— DRAFT of 29 Jan. 96 — do not distribute

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 1—Notation

The acronym CF is most often used in the hearing field to denote the “characteristic frequency” of a neuron, or the frequency at which a neuron is most sensitive or has the lowest threshold. If auditory filters were linear, CF would then be (at least very nearly) equivalent to the frequency at which the auditory filter has the highest gain, often called the “center frequency” of the filter. But the cochlea is known to be nonlinear, and the frequency of the peak gain of the linearized auditory filter is known to be level dependent.

Some authors extend the CF concept of greatest sensitivity to mechanical filtering as well as to neural data [H80]. Many authors say the CF shifts with level [E89, YCG83, WBW88], which is inconsistent with the use of CF as frequency of greatest sensitivity. Some use BF (best frequency) instead of CF for the frequency of greatest sensitivity of a neuron [RHAB71, DW94, PWJJ94, Pr88], and some use F_c [MN79]. Some use “characteristic” and “best” interchangeably [RS85], and allow level dependence [Pi88]. Some say CF is either the frequency of greatest firing rate (which is level dependent) or of lowest threshold [Y94]. Some use CF as “center” or “tuning” frequency of a mechanical iso-response tuning curve [K84] or as the “center” frequency of a modeled neuron or “channel” [MHS92] or of an auditory filter [MHS90].

The usages of CF and related terms are therefore not in fact standardized. In general, there seems to be confusion about how to separate these measures, one level-dependent and one a low-level limit and about when the neural and filter parameters need to be distinguished.

Since we’re writing about filters and not neurons, and since center frequency is a key measure of a filter’s transfer function, we prefer to use CF for center frequency, the frequency at which a filter’s gain is maximized, and to allow CF to depend on level. Then we use BF for the low-level limit; i.e., the best frequency is the frequency of greatest sensitivity or lowest threshold, from the point of view of the filter. We beg the indulgence of those

hearing scientists who would prefer to reserve the term CF for the characteristic or best frequency of a neuron.

The following symbols are used in this paper, and are collected here for reference:

Transfer functions:

$G(s)$ = gammatone filter

$H, H(s), H(i\omega), H(f)$ = all - pole gammatone filter

$H_1, H_1(s), H_1(i\omega), H_1(f)$ = filter stage or second - order section (SOS)

$H^{(OZ)}(s)$ = one - zero gammatone filter

$H^{(APFC)}(s)$ = all - pole filter cascade filter

Impulse responses:

$g(t)$ = gammatone filter

$g_c(t)$ = gammachirp filter

$h(t)$ = all - pole gammatone filter

$h_1(t)$ = filter stage or second - order section (SOS)

Functions:

$u(t)$ = Heaviside's unit step function

$\Gamma(x)$ = the gamma function

$\exp(x)$ = the exponential function e^x

$\log(x)$ = the natural logarithm, the inverse of exp

$\cos(x)$ = the cosine function $(\exp(ix) + \exp(-ix))/2$

$\sin(x)$ = the sine function $(\exp(ix) - \exp(-ix))/2i$

$\tan(x)$ = the tangent function $\sin(x)/\cos(x)$

$\cot(x)$ = the cotangent function $1/\tan(x)$

$\arccos(x)$ = the inverse function of cos

$\arcsin(x)$ = the inverse function of sin

$\arctan(y/x)$ = the angle between vectors (1,0) and (x,y) for $x > 0$

$\text{arccot}(x/y)$ = the angle between vectors (1,0) and (x,y) for $y > 0$

$J_n(x)$ = the Bessel function of the first kind

$j_n(x)$ = the Spherical Bessel function of the first kind

Variables, Parameters, Constants:

t = time, typically measured from an impulse input

s = complex frequency parameter of Laplace transform

i = the unit imaginary number $\sqrt{-1}$

π = Pi, 3.14159..., the number of radians in a half cycle

e = 2.71828..., the base of natural logarithms

N = order of compound poles in GTF, APGF, or OZGF

b = negative real part of pole location, rad / sec

ϕ = phase of the cosine tone in $g(t)$

A = amplitude factor in $g(t)$

c = phase factor in $g_c(t)$

K = amplitude factor in $H(s)$

ω = frequency, rad / sec, independent parameter of a transfer function

ω_r = ringing frequency, imag. part of pole location, rad / sec

ω_n = natural frequency, distance of pole location from origin, rad / sec

ω_c = center frequency (frequency of peak gain), rad / sec

p, p^* = an arbitrary complex number (pole location) and its conjugate

More Parameters, etc:

Q = quality factor of a resonant system

a = criterion power ratio at band edge (> 1)

f = normalized frequency ω/ω_n

f_c = normalized center frequency ω_c/ω_n

f_h = high band edge, where $H^2(f_h) = H^2(f_c)/a$

f_l = low band edge, where $H^2(f_l) = H^2(f_c)/a$

BW = bandwidth, frequency difference between high and low band edges

BW_{3-dB} = half - power bandwidth, BW for $a = 2$

Q_{3-dB} = CF / BW_{3-dB} for $a = 2$

Q_{10-dB} = CF / BW for $a = 10$

ERB = equivalent rectangular bandwidth

θ = polar pole parameter $\arcsin(1 - 1/(2Q^2))$

H_{\max} = the maximum gain of the APGF, assuming DC gain is unity

x = squared normalized frequency f^2 , extended to complex plane

x_p = complex x - plane pole $\sin \theta \pm i \cos \theta$

d = the distance in the x - plane from f^2 to x_p

p, t, w, r = parameters of the roex family of filters

$\alpha = b/\omega_r$, normalized GTF bandwidth parameter

Impulse response specific:

A_n = amplitude factor on gamma - tone n of N

$C_{n,N}$ = an integer numerator factor in A_n

D_N = an integer denominator factor in A_n

Group Delay, etc.

T_g = group delay of APGF, as a function of normalized frequency

$T_g^{(DC)}$ = group delay of APGF at DC

$T_g^{(\omega_n)}$ = group delay of APGF at natural frequency

$T_g^{(CF)}$ = group delay of APGF at center frequency

$T_g^{(max)}$ = maximum group delay of APGF

$f_{T_{peak}}$ = normalized frequency at which group delay is maximized

Particular to the SOS (1 - stage filter):

ϕ_1 = phase response of the SOS

$H_{1\max}$ = maximum gain of the SOS, assuming DC gain is unity

T_{g1} = group delay of SOS, as a function of normalized frequency

$T_{g1}^{(DC)}$ = group delay of SOS at DC

$T_{g1}^{(\omega_n)}$ = group delay of SOS at natural frequency

$T_{g1}^{(CF)}$ = group delay of SOS at center frequency

$T_{g1}^{(\max)}$ = maximum group delay of SOS

All - pole filter - cascade specific:

N_e = number of stages per e - fold change in natural frequency

H_{\max}^{APFC} = peak gain of the filter cascade

ω_c^{APFC} = center frequency of the filter cascade

T_g^{APFC} = group delay of the filter cascade

$T_{g(DC)}^{APFC}$ = DC group delay of the filter cascade

$T_{g(CF)}^{APFC}$ = group delay of the filter cascade at CF

ω_n^{tap} = local natural frequency associated with the output tap

ω_c^{tap} = local center frequency associated with the output tap

— DRAFT of 29 Jan. 96 — do not distribute

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 2—Where the Gammatone Filter Gets Its Zeros

The GTF is defined in terms of its impulse response $g(t)$, a “gamma-tone” or “ γ -tone” [AJ80], the product of a gamma distribution $At^{N-1}\exp(-bt)$ (or zero for negative time) times a tone $\cos(\omega_r t + \phi)$:

$$g(t) = At^{N-1}\exp(-bt)\cos(\omega_r t + \phi)$$

For the gamma-distribution factor to be an actual probability distribution (i.e., to integrate to unity), the factor A needs to be $b^N/\Gamma(N)$ [HC95]. The gamma-function is defined for integers as the factorial of the next lower integer: $\Gamma(N) = (N-1)!$, or more generally, but tautologically in this context, as $\Gamma(N) = \int_0^\infty t^{N-1}\exp(-t)dt$. In practice, A is used as an arbitrary factor in the filter response, and is typically chosen to make the peak gain equal to unity. Following Patterson et al. [P88], we call the scaled gamma distribution a “gamma envelope.” Figure A2.1 illustrates the gamma envelope, the tone, and the gamma-tone.

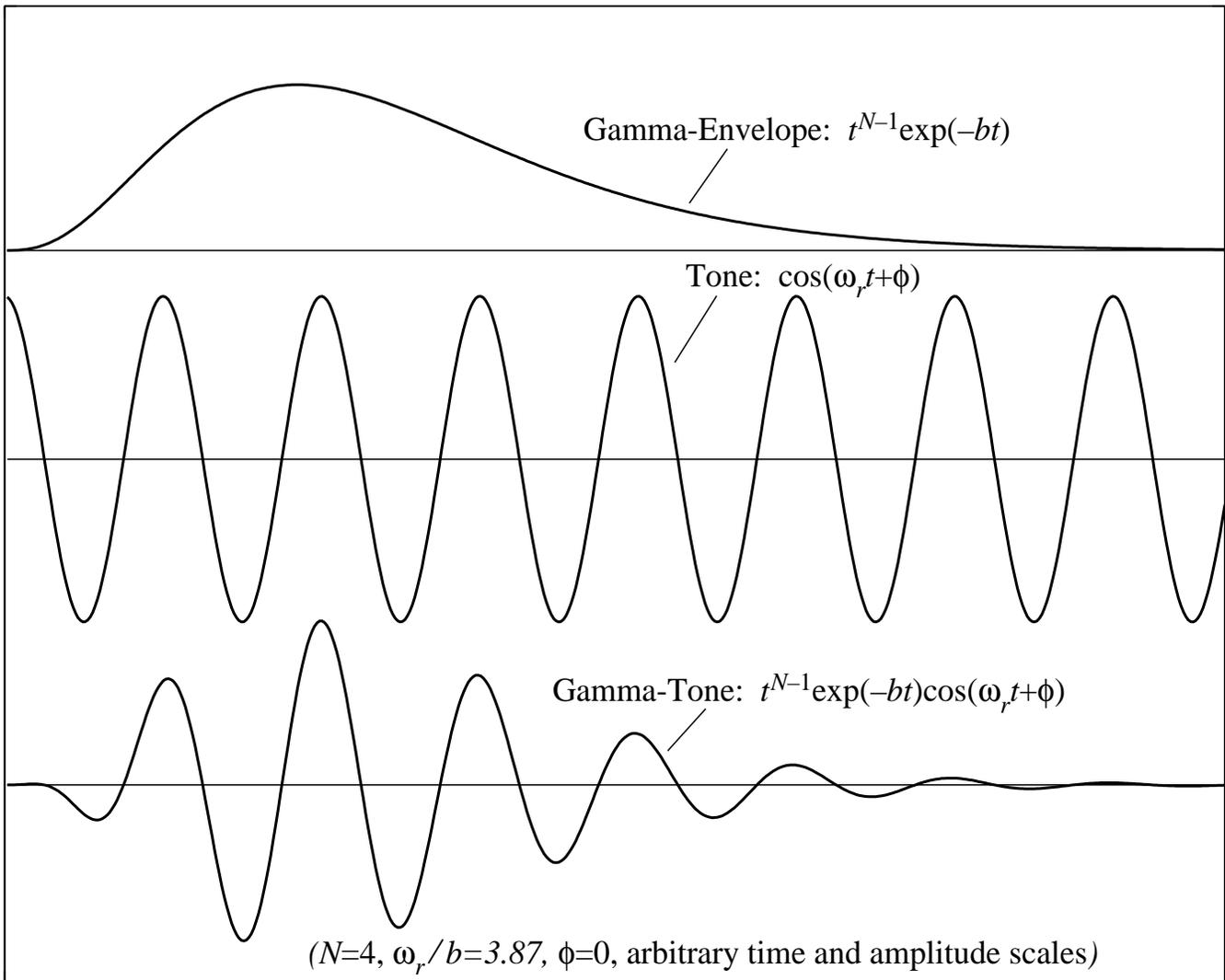


Figure A2.1. The components of a gamma-tone. The gamma-envelope (top), the sinusoidal tone (middle), and their product the gamma-tone (bottom). For all impulses in this paper, multiplication by the Heaviside function (unit step function) $u(t)$ is implied, so that responses are identically zero before the time origin.

The parameters, integer order N , ringing frequency ω_r radians per second, starting phase ϕ radians, and one-sided pole bandwidth b radians per second, complete the description of the particular GTF; more typically it is parameterized with ringing frequency and bandwidth in Hertz, rather than radians per second, but that clutters the formulae with too many 2π 's.

For all impulses in this paper, multiplication by the Heaviside function (unit step function) $u(t)$ is implied; i.e., the filters are causal and Laplace transforms defined only on positive time can be used. By expanding the cosine tone into a sum of complex exponentials, we

arrive at two terms with very easy Laplace transforms, or frequency-domain characterizations:

$$g(t) = At^{N-1} \exp(-bt) \frac{\exp(i\omega_r t + i\phi) + \exp(-i\omega_r t - i\phi)}{2}$$

$$= \frac{A}{2} \left[\exp(i\phi) t^{N-1} \exp((-b + i\omega_r)t) + \exp(-i\phi) t^{N-1} \exp((-b - i\omega_r)t) \right]$$

Using any Laplace transform table with the relation $t^{N-1} \exp(pt) \xrightarrow{L} \Gamma(N)/[s-p]^N$, identifying p with the complex pole locations $-b + i\omega_r$ and its complex conjugate, we arrive at the gamma-tone's Laplace transform, or GTF transfer function:

$$G(s) = \frac{A\Gamma(N)}{2} \left[\frac{\exp(i\phi)}{[s - (-b + i\omega_r)]^N} + \frac{\exp(-i\phi)}{[s - (-b - i\omega_r)]^N} \right]$$

The two terms each have only poles (denominator roots in s); one term has N poles at $s = -b + i\omega_r$ and other has N poles at the complex-conjugate location $s = -b - i\omega_r$. But the two N th-order complex pole terms add with constructive or destructive interference, depending on their relative phases, at various points in the s plane. If their sum is zero at some point, then that point is a zero of the GTF.

It is easy enough to combine the two terms over a common denominator to put the filter into a rational function form, such that the zeros of the GTF are the zeros of the numerator polynomial (but let's drop the constant factor $A\Gamma(N)/2$):

$$G(s) = \frac{\exp(i\phi)[s - (-b - i\omega_r)]^N + \exp(-i\phi)[s - (-b + i\omega_r)]^N}{[(s+b)^2 + \omega_r^2]^N}$$

The common denominator is quite simple, but the resulting numerator polynomial looks a little more formidable. It is not hard, given adequate tools, to solve for the zeros numerically or symbolically for particular values of N , as shown by Slaney [S93].

Alternatively, the zeros can be found graphically by observing where the two all-pole terms are equal in magnitude and opposite in phase, and thereby cancel each other. The magnitudes are equal all along the line that is equidistant from the poles, which is the real axis in the s plane. So the question is simply, where on the real axis are the two terms exactly out of phase? Notice that the two terms are complex conjugates of each other when

evaluated at real s , so they are out of phase when they are pure imaginary—i.e., where one has phase 90° ($\frac{\pi}{2}$ radians) and the other 270° ($\frac{3\pi}{2}$ radians).

Therefore a plot of the phase contours of the N th-order pole terms, which are “spokes” radiating from the pole locations, will immediately show the locations of the zeros where the 90° and 270° spokes meet on the real axis. Figure A2.2 illustrates this construction for a particular combination of N and ϕ . It becomes clear that a zero can be placed anywhere on the real axis by varying ϕ , and that there are N total zeros—except in the “sine phase” condition where ϕ is 90° or 270° , in which case one of the zeros moves out to infinity, leaving $N - 1$ zeros. The pole-zero plots of Flanagan [F62a] (order 3, sine phase) and Slaney [S93] (order 4, cosine phase) are consistent with this graphical analysis, but do not lead to a general view of the zeros for different parameters.

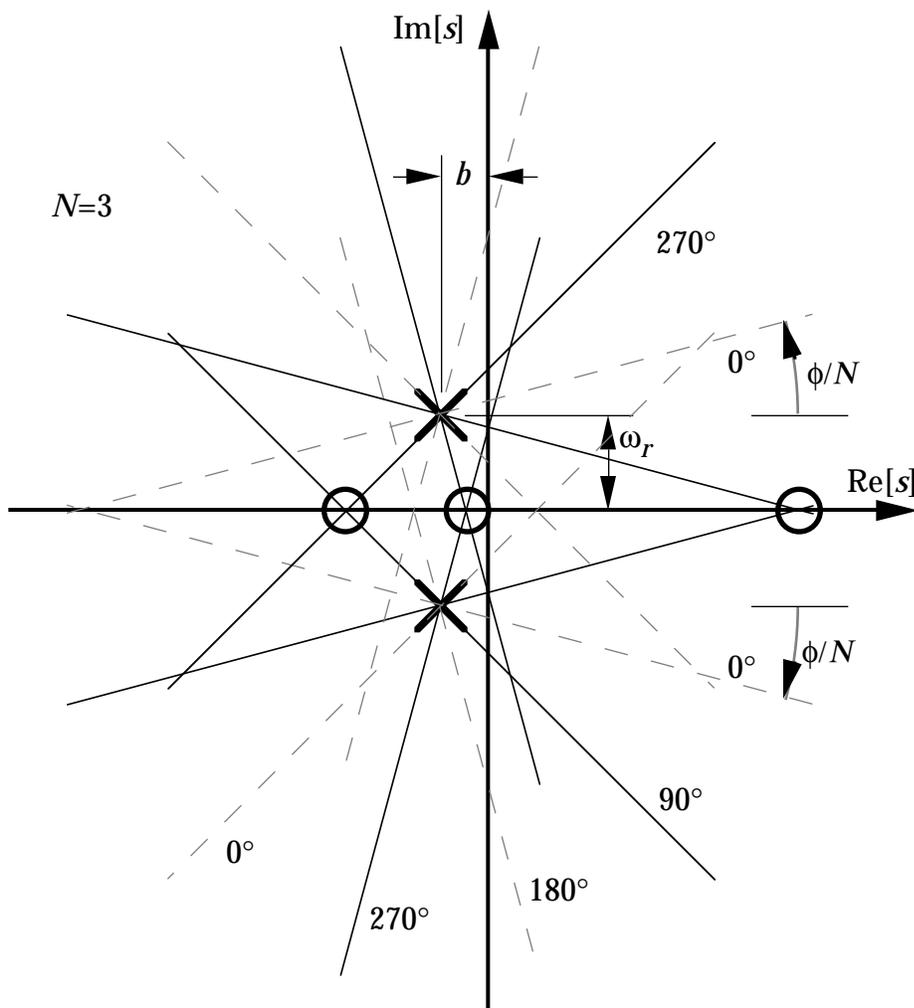


Figure A2.2. The s -plane poles and zeros of the order 3 gammatone filter for $\phi = 45^\circ$ ($\pi/4$). The zeros occur on the real axis where the "pure imaginary" constant-phase contours of the phase-shifted compound pole terms meet. The radial contours of pure-imaginary values (90° and 270°), which lead to the zeros, are shown in solid lines, while the pure-real contours (0° and 180°) are shown as dashed lines. The 0° contours that would extend to the right of the poles for $\phi = 0$ have been rotated in opposite directions by an amount ϕ/N . Increasing ϕ to 90° ($\pi/2$) would shift the rightmost zero out to infinity, leaving two zeros, in agreement with the sine-phase gamma-tone example shown by Flanagan [1960], with one zero at or near the DC point (the origin $s=0$). The Cartesian parameters b and ω_r of the pole positions are indicated.

What is the effect of these zeros? First, notice that the Fourier frequency response, the evaluation of the Laplace transform along the imaginary axis $s = i\omega$, may have a zero at $\omega = 0$ for particular phases of the gammatone filter. On a dB scale, the tail of the frequency response may therefore head for minus infinity at DC, depending on the particular relation among ϕ , N , b , and ω_r . Second, notice that at very high frequencies, the N zeros "mask" N of the poles—the ultimate falloff rate is only the $6N$ dB/octave of the individual N -pole terms, rather than the $12N$ dB/octave that would be possible with all $2N$ poles without zeros.

Figure 1 in the main paper shows several GTF, APGF, and OZGF transfer functions to illustrate these effects; notice that for the GTF transfer function, the ordering of gains at the peak is unrelated to the order in the tail, and that the orders are quite different for the cosine-phase ($\phi = 0$) and sine-phase ($\phi = \frac{\pi}{2}$) conditions. Figure 2 shows the corresponding impulse responses, where the differences in tail behaviors are not apparent.

Another effect of the zeros is that the GTF is usually not quite minimum phase, since a minimum-phase filter must have all of its zeros in the left half plane. This effect is not highly significant, but is contrary to the expressed auditory filter design goals of Patterson et al. [P88].

The possibility of interference between the two N -pole terms of the GTF in the frequency domain has been noted by several authors [H88, BK90, D93], but has generally been written off as a small and inconsequential effect for parameters in the range of interest. This is true for the range of parameters typically used to characterize revcor functions (neural reverse-correlation-derived impulse responses), since revcor functions provide a reasonable

characterization of auditory filters only when the passband response is large compared to the low-frequency response. This condition is satisfied only to about 40 to 60 dB above the threshold of hearing [BJ78, E89, BK90]; at higher levels, the response as observed on the auditory nerve becomes increasingly lowpass in nature, and the assumptions underlying revcor analysis begin to break down. To break out of this circular consistency constraint and extend the applicability of the GTF to high SPL's, where the filtering is more nearly low-pass, the interference needs to be understood and dealt with—one motivation for the APGF.

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 3—Review of Two-Pole Systems

A system with two poles and no zeros is of special interest, as it is both a GTF ($N = 1$, $\phi = \frac{\pi}{2}$) and an APGF ($N = 1$), as well as the basic building block of higher order filters. We refer to this special case as a “stage,” or as a “second-order section” (SOS), a terminology common in the DSP field (not to be confused with a second-order gamma-tone, which is a fourth-order filter with a pair of second-order poles). We denote this $N = 1$ system using the subscript 1 on the impulse response, transfer function, and related quantities.

SOS Transfer Function

The transfer function of the SOS, with numerator adjusted for unity gain at DC ($s = 0$), in terms of the Cartesian parameterization, is:

$$H_1(s) = \frac{b^2 + \omega_r^2}{(s + b)^2 + \omega_r^2}$$

The equivalent expression parameterized by Q and natural frequency is

$$\begin{aligned} H_1(s) &= \frac{\omega_n^2}{s^2 + s\omega_n/Q + \omega_n^2} \\ &= \frac{1}{(s/\omega_n)^2 + s/(\omega_n Q) + 1} \end{aligned}$$

where the natural frequency $\omega_n = \sqrt{b^2 + \omega_r^2}$ is the distance of the pole from the origin in the s plane, and $Q = \omega_n/(2b)$ is the standard definition of the quality factor of a second-order system. This “polar” parameterization is sometimes preferred because, in physical systems, the natural frequency is a function of physical energy-storage parameters, independent of damping, and Q is independent of frequency and depends only on relative damping

(damping factor $\zeta = 1/(2Q)$ is sometimes used instead of Q in describing such systems). In variable-damping physical auditory models, it may be more appropriate to fix the natural frequency than the ringing frequency.

SOS Impulse Response

The impulse response of this system is a degenerate gamma-tone:

$$h_1(t) = A \exp(-bt) \sin(\omega_r t)$$

where $A = (1 + b^2/\omega_r^2)\omega_r$ for unity gain at DC.

SOS Gain Magnitude

The gain magnitude is given by:

$$\begin{aligned} |H_1(i\omega)| &= \sqrt{H_1(i\omega)H_1^*(i\omega)} = \sqrt{H_1(i\omega)H_1(-i\omega)} \\ &= \frac{1}{\sqrt{1 + (1/Q^2 - 2)(\omega/\omega_n)^2 + (\omega/\omega_n)^4}} \end{aligned}$$

It is easier to understand the magnitude gain in terms of the simple quadratic polynomial function of $(\omega/\omega_n)^2$ in the denominator of $|H_1(i\omega)|^2$:

$$\begin{aligned} |H_1(i\omega)|^2 &= \frac{1}{1 + (1/Q^2 - 2)(\omega/\omega_n)^2 + (\omega/\omega_n)^4} \\ &= \frac{1}{1 + (1/Q^2 - 2)x + x^2} \text{ where } x = (\omega/\omega_n)^2 \end{aligned}$$

This reduction of the quartic denominator to a quadratic introduces the normalized-squared-frequency domain of $x = f^2 = \omega^2/\omega_n^2$ of Sarpeshkar et al. [SM96], based on frequency normalized relative to natural frequency $f = \omega/\omega_n$. We use these normalized variables extensively.

The maximum of the gain magnitude occurs at a frequency we call the center frequency (CF), quantified in radians per second as ω_c .

Asymmetry from Symmetry

An elegant formulation and graphical construction for the SOS transfer function due to Sarpeshkar et al. [SM95] allows simple expressions in the normalized squared-frequency domain of $x = f^2 = \omega^2/\omega_n^2$. They introduce an alternative polar parameterization in terms of θ , an angle characterizing the position of the x -plane roots of the quadratic in the denominator of the magnitude transfer function.

Since the denominator of the squared gain is simply a quadratic function of x , as shown above, it is a symmetric parabola in that squared-frequency domain. A graphical evaluation of the denominator polynomial can be constructed in the complex x plane, just as polynomials in s are evaluated using products of distances to roots in the s plane; see Figure A3.1. The roots (poles) are at $x_p = (1 - 1/(2Q^2)) \pm i\sqrt{1 - 1/(4Q^2)}/Q$. Since the coefficients of the constant and quadratic terms in $1 + (1/Q^2 - 2)x + x^2$ are both 1 and the poles are complex, the poles lie on the unit circle and are much easier to interpret as $x_p = \sin \theta \pm i \cos \theta$, Sarpeshkar's key observation. This formula for the poles gives one definition of the "theta" parameterization:

$$\theta = \arcsin(1 - 1/(2Q^2))$$

The graphical evaluation of the polynomial is just the product of the Euclidean distances from the "frequency point" $(x, 0)$ to the poles $(\sin \theta, \pm \cos \theta)$. Distances from the real axis to two complex-conjugate poles are equal; call them d , which by the Pythagorean theorem are:

$$d^2 = (x - \sin \theta)^2 + \cos^2 \theta$$

Since the distances are equal, the graphical construction can be reduced to use a single pole $(\sin \theta, \cos \theta)$, called the " Q -point" since its location only depends on Q , and a single distance d , yielding $|H_1(i\omega)| = 1/d$, a symmetric function of x .

Figure A3.1 illustrates the theta parameterization and the construction just described, and illustrates the relation between the symmetric response in the x domain and the asymmetric frequency response $|H_1(i\omega)|$. Notice that the θ parameter may be defined directly in the x plane, or equivalently as twice the angle in the s plane between the radial to the pole and the radial at 45 degrees, as the reader may verify.

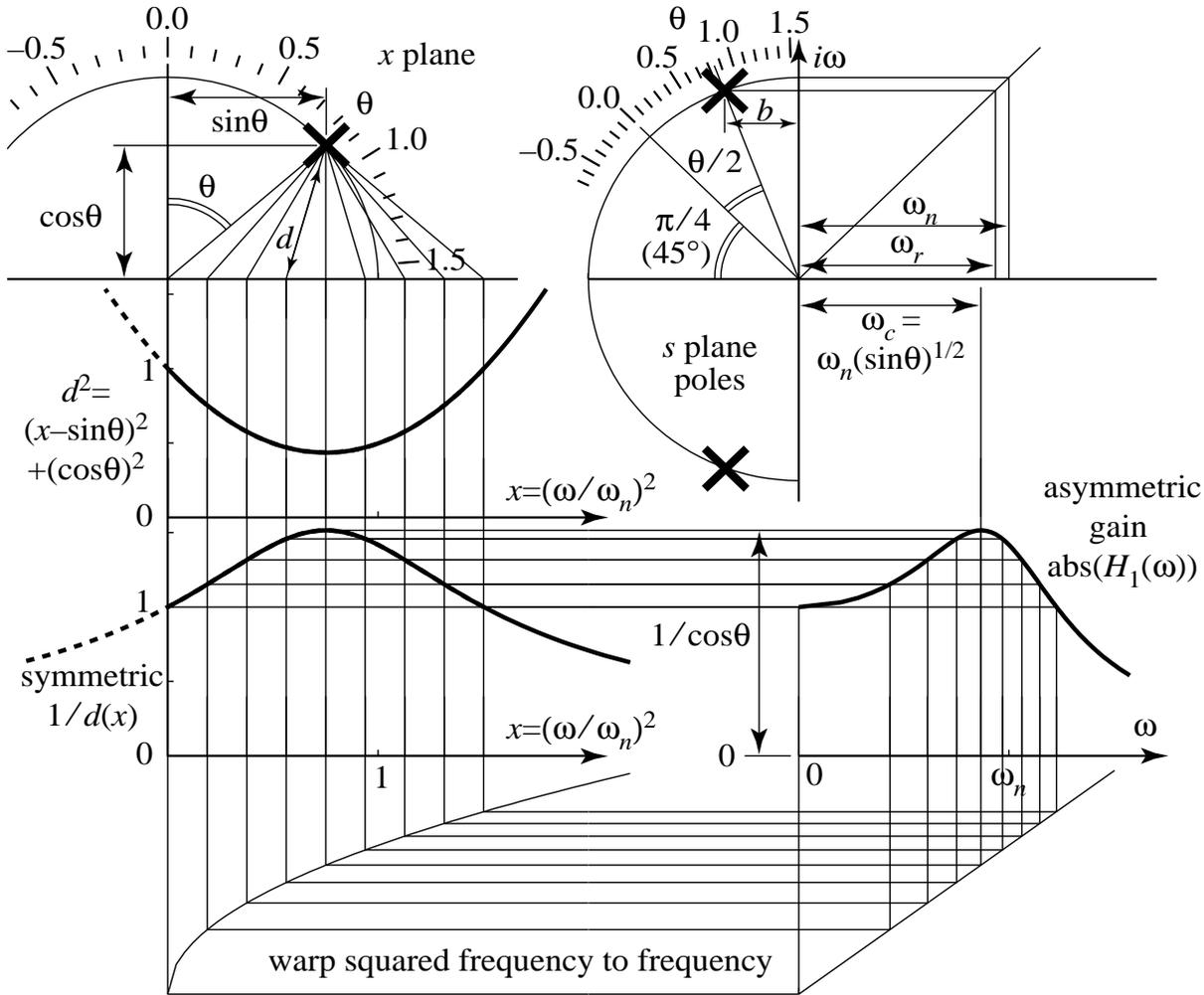


Figure A3.1. The “theta” parameterization of a two-pole system. The left hand side of the figure introduces the “normalized squared frequency” space and the complex x plane, for $x = (\omega/\omega_n)^2$, while the right hand side of the figure shows the more conventional s plane and frequency domain. The x -plane pole shown is one of the complex conjugate pair of roots of the denominator polynomial of the squared magnitude $|H_1(i\omega)|^2 = 1/(1 + (1/Q^2 - 2)x + x^2) = 1/d^2$. The graphical evaluation of the real quadratic polynomial involves the product of two identical distances d from the x axis to a pole, so only one is shown. The connection between the two sides of the figure is via the abscissa warp shown at the bottom, and via the angles θ and $\theta/2$ shown at the top. In the x domain, the magnitude response is exactly symmetric, explaining the asymmetry in the frequency domain. Negative values of x are not particularly meaningful, but curves are shown dashed in that region to emphasize the symmetry in x . The center frequency ω_c does not align with the natural frequency ω_n or the ringing frequency ω_r , but is a simple function of θ .

Peak Gain and CF

As shown in Figure A3.1, for $\theta \geq 0$ the transfer function has a peak at a nonzero frequency. In the theta parameterization, the frequency of the peak corresponds to the minimum of d , where x aligns with the pole location's real part, $x = (\omega/\omega_n)^2 = \sin \theta$, resulting in these solutions for center frequency:

$$\begin{aligned}\omega_c^2 / \omega_n^2 &= \sin \theta = 1 - 1/(2Q^2) \\ \omega_c &= \omega_n \sqrt{\sin \theta} = \omega_n \sqrt{1 - 1/(2Q^2)}\end{aligned}$$

The relations $\sqrt{1 - 1/(4Q^2)} = \omega_r/\omega_n$ and $\sqrt{1 - 1/(2Q^2)} = \omega_c/\omega_n$ should be kept in mind, as they are very handy in constructing and interpreting mixed representations. At CF, where $x = \sin \theta$ and $d = \cos \theta$, the gain is given by the following expressions in terms of the various parameterizations (theta, Q, Cartesian, and mixed):

$$H_{1\max} = \frac{1}{\cos \theta} = \frac{Q}{\sqrt{1 - 1/(4Q^2)}} = \frac{\omega_r}{2b} \left(1 + \frac{b^2}{\omega_r^2} \right) = Q \frac{\omega_n}{\omega_r}$$

As θ goes to zero, corresponding to the second-order Butterworth or maximally flat condition, the peak gain goes to unity, while the frequency of the peak goes to zero; so theta is a zero-based measure of the peak gain of an SOS (the stage gain in an APGF). While negative values of θ , corresponding to $Q < \frac{1}{\sqrt{2}}$, have no sensible interpretation in these peak formulae, the general gain formula in the theta parameterization still works for negative θ , down to $\theta = -\frac{\pi}{2}$ or $Q = \frac{1}{2}$:

$$|H_1(i\omega)|^2 = \frac{1}{d^2} = \frac{1}{\left((\omega/\omega_n)^2 - \sin \theta \right)^2 + \cos^2 \theta}$$

Parameterizing by Peak Gain

Given a peak gain value as an independent parameter, one can back-solve for Q or θ as a dependent intermediate parameter.

$$\begin{aligned}
Q &= \sqrt{\frac{H_{1\max}^2 + \sqrt{H_{1\max}^4 - H_{1\max}^2}}{2}} \\
&= \frac{1}{\sqrt{2 - 2\sqrt{1 - H_{1\max}^{-2}}}} \\
\theta &= \arccos\left(\frac{1}{H_{1\max}}\right)
\end{aligned}$$

Substituting Q back into the magnitude-gain expression gives this result parameterized directly by $H_{1\max}$:

$$|H(i\omega)|^2 = \frac{1}{1 - 2\sqrt{1 - H_{1\max}^{-2}}(\omega/\omega_n)^2 + (\omega/\omega_n)^4}$$

One can parameterize directly by center frequency, too, instead of by natural frequency, which gives a particularly nice expression:

$$|H_1(i\omega)|^2 = \frac{1}{1 + (1 - H_{1\max}^{-2})((\omega/\omega_c)^4 - 2(\omega/\omega_c)^2)}$$

We will use direct parameterization by H_{\max} extensively in APGF results, especially as the independent variable on plots.

SOS Bandwidth

Using the symmetric x domain of the theta parameterization, we easily obtain the high and low band edge frequencies at which the gain-squared crosses the $1/a$ power points (e.g. $a = 2$ for half-power or 3-dB points, or $a = 10$ for 10-dB points). In terms of frequencies f normalized to natural frequency ($f = \omega/\omega_n = \sqrt{x}$):

$$\begin{aligned}
\{f_h^2, f_l^2\} &= \sin \theta \pm \cos \theta \sqrt{a-1} \\
&= f_c^2 \pm \frac{\sqrt{a-1}}{H_{1\max}} \\
&= 1 - 1/(2Q^2) \pm \frac{1}{Q} \sqrt{1 - 1/(4Q^2)} \sqrt{a-1}
\end{aligned}$$

The ubiquitous factor $\sqrt{a-1}$ is just unity for the half-power (3-dB) bandwidth of an SOS, but we keep it around because we'll need to generalize it to $\sqrt{a^{1/N}-1}$ for the APGF.

The bandwidth is just the difference of the band edges (multiply these normalized expressions by natural frequency in radians per second or Hertz to get actual bandwidth):

$$\begin{aligned} f_h - f_l &= \sqrt{\sin \theta + \cos \theta \sqrt{a-1}} - \sqrt{\sin \theta - \cos \theta \sqrt{a-1}} \\ &= \sqrt{f_c^2 + \frac{\sqrt{a-1}}{H_{1\max}}} - \sqrt{f_c^2 - \frac{\sqrt{a-1}}{H_{1\max}}} \\ &= \sqrt{1-1/(2Q^2) + \frac{1}{Q}\sqrt{1-1/(4Q^2)}\sqrt{a-1}} - \sqrt{1-1/(2Q^2) - \frac{1}{Q}\sqrt{1-1/(4Q^2)}\sqrt{a-1}} \end{aligned}$$

To an excellent approximation over much of the range of interest, $\sqrt{1+\varepsilon} - \sqrt{1-\varepsilon} \approx \varepsilon$, so we approximate these as:

$$\begin{aligned} f_h - f_l &\approx \sqrt{\sin \theta \cot \theta \sqrt{a-1}} \\ &\approx \sqrt{a-1}/(H_{1\max} f_c) \\ &\approx \frac{\sqrt{1-1/(4Q^2)}}{Q\sqrt{1-1/(2Q^2)}} \sqrt{a-1} \end{aligned}$$

The CF-relative bandwidth, i.e., $1/Q_{3\text{-dB}}$ or $1/Q_{10\text{-dB}}$ depending on the choice of a , is

$$\begin{aligned} \frac{(f_h - f_l)}{f_c} &= \sqrt{1 + \cot \theta \sqrt{a-1}} - \sqrt{1 - \cot \theta \sqrt{a-1}} \\ &= \sqrt{1 + \frac{\sqrt{a-1}}{H_{\max} f_c^2}} - \sqrt{1 - \frac{\sqrt{a-1}}{H_{\max} f_c^2}} \\ &= \sqrt{1 + \frac{\sqrt{1-1/(4Q^2)}}{Q(1-1/(2Q^2))} \sqrt{a-1}} - \sqrt{1 - \frac{\sqrt{1-1/(4Q^2)}}{Q(1-1/(2Q^2))} \sqrt{a-1}} \end{aligned}$$

Using the difference of roots approximation again we find the approximate relative bandwidth; we avoid the usual $1/Q$ approximation, since in our APGF application where $\sqrt{a-1}$ is replaced by $\sqrt{a^{1/N}-1}$, Q is typically too low for this approximation to be useful:

$$\begin{aligned}
\frac{(f_h - f_l)}{f_c} &\approx \cot \theta \sqrt{a-1} && \text{(theta parameterization)} \\
&\approx \frac{\sqrt{a-1}}{(H_{\max} f_c^2)} && (H_{\max}, \text{ CF parameterization}) \\
&\approx \frac{\sqrt{1-1/(4Q^2)}}{Q(1-1/(2Q^2))} \sqrt{a-1} && (Q \text{ parameterization}) \\
&\approx \frac{\omega_r \omega_n}{Q \omega_c^2} \sqrt{a-1} && \text{(mixed parameterization)}
\end{aligned}$$

We might prefer the ratio or log-ratio of band edges, rather than their difference, so we define the (exact) bandwidth in octaves as follows:

$$\begin{aligned}
\text{BW in octaves} &= \log_2(f_h/f_l) = \frac{1}{2} \log_2(f_h^2/f_l^2) \\
&= \frac{1}{2} \log_2 \left(\frac{\sin \theta + \cos \theta \sqrt{a-1}}{\sin \theta - \cos \theta \sqrt{a-1}} \right)
\end{aligned}$$

SOS Equivalent Rectangular Bandwidth

The “equivalent rectangular bandwidth” (ERB) of a filter is defined as the width of a rectangular filter whose height equals the peak gain of the filter and which passes the same total power as the filter (given a flat-spectrum input such as white noise or an impulse); i.e., in natural units for the SOS:

$$\begin{aligned}
\text{ERB} &= \frac{\int_0^\infty |H_1(i\omega)|^2 d\omega}{H_{1\max}^2} = \frac{\omega_n Q \pi / 2}{\left(Q / \sqrt{1-1/(4Q^2)}\right)^2} \\
&= \frac{\omega_n \pi (1-1/(4Q^2))}{2Q} && (Q \text{ parameterization}) \\
&= \frac{\omega_n \pi \cos^2 \theta}{2\sqrt{2}(1-\sin \theta)} && \text{(theta parameterization)}
\end{aligned}$$

Relative to the 3-dB bandwidth, the ERB is significantly wider, by a factor of nearly $\frac{\pi}{2}$ (but below Q of about 1.4, the 3-dB bandwidth is not well defined, because the peak gain is less than 3-dB):

$$\frac{\text{ERB}}{\text{BW}_{3\text{-dB}}} \approx \frac{\pi}{2} \sqrt{1 - 1/(4Q^2)} \sqrt{1 - 1/(2Q^2)}$$

Extending these results to the APGF involves integrals we don't know how to do in closed form for general N , but which we can evaluate for particular values of N up to at least 16. The ERB is much closer to the 3-dB bandwidth for high N , as will be shown numerically.

SOS Phase and Delay

Besides the magnitude, the phase of the transfer function is also of interest. The most useful view of phase is its derivative versus frequency, known as the group delay, which is closely related to the magnitude and avoids the need for trigonometric functions:

$$\begin{aligned} T_{g1} &= \frac{1+x}{\omega_n Q (1 + (1/Q^2 - 2)x + x^2)} \\ &= \frac{1+f^2}{\omega_n Q} |H_1|^2 \end{aligned}$$

The corresponding phase expression is never very simple; the best form we know without a branch cut in the middle of it is:

$$\phi_1 = -\text{arccot}\left(\frac{(1-f^2)Q}{f}\right)$$

By normalizing group delay relative to natural frequency, the delay can be made non-dimensional, or in terms of natural time units of the system (radians at ω_n), leading to a variety of simple expressions for delay at particular frequencies; notice that the maximum group delay only slightly exceeds the delay at CF:

$$\begin{aligned} \text{General Group Delay: } T_{g1} \omega_n &= \frac{1+f^2}{Q(1+(1/Q^2-2)f^2+f^4)} \\ &= \frac{1+f^2}{Q} |H_1|^2 \end{aligned}$$

$$\text{Group Delay at DC: } T_{g1}^{(DC)} \omega_n = 1/Q$$

$$\text{Group Delay at } \omega_n: T_{g1}^{(\omega_n)} \omega_n = 2Q$$

$$\text{Group Delay at CF: } T_{g1}^{(CF)} \omega_n = 2Q$$

$$\begin{aligned} \text{Maximum Group Delay: } T_{g1}^{(max)} \omega_n &= \frac{2Q}{2 - 8Q^2 \left(1 - \sqrt{1 - 1/(4Q^2)}\right)} \\ &\approx \frac{2Q}{1 - 1/(16Q^2)} \end{aligned}$$

$$\begin{aligned} \text{Norm. Freq. of Max. Group Delay: } f_{T_{peak}} &= \sqrt{2\sqrt{1 - 1/(4Q^2)} - 1} \\ &= \sqrt{2\omega_r/\omega_n - 1} \approx \omega_r/\omega_n \\ 0.644 &\approx \sqrt{\sqrt{2} - 1} < f_{T_{peak}} < 1 \end{aligned}$$

The similar expressions normalized to CF rather than to natural frequency look a bit more complicated, but may be more appropriate to consider since CF is a very useful characterization of a filter:

$$\text{Group Delay at DC: } T_{g1}^{(DC)} \omega_c = \sqrt{1 - 1/(2Q^2)} / Q$$

$$\text{Group Delay at CF: } T_{g1}^{(CF)} \omega_c = 2Q\sqrt{1 - 1/(2Q^2)}$$

$$\text{Maximum Group Delay: } T_{g1}^{(max)} \omega_c \approx \frac{2Q\sqrt{1 - 1/(2Q^2)}}{1 - 1/(16Q^2)}$$

SOS Dispersion

The difference between group delay at CF and at DC is what we call the low-side dispersion, which we also normalize relative to natural frequency or to CF:

$$\begin{aligned} \text{Low - side Dispersion: } (T_{g1}^{(CF)} - T_{g1}^{(DC)}) \omega_n &= (2Q - 1/Q) \\ (T_{g1}^{(CF)} - T_{g1}^{(DC)}) \omega_c &= (2Q - 1/Q)\sqrt{1 - 1/(2Q^2)} \end{aligned}$$

This measure of dispersion is the time spread (in normalized or radian units) between the arrival of low frequencies in the tail of the transfer function and the arrival of frequencies near CF, in response to an impulse. The maximum dispersion, to frequencies slightly above CF, is slightly greater, but more complicated to quantify.

These various observations on the SOS lead directly to corresponding features of the APGF.

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 4—Observations on the APGF Response

The APGF is just a cascade (i.e., function composition, or product of transfer functions) of N identical SOS stages, so the transfer function is just the N^{th} power of $H_1(s)$:

$$\begin{aligned} H(s) &= [H_1(s)]^N \\ &= \frac{[b^2 + \omega_r^2]^N}{[(s+b)^2 + \omega_r^2]^N} \quad (\text{Cartesian parameterization}) \\ &= \frac{1}{[(s/\omega_n)^2 + s/(\omega_n Q) + 1]^N} \quad (Q \text{ parameterization}) \end{aligned}$$

The real and imaginary components of the log gain, both log magnitude (or dB) and phase (or delay), are simply multiplied by N relative to the SOS.

As with the SOS, neither ω_n nor ω_r is the CF at which the APGF transfer function peaks (though ω_r is very close to the peak for the GTF).

As Flanagan [F60] discussed, the transfer function $H(s)$ can be implemented by a cascade of N identical two-pole filters ($N=3$ in his case). Assuming the constraint of unity gain at DC, the cascaded filters have only two free parameters: a frequency (ω_n or ω_r) and a Q or bandwidth (Q or b). The number of filters cascaded (or the number of times a filtering operation is applied), N , is the third parameter needed to characterize the APGF.

The APGF, therefore, is completely characterized as a complex-conjugate pair of order- N pole locations. It does not have the complexity of an additional phase parameter or corresponding zero locations. It also does not have a free gain parameter if we make the usual constraint of unity gain at DC.

The APGF can also be completely characterized by its magnitude transfer function (since it is a minimum-phase filter, the magnitude is a complete characterization), which we give here in the various parameterizations:

$$\begin{aligned}
|H(i\omega)| &= \sqrt{H(i\omega)H^*(i\omega)} \\
&= \frac{[\omega_r^2 + b^2]^N}{\left[(\omega_r^2 + b^2)^2 - 2\omega^2(\omega_r^2 - b^2) + \omega^4\right]^{N/2}} && \text{(Cartesian parameterization)} \\
&= \frac{1}{\left[1 + (1/Q^2 - 2)(\omega/\omega_n)^2 + (\omega/\omega_n)^4\right]^{N/2}} && \text{(} Q \text{ parameterization)} \\
&= \frac{1}{\left[\cos^2 \theta + \left((\omega/\omega_n)^2 - \sin \theta\right)^2\right]^{N/2}} && \text{(theta parameterization)} \\
&= \frac{1}{\left[1 - 2\sqrt{1 - H_{\max}^{-2/N}}(\omega/\omega_n)^2 + (\omega/\omega_n)^4\right]^{N/2}} && \text{(} H_{\max} \text{ parameterization)} \\
&= \frac{1}{\left[1 + \left(1 - H_{\max}^{-2/N}\right)\left((\omega/\omega_c)^4 - 2(\omega/\omega_c)^2\right)\right]^{N/2}} && \text{(} H_{\max}, \text{ CF parameterization)}
\end{aligned}$$

which are all exact at all frequencies ω .

The GTF, on the other hand, has a much more difficult form for its magnitude transfer function, due to the phase-dependent interference of two simple terms; as a result, it is usually characterized only approximately, as by Holdsworth [H88], de Boer and Kruidenier [BK90], and Carney [C93]. The exact GTF magnitude formula by Darling [D91] shows the approximation plus a complicated cross term that is difficult to understand.

The time-domain impulse response $h(t)$, or inverse transform of $H(s)$, is not as simple as a gamma-tone, but is still tractable. It can be described as a Bessel function of the first kind times a power of t times an exponential decay (structurally very similar to a gamma-tone, but with Bessel instead of cosine), or as a sum of N gamma-tones of all orders from 1 to N with phases rotating in multiples of 90° . Both interpretations will prove useful. See Appendix 6 on impulse responses for details.

Asymmetric Auditory Filters

One of the most striking features of auditory tuning curves is the asymmetry between the low-frequency and high-frequency “tails” or “skirts.” In addition, the degree of asymmetry,

the tip-to-tail ratio and the bandwidth are known to vary with signal level [GM90]—i.e., the filter is nonlinear—while the low-frequency tail behaves nearly linearly, maintaining an almost constant gain across levels [R71, R85, RRR92]. An advantage of the APGF is that it displays corresponding effects as one filter parameter is varied with level.

Auditory filter descriptions often include explicit asymmetry effects, as in the rounded exponential family of filters of Patterson et al. [PN80, PM86]. With respect to gamma-tones in particular, asymmetry has been recognized as an issue. Patterson et al. [P88] observe that “the gammatone filter has one notable disadvantage: the amplitude characteristic is virtually symmetric for orders equal to or greater than two, and there is no obvious way to introduce asymmetry.” And de Boer and Kruidenier [BK90] remark that “the remaining asymmetry indicates that the description in terms of the gamma tone is actually too simple.”

The APGF magnitude frequency response $|H(i\omega)|$, as noted above, is just a power of a quadratic function of “frequency squared,” and is therefore exactly symmetric about the peak in that space. In the (not squared) frequency domain, it has a predictable degree of asymmetry which is at least a qualitatively reasonable match to measured auditory filters. The degree of asymmetry observed within a limited range, e.g. within 30 dB of the peak, can be easily varied by the parameters of the APGF. See Figure A4.1 for an example.

The symmetry of the APGF on a frequency-squared scale, and its warping to the frequency scale, is analogous to the piecewise-linear frequency warping used with the original roex filter of Patterson and Nimmo-Smith [PN80] (“...upper and lower halves of the filter have the same shape but the frequency scale below the CF is expanded with respect to that above it”) or the roex(p,w,t) filter by Patterson and Moore [PM86] (“...these constraints characterize the asymmetry as a stretching of the frequency axis below the centre frequency of the filter relative to that above the centre frequency, and they effectively reduce the number of free parameters from six to four, which makes the fitting procedure more stable”). They are speaking of four parameters besides CF and gain, which are normalized. The GTF has only three such parameters (order, bandwidth, and phase).

The recently developed “Gammachirp Filter” (GCF) of Irino [I95], approaching from a completely different perspective, is a close relative of the GTF with one additional parameter, so it also has four parameters beyond CF and gain: “The gammachirp function fits the experimental results better than the gammatone function. This is predictable since the number of parameters increases. However, the gammachirp function is derived as the optimal function under the hypothesis and, thus, is a more reasonable approximation than other functions with additional unknown parameters.”

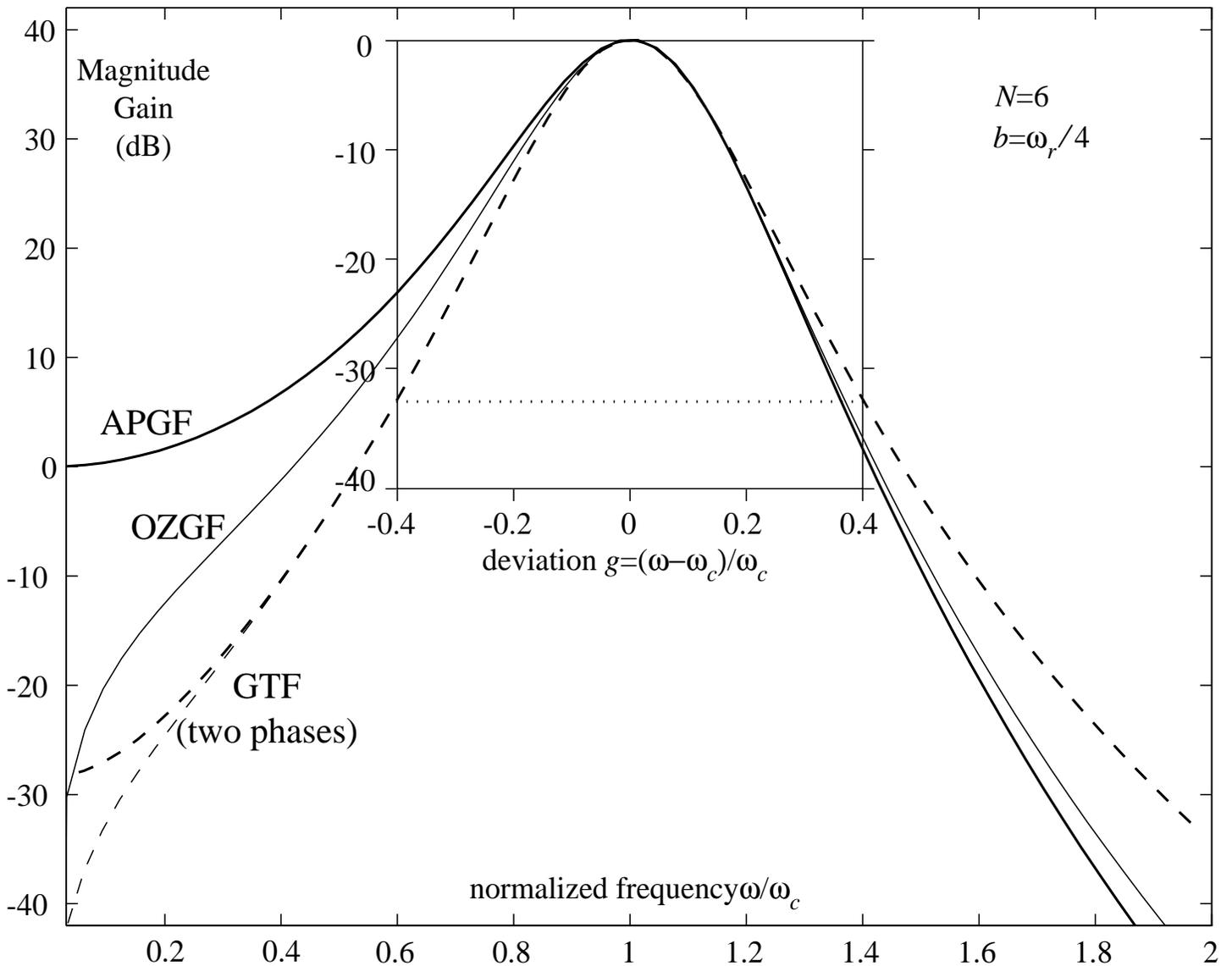


Figure A4.1. Comparison of magnitude transfer functions of the nearly symmetric GTF and the asymmetric APGF and OZGF, on a linear frequency scale normalized to CF. The inset axes show the typical range of data available from psychophysical methods; in this limited range, the apparent asymmetry is small.

By contrast, the APGF or OZGF achieves a result very much like the GCF, in terms of time-domain chirping and frequency-domain asymmetry, using only two free parameters besides CF and gain: N and Q . The level dependence of gain, bandwidth, symmetry, and chirping are correctly coupled via Q variation.

The conventional GTF magnitude frequency response is generally approximated as being symmetric about the peak in frequency space. Varying the phase parameter can make it asymmetric in either direction, but only by a very little as Patterson and Nimmo-Smith

observed [P88 Annex B] “for orders equal to or greater than two.” Varying the bandwidth parameter has a similarly small and non-monotonic effect on the asymmetry. In either case, the greatest relative variation occurs in the low-frequency tail of the response, where the ideal auditory filter would be least sensitive to varying parameters. Figure A4.1 shows that the typical amount of asymmetry that would be noticed within the frequency range of psychophysical data is not extreme.

Peak Gain or Tip-to-Tail Ratio

The low-frequency tail of the APGF has approximately constant gain (unity gain by our conventions), independent of other parameters. Variation of the bandwidth or Q parameter will change the peak gain, the tip-to-tail ratio, the asymmetry, and the bandwidth, all without affecting the approximately constant-gain linear behavior of the tail, just as observed in cochlear mechanics. By defining the DC gain to be unity, we unify the concepts of peak gain and tip-to-tail ratio into a single numerical parameter that we call H_{\max} , which can serve as a useful alternative to Q or b or θ for characterizing the degree of damping, sharpness, or asymmetry of an auditory filter. Auditory filter formulations that do not have a stable low-frequency tail cannot be similarly characterized. H_{\max} relates simply to the other parameters, and can be thought of the degree of “active gain” of the cochlea, which might be around 50 dB at low signal levels and reduce to near 0 dB at high signal levels.

Filter Shape vs. Order

Figure A4.2 shows a plot of several different orders of APGF for a fixed H_{\max} (with frequencies ω_r or ω_n adjusted to make the CF peaks coincide exactly). Starting with $N = 1$, a Q of about 100 is needed to give a peak gain of 100 (40 dB), so the peak is extremely narrow. Successive doublings of N give broader peaks, approaching a limiting shape. The shape of an empirical magnitude transfer function provides very little constraint on the tradeoff between Q and N unless the absolute H_{\max} is known, and then only a relatively weak constraint when the filter width is near the limiting width.

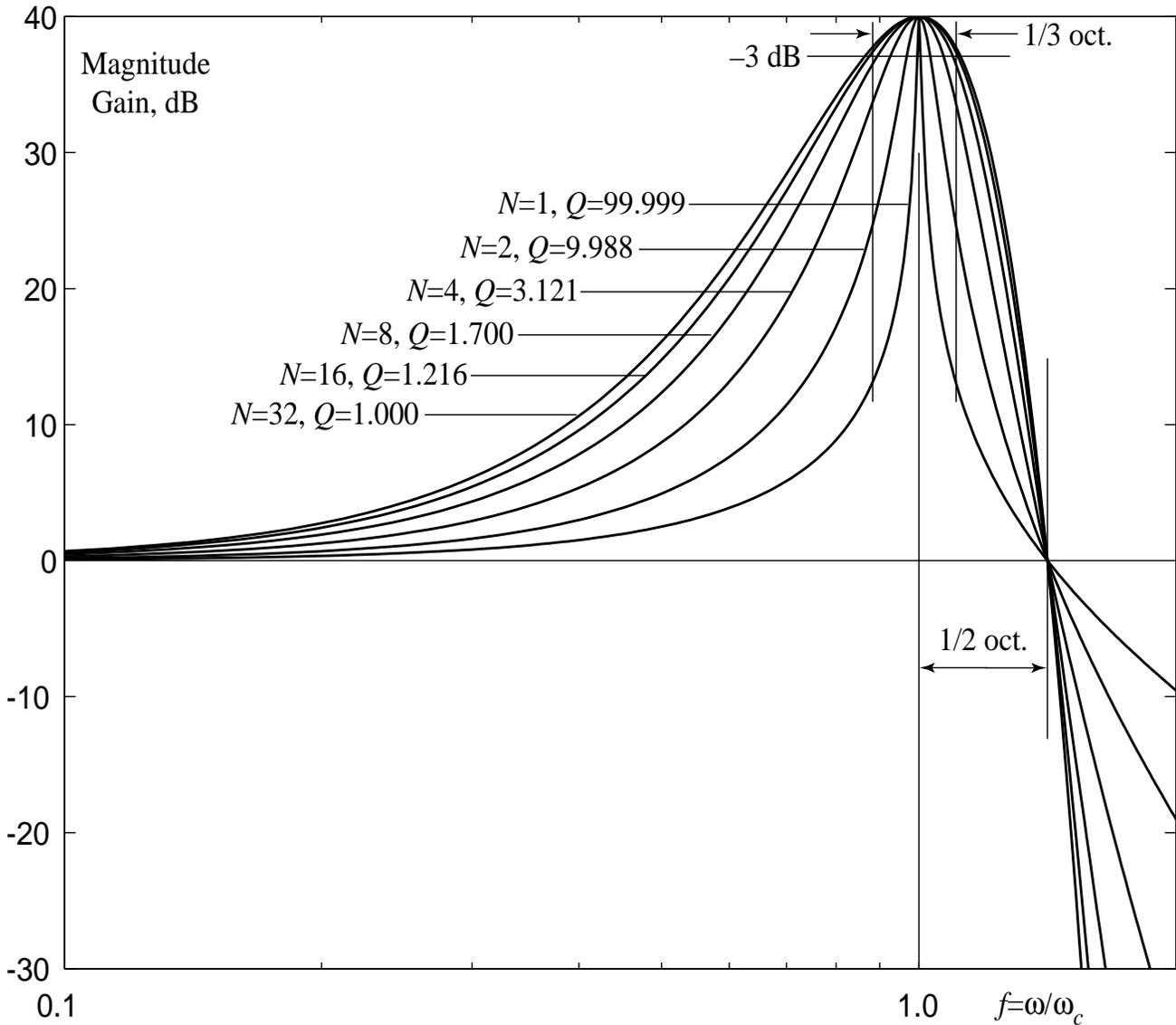


Figure A4.2. A plot of several different orders of APGF for a fixed H_{\max} of 100 (40 dB), with frequencies ω_r or ω_n adjusted to make the CF peaks coincide exactly. Starting with $N = 1$, a Q of about 100 is needed to give a peak gain of 100 (40 dB), so the peak is extremely narrow. Successive doublings of N give broader peaks, approaching a limiting shape. The 3-dB band edge level and a representative width of 1/3 octave are indicated, as is the 1/2 octave distance from CF to the point where the gain matches the DC gain.

The OZGF and Tail Slope

The flat unity gain tail at very low frequencies is a consequence of the all-pole formulation, but it is not necessarily desirable. A sloped but otherwise linear tail can be obtained by

adding an additional fixed filter to a nonlinearly-parameterized APGF. For example, a better fit to basilar membrane velocity, as opposed to displacement, would be obtained by adding a differentiator—a zero at $s = 0$. This filter is called the One-Zero Gammatone Filter (OZGF), as it corresponds to a GTF with all but one of the zeros removed (for some ϕ). With $N = 3$, the OZGF is one of the models used by Flanagan [F60].

More generally, the uncontrolled zero placement of the GTF can be replaced by one or more carefully selected zeros—or by any fixed filter, perhaps designed to model the outer and middle ears.

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 5— Gain/Bandwidth/Delay/Order Relations

Since the APGF is simply the N th power of the well-studied two-pole filter, equivalent to a cascade of N identical two-pole stages, it is straightforward to characterize it in terms of how its center frequency, peak gain, delay, bandwidth, etc., depend on its parameters.

Peak Gain

The peak gain H_{\max} is simply the N th power of the peak gain of each of N stages, $H_{1\max}$:

$$\begin{aligned}
 H_{\max} &= \left[\frac{\omega_r}{2b} \left(1 + \frac{b^2}{\omega_r^2} \right) \right]^N && \text{(Cartesian parameterization)} \\
 &= \left[\frac{Q}{\sqrt{1 - 1/(4Q^2)}} \right]^N && \text{(Q parameterization)} \\
 &= \left[Q \frac{\omega_n}{\omega_r} \right]^N && \text{(mixed parameterization)} \\
 &= \left[\frac{1}{\cos \theta} \right]^N && \text{(theta parameterization)}
 \end{aligned}$$

The expressions are all N th powers of the stage gain, which is near Q or near $\omega_r/(2b)$ for high enough Q , so the first three expressions are in terms of Q or $\omega_r/(2b)$, times correction factors somewhat greater than 1. Typically N is high enough that Q is low and the correction factor is significant. Regarding Q as a descriptor of a “polar” parameterization, the first three expressions are pure Cartesian, pure polar, and mixed, respectively; as often happens in these things, the mixed expression looks simplest. The fourth, the “theta” or “true polar” parameterization, gives an even simpler result.

The frequency at which the peak occurs, or center frequency ω_c , is the same as for the underlying SOS and has similarly many possible forms:

$$\begin{aligned}
 \omega_c &= \omega_r \sqrt{1 - b^2 / \omega_r^2} && \text{(Cartesian parameterization)} \\
 &= \omega_n \sqrt{1 - 1 / (2Q^2)} && \text{(} Q \text{ parameterization)} \\
 &= \omega_r \sqrt{1 - 1 / (4Q^2 - 1)} && \text{(mixed parameterization)} \\
 &= \omega_n \sqrt{\sin \theta} && \text{(theta parameterization)}
 \end{aligned}$$

The expressions are in terms of the approximate center frequency ω_r or ω_n , times a correction factor that depends on the bandwidth, in Cartesian, Q , mixed, and theta representations. Notice that the center frequency is less than ω_r —by about half as much as it is below ω_n . In the GTF, by contrast, (except for the case $N = 1$) the center frequency is very close to ω_r .

The Q and order can be traded off to maintain a constant peak gain H_{\max} by keeping a constant $N \log(\cos \theta)$ or $N \log(Q / \sqrt{1 - 1 / (4Q^2)})$.

Bandwidth

The various bandwidth results presented for the SOS carry over directly for the APGF, with the power ratio a replaced by $a^{1/N}$; e.g. a power ratio of 2 for the 3-dB bandwidth of an APGF corresponds for a ratio $2^{1/N}$ per stage of the order- N APGF; therefore, we do not repeat them here.

The formula for the bandwidth in octaves is exact for the APGF, but blows up when the gain peak gets so low that the lower band edge goes to zero frequency. A useful approximation that doesn't blow up considers only the ratio of the upper band edge to the center frequency, and may be more appropriate to the more nearly symmetric OZGF, in which the low band edge never goes to zero:

$$\begin{aligned}
 BW \text{ in octaves} &\approx 2 \log_2(f_h / f_c) = \log_2(f_h^2 / f_c^2) \\
 &= \log_2 \left(\frac{\sin \theta + \cos \theta \sqrt{a^{1/N} - 1}}{\sin \theta} \right) \\
 &= \log_2 \left(1 + \cot \theta \sqrt{a^{1/N} - 1} \right)
 \end{aligned}$$

This approximation goes to exactly 1 octave under the conditions that the exact APGF formula above goes to infinity, since the high side gain of a two-pole filter matches the DC gain at a frequency exactly one-half octave above the peak frequency (i.e., $f_h^2 = 2f_c^2$ when $f_l = 0$).

For bandwidths much less than 1 octave, the log is small enough that a first-order Taylor expansion is a useful approximation. A similar approximation converts the N^{th} root into a more useful form, for this approximate result:

$$BW \text{ in octaves} \approx \frac{\sqrt{(\log a)/\log 2}}{\sqrt{N} \tan \theta}$$

The latter approximation makes it clear that $\tan \theta$ can be traded off against \sqrt{N} to maintain a particular bandwidth—i.e. that the product $\sqrt{N} \tan \theta$ is inversely proportional to the bandwidth in octaves.

For a wide range of orders and gains, the bandwidth will be between 1/4 and 1/2 octave. Evaluating bandwidth change across a wide range of levels and gains should help to constrain the model order; if the order is too low, bandwidth will change too much with level, and if it's too high, bandwidth will be too high or change too little.

Equivalent Rectangular Bandwidth

The “equivalent rectangular bandwidth” (ERB) of an APGF is a little higher than the 3-dB bandwidth, except when the lower band edge goes to DC. We plot their ratio in Figure A5.1.

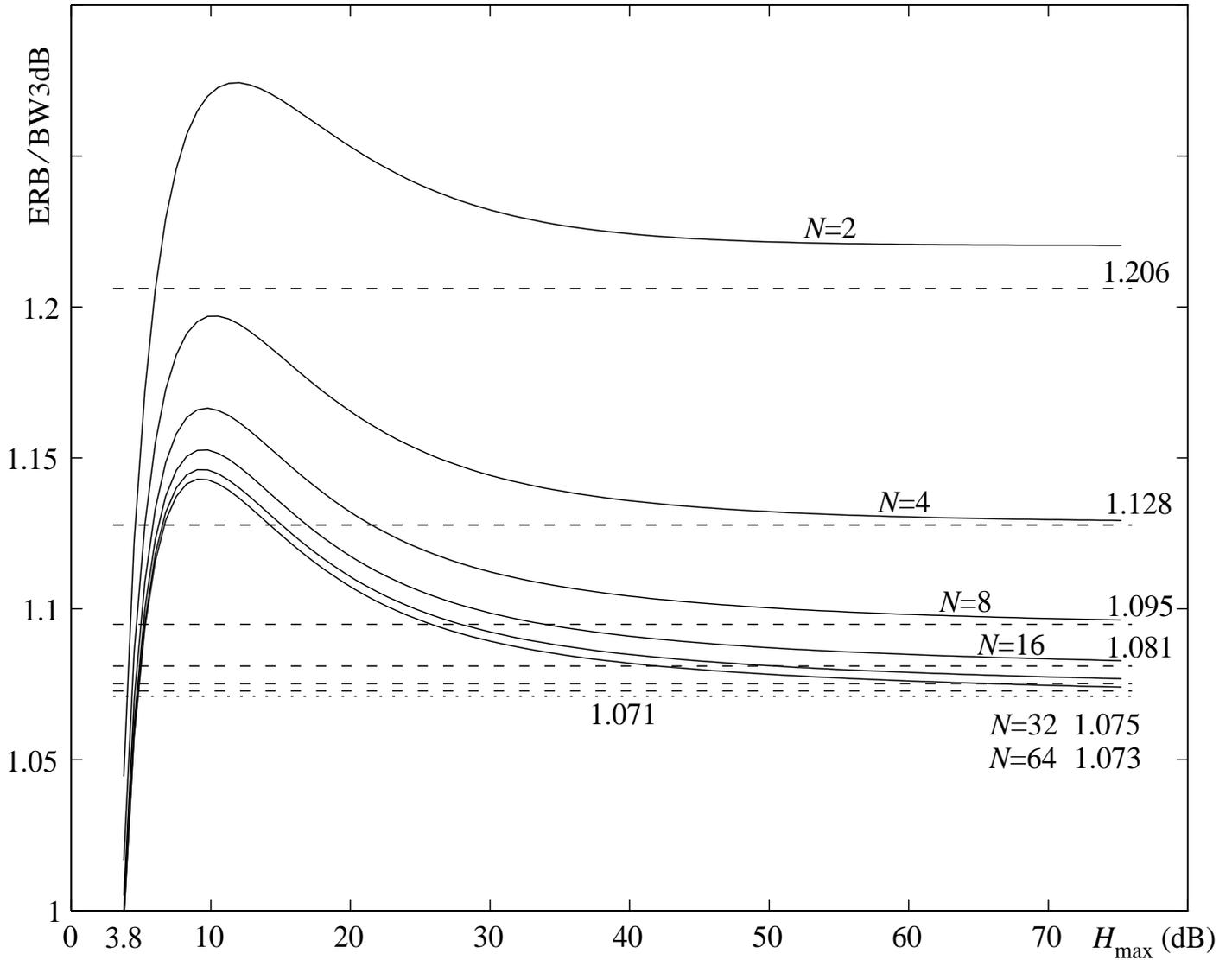


Figure A5.1. The ratio of equivalent rectangular bandwidth (ERB) of the APGF to the 3-dB bandwidth, vs. the maximum gain H_{\max} , with parameter N . The horizontal dashed lines are empirical asymptotic ratios given by the formula $1.071(1 + 0.3/N^{1.25})$, which appear to fit reasonably well for $N > 2$; the limit for high N is shown dotted. A good first-order “rule-of-thumb” is that the ERB is about 10% greater than the 3-dB bandwidth.

The general ERB formula is in terms of a polynomial $P_N(Q)$, found by contour integration of $H(s)H(-s)$, which we have calculated up to $N = 16$ (the general pattern of coefficients is not evident and needs more work to derive):

$$\begin{aligned} \text{ERB} &= \frac{\int_{-\infty}^{\infty} |H(i\omega)|^2 d\omega}{2H_{\max}^2} = \frac{\omega_n \pi P_N(Q)}{2\left(Q/\sqrt{1-1/(4Q^2)}\right)^{2N}} \\ &= \frac{\omega_n P_N(Q) \pi \left(1-1/(4Q^2)\right)^N}{2Q^{2N}} \quad (Q \text{ parameterization}) \end{aligned}$$

$$P_1(Q) = \frac{1}{1}Q$$

$$P_2(Q) = \frac{2}{4}(Q + Q^3)$$

$$P_3(Q) = \frac{3}{16}(2Q + 2Q^3 + 2Q^5)$$

$$P_4(Q) = \frac{4}{64}(5Q + 5Q^3 + 6Q^5 + 5Q^7)$$

$$P_5(Q) = \frac{5}{256}(14Q + 14Q^3 + 18Q^5 + 20Q^7 + 14Q^9)$$

$$P_N(Q) = \frac{N}{4^{N-1}}(K_N Q + K_N Q^3 + ?Q^5 + ?\dots? + K_N Q^{2N-1})$$

where $K_N = \{1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, \dots\}$

$$= \frac{\prod_{n=2}^N (4n-6)}{N!} \quad (\text{empirically, for } N \geq 2)$$

It is difficult to imagine a good approximate form for the ERB, since all the terms of the polynomial matter about equally for Q values near unity, which is the region of interest, and we don't even know the form of the coefficients.

Darling [D91] was able to produce an analytic formula for the ERB of the GTF, but then, due to the complexity, fell back on the usual symmetric approximation, yielding simply a function of N times b , as in the approximate 3-dB bandwidth; the APGF is not quite so simple, due to its asymmetry and tail.

Delay and Dispersion

Since a differentiator has zero group delay, the APGF and OZGF have identical group delays, equal to N times the group delay of a single two-pole stage. Using the expression for group delay of a stage, multiplying by N , and normalizing to natural frequency:

$$\begin{aligned} T_g \omega_n &= \frac{N(1+f^2)}{Q[1+(1/Q^2-2)f^2+f^4]} \\ &= \frac{N(1+f^2)}{Q} |H_1|^2 \end{aligned}$$

The most appropriate characterization of the delay of a bandpass filter is the group delay at CF. A two-pole stage has a normalized group delay at CF of simply $2Q$, as may be verified by substituting the frequency of the gain peak into the above expression. An alternate normalization, by CF rather than by natural frequency, leads to a slightly more complex expression that shows how Q and order may be traded off to maintain a constant delay in a constant-CF APGF:

$$T_g^{(\text{CF})} \omega_n = 2NQ$$

$$T_g^{(\text{CF})} \omega_c = 2NQ\sqrt{1-1/(2Q^2)}$$

Otherwise similar models can have very different overall delays, so it may be more useful to characterize a filter by how the delay changes across the passband; this difference in delay at different frequencies is known as dispersion. For auditory filters, a simple and useful characterization is the difference between the group delay at CF and in the low-frequency tail—call this the low-side dispersion; the change in delay above CF may also be useful, but is much smaller. A two-pole stage has a normalized group delay at DC (representative of the low-frequency tail) of $1/Q$. The APGF low-side dispersion is thus:

$$\left(T_g^{(\text{CF})} - T_g^{(\text{DC})}\right)\omega_n = N(2Q - 1/Q)$$

$$\left(T_g^{(\text{CF})} - T_g^{(\text{DC})}\right)\omega_c = N(2Q - 1/Q)\sqrt{1-1/(2Q^2)} = 2NQ\left(1-1/(2Q^2)\right)^{3/2}$$

The constant-dispersion tradeoff between Q and N resembles that for delay, but with a steeper Q dependence. For $Q < 1$, the low-side dispersion is less than half the CF delay. For fixed peak gain and very high order, $Q \approx \frac{1}{\sqrt{2}}$ and the dispersion per stage goes to zero much more rapidly than the delay per stage. Yet the total dispersion continues to increase, slowly, for high N .

In Figure A5.2 we show dispersion vs. peak gain for various N . In this case, the behavior for high N is not asymptotic; rather, the total dispersion continues to increase with N once N is high enough. Dispersion data, obtainable by mechanical and physiological (revcor) methods, therefore provides a potentially useful additional constraint on the order fit. The curves for orders 8 to 32 are very close to each other, while both higher and lower orders have significantly different dispersion/gain relations.

Figure A5.3 shows the frequency-dependent group delay that the dispersion values of Figure A5.2 are computed from.

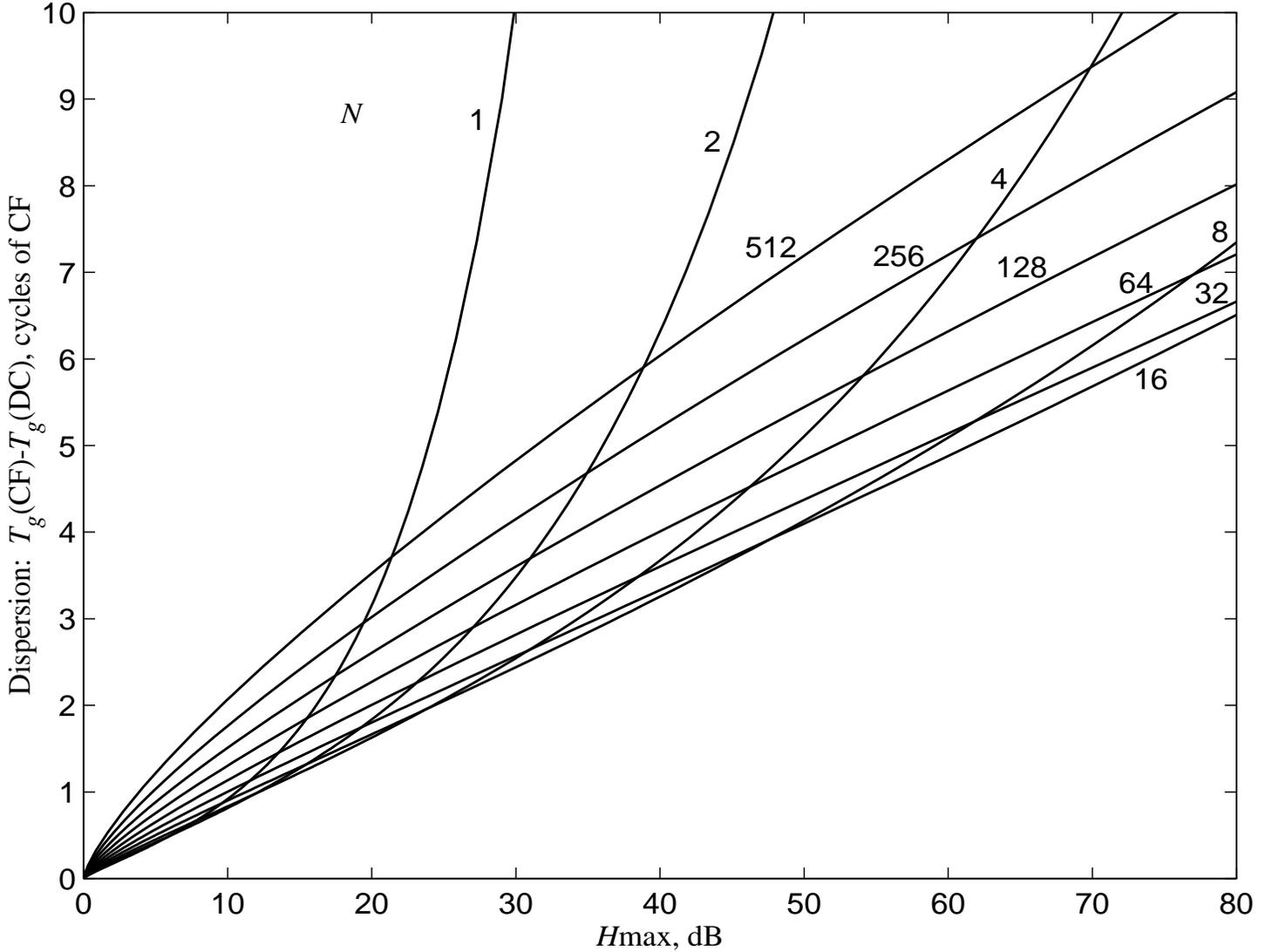


Figure A5.2. Dispersion vs peak gain for various N . The behavior for high N is not asymptotic; rather, the total dispersion continues to increase with N once N is high enough for the particular peak gain value.

The behavior of group delay above CF may also provide an additional model-fitting constraint, since the APGF group delay has a peak at a frequency above CF by a ratio that depends on Q :

$$\frac{f_{Tpeak}}{f_{Hpeak}} = \frac{\sqrt{2\sqrt{1-1/(4Q^2)}-1}}{\sqrt{1-1/(2Q^2)}}$$

However, it is difficult to get good phase or delay data far enough above CF to know whether this feature of the model will be useful. So far we have not seen any clear indication of a group delay maximum above CF, which only means that the APGF model order should not be low.

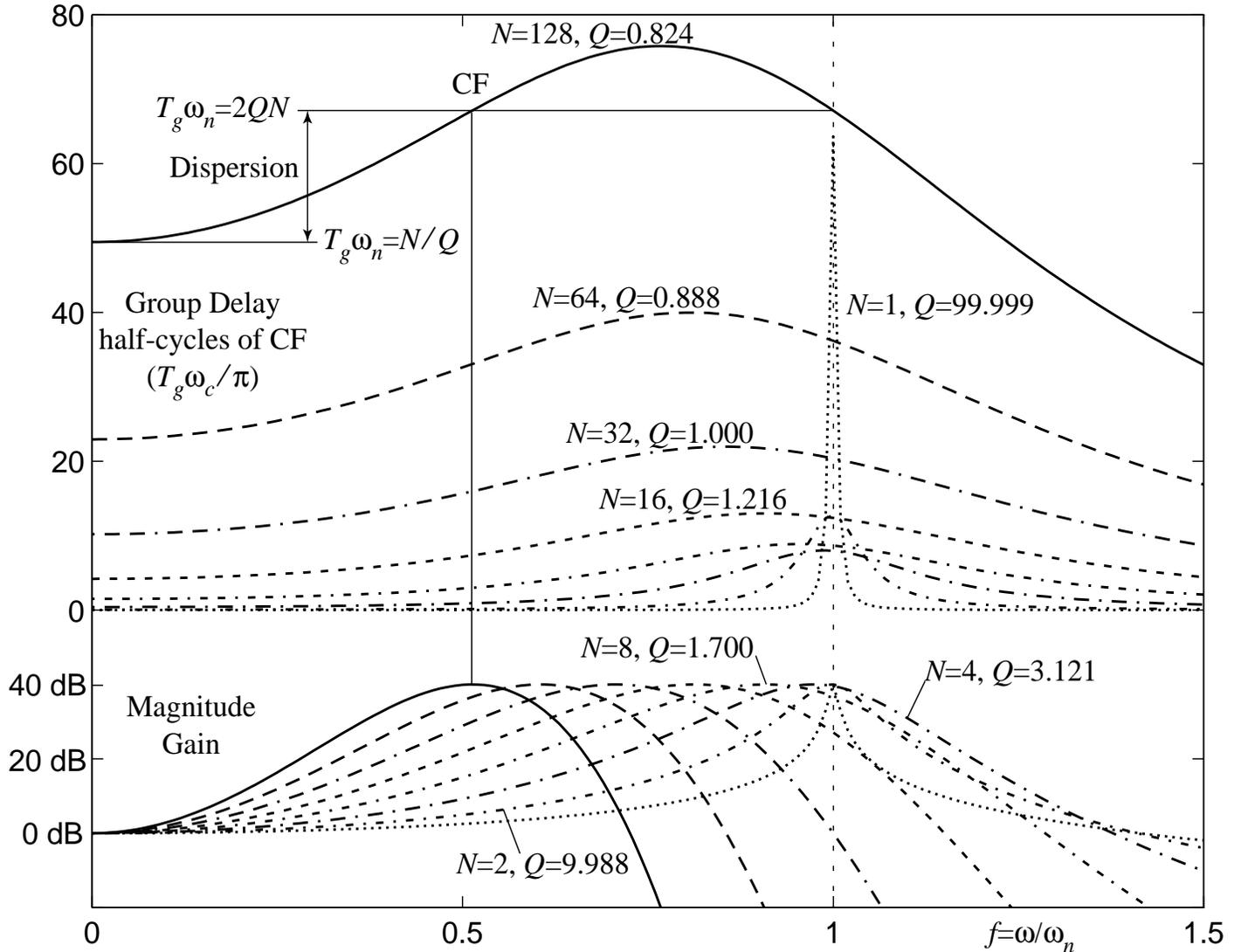


Figure A5.3. Group Delay and Magnitude Gain of the APGF for N being powers of 2 from 1 to 128, as a function of frequency; conditions similar to Figure A4.2, but with natural frequencies aligned instead of center frequencies. The equality relation between group delay points at natural frequency and at center frequency is shown, and the dispersion value used in calculations is shown. Note that group delay and dispersion continue to increase with large N , even after the delay values are corrected for the decreasing center frequency.

Order Constraints and Tradeoffs

Figure A5.4 shows contours, in the N vs. Q space, of constant gain, constant bandwidth, and constant dispersion. Over a wide range, they are very nearly parallel, which means that any compatible combination of these constraints that can be satisfied at some order can be nearly satisfied for a wide range of orders.

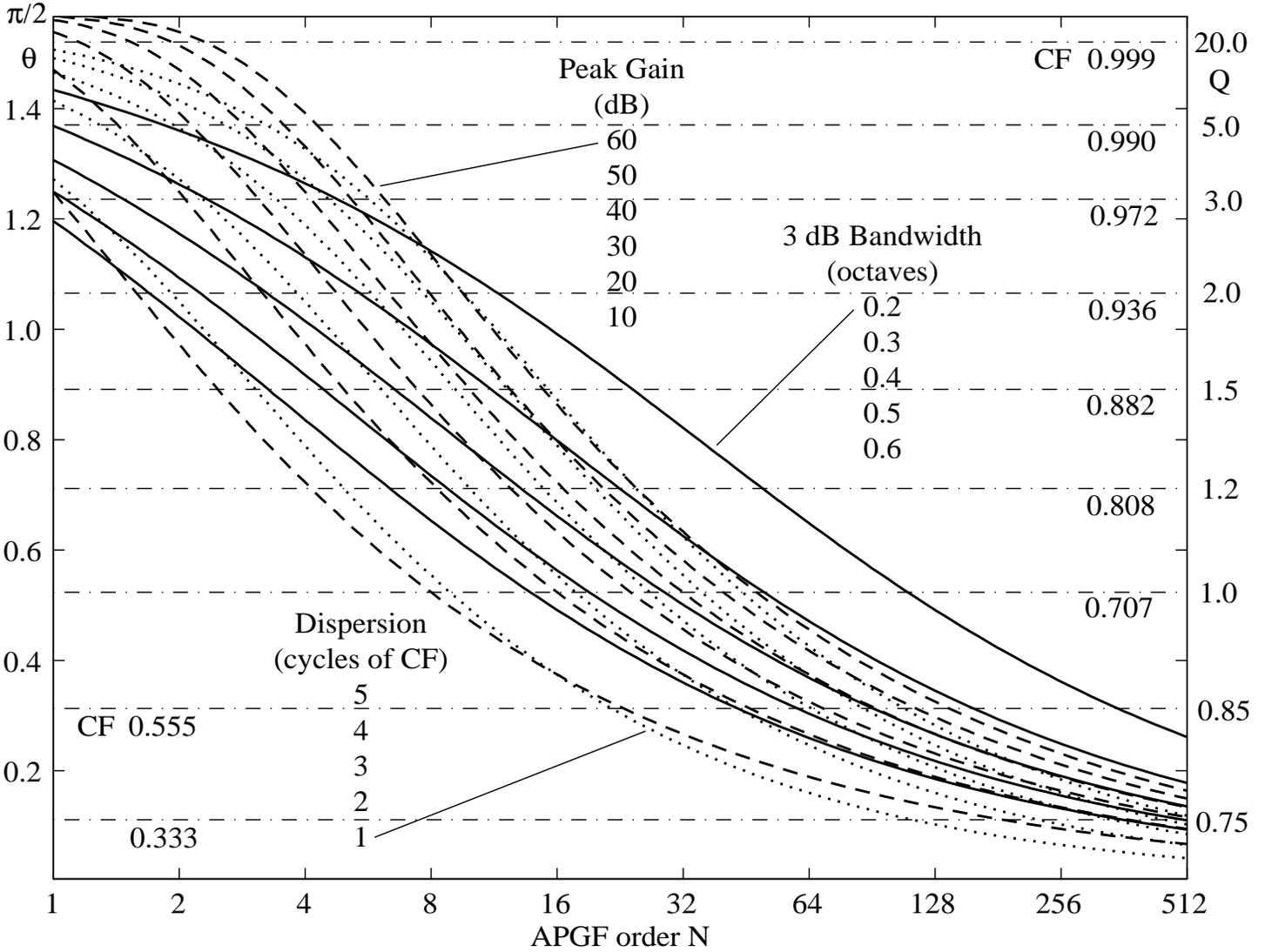


Figure A5.4. Contours, in the N vs. Q space, of constant bandwidth (solid), constant gain (dashed), and constant dispersion (dotted). The vertical axis is scaled linearly with θ , twice the angle of the poles from a 45-degree line. Selected Q values and corresponding relative CF values are indicated on horizontal contours.

Fitting a whole family of APGF's to a family of auditory filters measured at different sound levels may lead to an order estimate that works throughout the family. To do a good job, we probably need at least a good model of how the filter's peak gain varies with level, as obtainable from Basilar Membrane mechanics measurements, in combination with bandwidth measurements from mechanical, physiological, or psychoacoustic methods. For example, in Figure A5.5 we plot gain against bandwidth for several values of N , such that the curve with the most appropriate shape could be chosen to fix an order; however, the curves approach an asymptote for high order, so if the best curve is the asymptotic curve, any sufficiently high value of N will do.

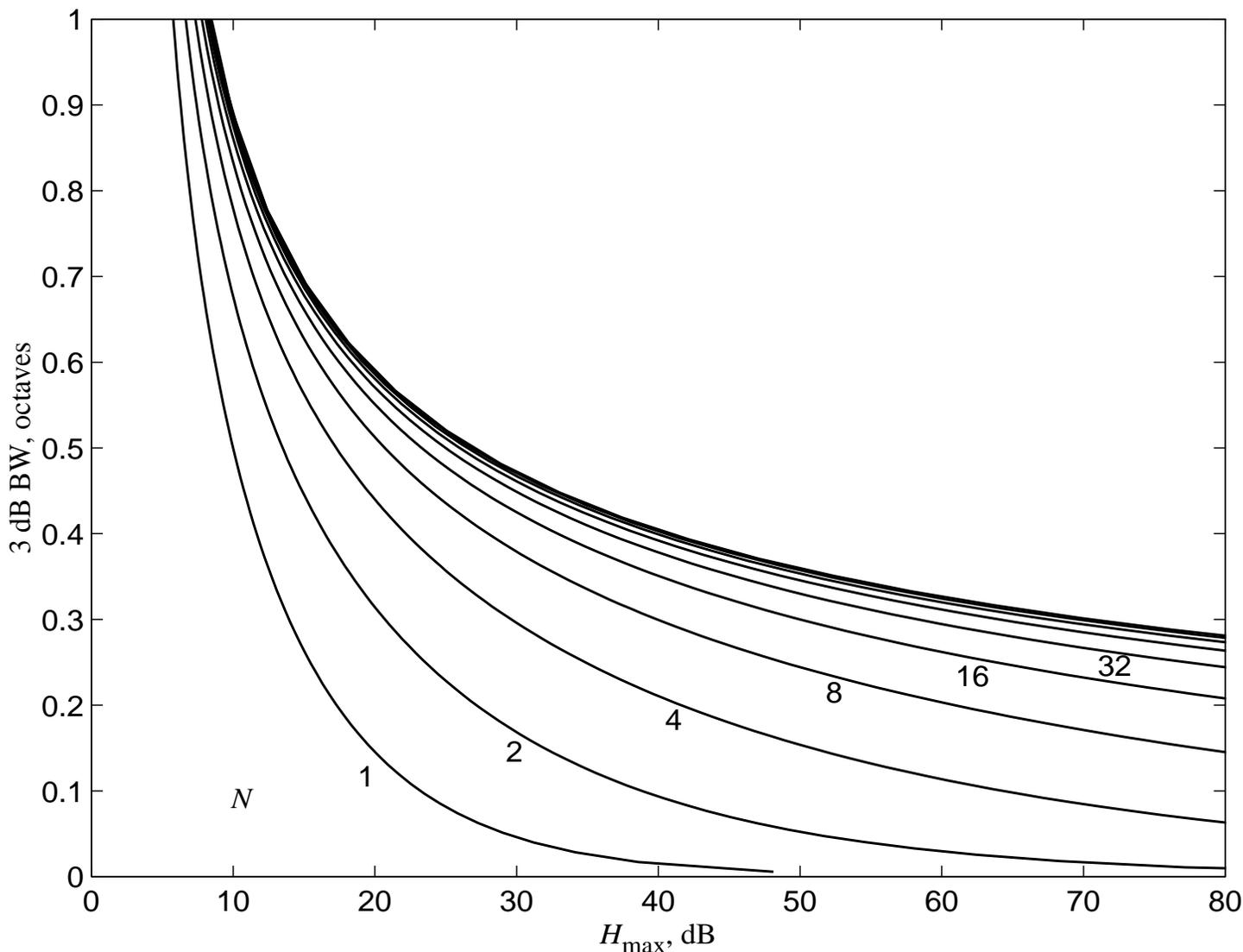


Figure A5.5. Bandwidth versus peak gain for several values of N . The curves approach an asymptote for high order.

In terms of psychophysical data, the 3-dB bandwidth of an auditory filter may be as low as 11% of CF, or about 0.14 octave. Then according to Figure A5.5, an APGF with $N = 4$ gives bandwidths that are too narrow for $H_{\max} > 50$ dB, while $N = 8$ yield bandwidths that are possibly a little too wide.

The variation of CF with peak gain may also be useful in fitting experimental data, once the peak gain has been related to level. Figure A5.6 shows CF (relative to ω_n), and corresponding pole Q , versus peak gain H_{\max} , for various model orders.

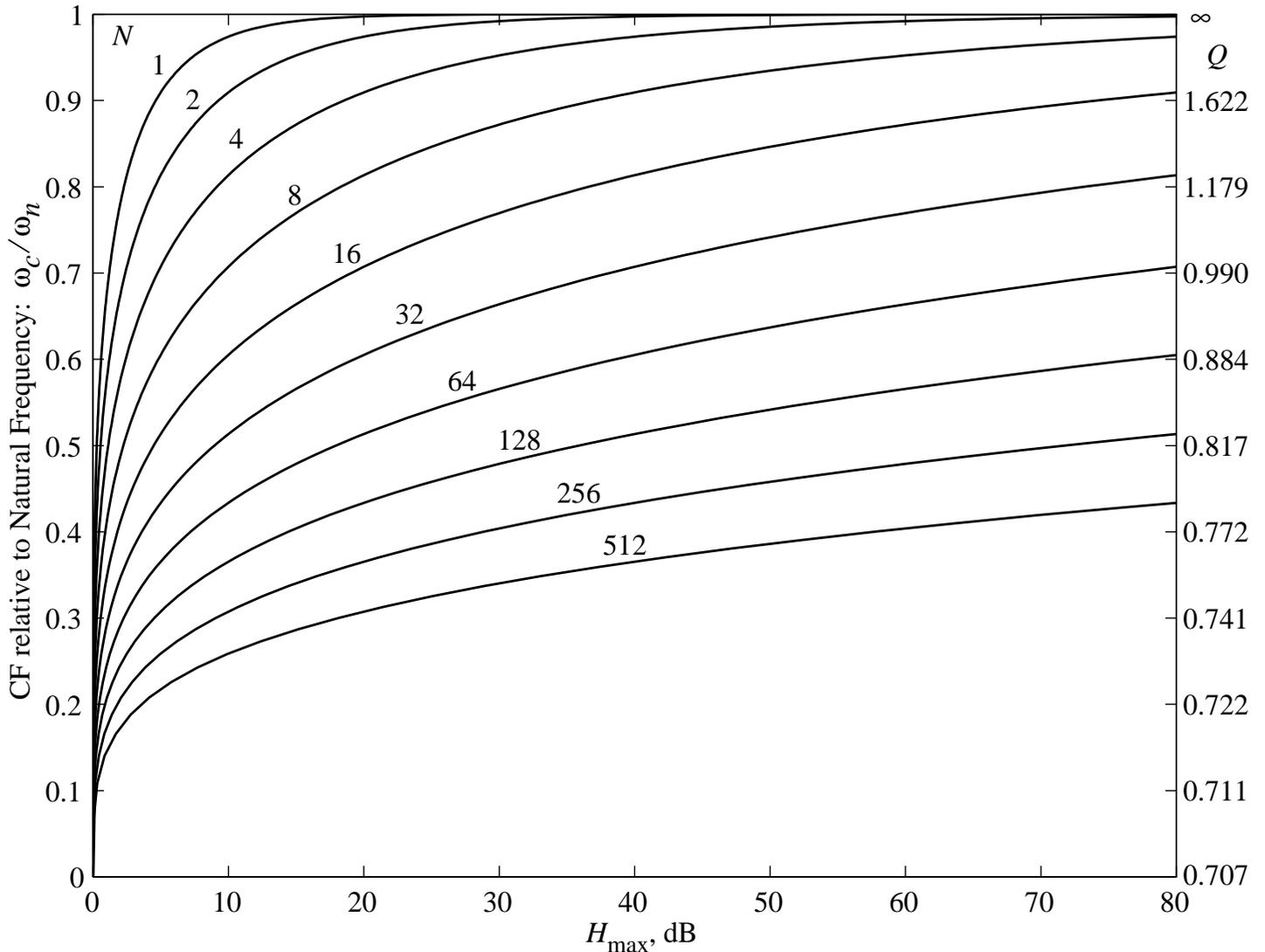


Figure A5.6. CF versus peak gain for several values of N , illustrating a range of possible dependences of CF on gain, and hence indirectly on level, under the assumption of constant natural frequency.

ringing versus Chirping

In a dispersive medium, the energy of an impulse input is smeared out in time, such that different frequencies come out at different times during the impulse response. Dispersion can be characterized in two ways: as a variation in group delay as a function of frequency, as discussed above, or as a variation in instantaneous frequency as a function of time, especially if group delay varies monotonically within the passband of the system. This latter view characterizes the impulse response as a chirp: a sine-like waveform with continuously varying frequency.

The impulse response of the GTF “rings” with a constant interval between zero crossings, and therefore approximately a constant instantaneous frequency. To within an excellent approximation, the peak of the group delay occurs at the same frequency as the peak of the magnitude response, so frequencies above and below CF arrive together and average out. The APGF and OZGF, on the other hand, have increasing group delay through the passband, so the impulse response starts with low frequencies and “glides” from well below CF to somewhat above CF, approaching the ringing frequency of the poles.

To compute instantaneous frequency of a real waveform, the usual approach is to take the rate of change of the phase of the complex Hilbert transform, or analytic extension of the real waveform to the complex domain. The “analytic envelope” can be computed at the same time, as the magnitude of the complex Hilbert transform.

Figure A5.7 shows the glide of the APGF's chirping impulse response for various orders with H_{\max} of 100 (40 dB); analytic envelopes of the impulse responses are also shown. The increasing dispersion at high order is seen as a glide over a wider frequency region during the time that the envelope is significantly nonzero.

Frequency glides, or chirps, have been observed in cochlear mechanics impulse response data by Ruggero [R85], in revcor data by Møller and Nilsson [MN79], and in various mechanical and physiological data sets by de Boer and Nuttall [BN96???]. Matching de Boer and Nuttall's observed glides requires a fairly high order APGF (perhaps $N = 32$ or higher).

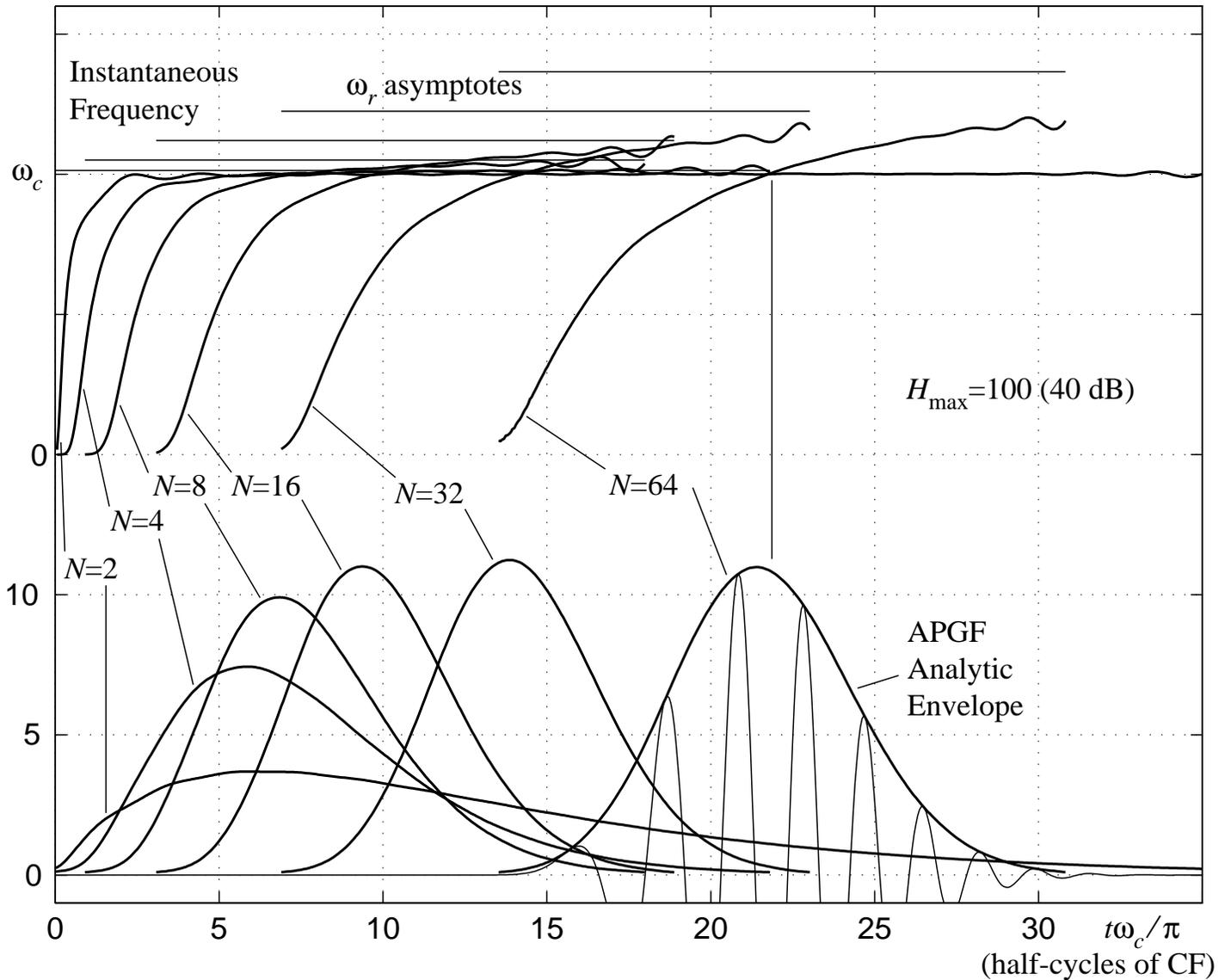


Figure A5.7. Instantaneous frequency and analytic envelopes of APGF impulse responses, for varying order and for a fixed H_{\max} of 100 (40 dB), with frequencies ω_r or ω_n adjusted to make the CF values coincide exactly. These are the same conditions used for the filter shape comparison of Figure A4.2, except that the $N=1$ case is omitted here due to its abrupt onset that can not be analytically extended reasonably. Higher N values are also included to show that the dispersion does continue to increase, slowly, with N . The ringing frequencies shown for $N > 2$ are approximate asymptotes of the instantaneous frequency. Curves are trimmed to the time region for which the envelope exceeds 0.1, or about 40 dB below the peak amplitude. One APGF impulse is shown (bottom, light curve) to show the relation between zero-crossing intervals and instantaneous frequency. The “glide” in instantaneous frequency is an alternative view of the dispersion, or variation of time delay with frequency, that was plotted in Figures A5.2 and A5.3.

— DRAFT of 29 Jan. 96 — do not distribute

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
Apple Computer, Inc.
Cupertino, CA 95014 USA

email: lyon@apple.com
fon: (408) 974-4245
fax: (408) 974-8414

Appendix 6— Impulse Response of the APGF

The time-domain impulse response $h(t)$, or inverse transform of $H(s)$, is not as simple as the gamma-tone. For general N , Mathematica™ [W91] will tell you that it's a Bessel function of the first kind times a power of t times an exponential decay—structurally very similar to a gamma-tone, but with Bessel instead of cosine; or see Abramowitz and Stegun [AS72 transform 29.3.57 in combination with property 29.2.12]:

$$h(t) = At^{N-\frac{1}{2}} \exp(-bt) J_{N-\frac{1}{2}}(\omega_r t) \quad \text{where } A = \frac{(b^2 + \omega_r^2)^N \sqrt{\pi}}{(2\omega_r)^{N-\frac{1}{2}} \Gamma(N)}$$

The Bessel function of the first kind of half-integer order is closely related to the Spherical Bessel function of the first kind: $j_n(z) = \sqrt{\frac{\pi}{2z}} J_{n+\frac{1}{2}}(z)$ [AS72 10.1.1], such that the impulse can be re-expressed without the fractional orders (from here on we use powers of the nondimensionally scaled time $\omega_r t$ to simplify other factors):

$$h(t) = A(\omega_r t)^N \exp(-bt) j_{N-1}(\omega_r t) \quad \text{where } A = \left[1 + b^2/\omega_r^2\right]^N \frac{\omega_r}{2^{N-1} \Gamma(N)}$$

Alternatively, for particular integer values of N , Mathematica™ [W91] will tell you that the inverse transform is a sum of N gamma-tones of all orders from 1 to N with phases rotating in multiples of 90° :

$$h(t) = \sum_{n=1}^N A_n (\omega_r t)^{n-1} \exp(-bt) \cos(\omega_r t - n \frac{\pi}{2})$$

The phases $-n \frac{\pi}{2}$ correspond to cycling through terms $\sin(\omega_r t)$, $-\cos(\omega_r t)$, $-\sin(\omega_r t)$, $\cos(\omega_r t)$, etc., in that order, for any N . The magnitude coefficients A_n of the gamma-tone terms are

not too difficult to derive given compiled relationships about Bessel functions [AS72 10.1.8]. For clarity, we factor A_n into components that each depend on a minimum number of parameters:

$$A_n = \left[1 + b^2/\omega_r^2\right]^N \frac{\omega_r C_{n,N}}{D_N}$$

where $\left[1 + b^2/\omega_r^2\right]^N$ keeps unity gain at DC

ω_r keeps gain constant with center frequency

$$C_{n,N} = \frac{(2N - n - 1)!}{(N - n)!(n - 1)!2^{N-n}}$$
 is an integer numerator for each term

and $D_N = 2^{N-1}(N - 1)!$ is an integer common denominator

The APGF impulse response can be expressed in successively more explicit forms:

$$\begin{aligned} h(t) &= \left[1 + b^2/\omega_r^2\right]^N \frac{\omega_r}{D_N} \sum_{n=1}^N C_{n,N} (\omega_r t)^{n-1} \exp(-bt) \cos(\omega_r t - n \frac{\pi}{2}) \\ &= \left[1 + b^2/\omega_r^2\right]^N \frac{\omega_r}{2^{N-1}(N-1)!} \sum_{n=1}^N \frac{(2N - n - 1)!}{(N - n)!(n - 1)!2^{N-n}} (\omega_r t)^{n-1} \exp(-bt) \cos(\omega_r t - n \frac{\pi}{2}) \end{aligned}$$

For example, values for $C_{n,N}/D_N$, for $N = 5$ are:

$$\frac{C_{1,5}}{D_5} = \frac{105}{384}, \quad \frac{C_{2,5}}{D_5} = \frac{105}{384}, \quad \frac{C_{3,5}}{D_5} = \frac{45}{384}, \quad \frac{C_{4,5}}{D_5} = \frac{10}{384}, \quad \frac{C_{5,5}}{D_5} = \frac{1}{384}$$

Notice that the coefficient of the order-5 gamma-tone term is quite small. This does not mean that the total will not much resemble the order-5 gamma-tone, but rather that the resemblance will grow with time; for $\omega_r t \gg 10$ radians, i.e. after a few cycles of the response, the order-5 term will dominate. Before then, other terms will dominate, making the phase vary from the regular oscillatory pattern of the gamma-tone. The resulting pattern of zero crossings is exactly the set of zeros of the Bessel function or the Spherical Bessel function: initially stretched out a bit, then converging to a constant frequency oscillation. These sequences of zero-crossing times are well-studied and tabulated [AS72 Table 10.6].

Figure A6.1 illustrates the sum of gamma-tones and the Bessel function in relation to the APGF impulse response.

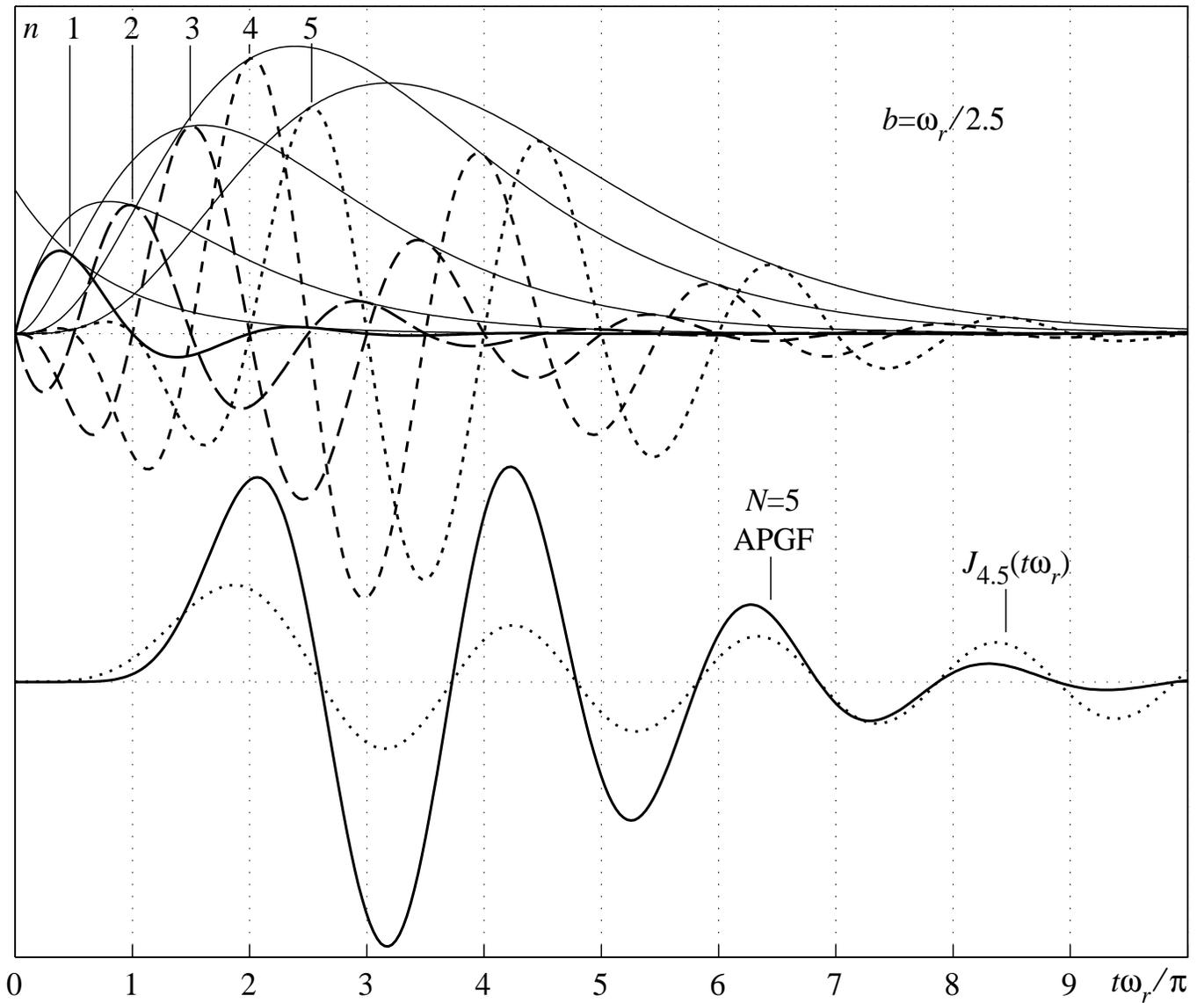


Figure A6.1. The sum of gamma-tones and the Bessel function in relation to the APGF impulse response. The N individual gamma-tones (top) add to make the APGF impulse response (bottom, solid). The gamma-envelopes (top, light curves) are also shown. The resulting impulse can be expressed in terms of a fractional-order gamma-envelope (not shown) times the fractional-order Bessel function (bottom, dotted).

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

Appendix 7—All-Pole Filter Cascade Models

One of our reasons for developing, exploring, and writing about the APGF is to reveal a close relationship between Lyon and Mead’s All-Pole Filter Cascade (APFC) model of cochlear dynamics [LM88, LM89a, LM89b] and the more widely known and accepted gammatone filter.

The APFC is a bank of filters formed by cascading two-pole filter stages, with an output channel or “tap” taken between each pair of stages. The natural frequencies of the stages are varied from high to low, as in the Kim et al. model [KMP73], but using many more stages covering the entire frequency range of hearing. The natural frequencies follow a geometric sequence to map place (stage number) to log frequency (or can follow some other sequence to modify the mapping, but that complicates the analysis).

The APFC can be interpreted as a uni-directional wave propagation model for a system with an underlying spatially-varying complex wave-number, proportional to the log of the complex transfer function of the local stage corresponding to a segment of the system of length Δx :

$$k(\omega, x) = \frac{\log H_1(\omega, x)}{i\Delta x} \quad \text{or} \quad H_1 = \exp(ik\Delta x)$$

$$H^{(APFC)} = \prod H_1(\omega, x) = \exp\left(\sum \log H_1(\omega, x)\right) \approx \exp\left(\int ikdx\right)$$

This relationship corresponds to the WKB approximation to wave propagation in a nonuniform medium, except for a slow amplitude factor that depends on the physics and can’t be reconstructed from the filter [LM89b]. There is no particular reason to believe, however, that the all-pole stage corresponds to the wave-numbers in a real cochlea, or that a pole-zero filter wouldn’t be a better model.

Conceptually, the APFC can start at arbitrarily high frequency (above the range of hearing), such that each output tap can be considered to be approximately the output of an infinite cascade of filters with geometrically varying natural frequency. It is by no means obvious that the infinite product of transfer functions converges to a limit that can be approximated to a reasonable degree by an APGF of finite order—but it does.

Because of the way the APFC has been implemented in sub-threshold analog VLSI using the declining voltage on a resistive line as input to exponential devices to set natural frequencies, we refer to the APFC model as a “tilted line.” “Line” refers both to the resistive line that programs the frequencies, and to the resulting delay line structure. “Tilted” refers to the declining voltage and natural frequency as a function of place. By setting the voltages on opposite ends to the resistive line to be equal, we can convert the circuit to a “flat line,” or cascade of stages with identical natural frequencies; i.e., a tapped succession of APGF’s of increasing order. Feinstein [F96] has explored the relationship between finite-order APGF “flat lines” and finite and infinite APFC “tilted lines.”

Matching an APFC with an APGF

Assuming an alignment of CF’s, an infinite APFC channel and an APGF have only two parameters available in attempting to match their responses: N and Q (where for the APFC we interpret N as N_e , the number of stages per factor of e variation in natural frequency, as opposed to the total number of stages, which is, or can be approximated as, infinite). Fortuitously, setting N and Q equal across the two filter configurations results in a pretty good first-order match between them; small adjustments can further improve the match in gain, delay, and dispersion, but matching delay and dispersion simultaneously turns out to be impossible. Since both systems are minimum phase, it should not be possible to match their impulse responses and still have a delay discrepancy, yet that is what happens, to within a good approximation; the details that differ between them are primarily in the high-side cutoff.

When the peak gains and CF’s are matched, the APFC has slightly wider ERB and 3-dB BW, due to the less steep high-side rolloff.

APFC Peak Gain and CF

The peak gain of the infinite APFC may be found via a closed-form evaluation of an excellent integral approximation to the infinite sum of log gains [F96]. It is most

conveniently expressed in the theta parameterization, where its close approximate relationship to the gain of the corresponding APGF for small θ is evident:

$$H_{\max}^{APFC} = \exp(N_e \theta^2 / 2) = \left[\exp(\theta^2 / 2) \right]^{N_e} \approx [1 / \cos \theta]^{N_e} = H_{\max}^{APGF}$$

$$\text{where } \exp(\theta^2 / 2) \approx 1 + \frac{1}{2} \theta^2 + \frac{3}{24} \theta^4 + \dots$$

$$\text{and } 1 / \cos \theta \approx 1 + \frac{1}{2} \theta^2 + \frac{5}{24} \theta^4 + \dots$$

Each stage and each tap between stages can be characterized by a local natural frequency, center frequency, etc. Since stages and taps alternate, it is most natural to associate with each tap a natural frequency half-way between the natural frequencies of the stages before and after it (i.e., their geometric mean). The CF associated with a tap is defined by applying the formula for the CF of a stage, but using the tap's natural frequency; we say "associated with" because there is no actual filter with such a CF value. Using this convention, the frequency of the peak of the APFC (ω_c^{APFC}) is almost exactly a half-octave higher than the CF associated with the tap (a further correction factor of about $1 - 1/(2N)$ would have been needed if we had instead used the CF of the last stage before the tap):

$$\begin{aligned} \omega_c^{APFC} &= \omega_c^{tap} \sqrt{2} \\ &= \omega_n^{tap} \sqrt{2} \sqrt{1 - 1/(2Q^2)} \\ &= \omega_n^{tap} \sqrt{2 - 1/Q^2} \end{aligned}$$

Notice that the CF of the cascade can be either above the natural frequency of the tap (for $Q > 1$), or below.

APFC Delay and Dispersion

The low-frequency delay is just the summed infinite geometric series of stage delays, totaling about N_e times the delay associated with the tap, or $\sqrt{2}$ times the low-frequency delay of the corresponding CF-matched APGF:

$$\begin{aligned} T_{g(DC)}^{APFC} \omega_n^{tap} &= N_e / Q \\ T_{g(DC)}^{APFC} \omega_c^{tap} &= N_e \sqrt{1 - 1/(2Q^2)} / Q \\ T_{g(DC)}^{APFC} \omega_c^{APFC} &= N_e \sqrt{2 - 1/Q^2} / Q \end{aligned}$$

For the more general case of group delay vs. frequency, we recognize that the phase shift of a stage (or “associated with” a tap) is proportional to the spatial derivative of the phase of the cascade. Since group delay is the frequency derivative of the phase of the cascade, and since frequency and place are closely related in a family of similar functions, we can write the group delay of the cascade in terms of the phase delay at the tap:

$$T_g^{APFC} \omega_n^{tap} = \frac{N_e}{f} \operatorname{arccot} \left(\frac{(1-f^2)Q}{f} \right)$$

$$T_g^{APFC} \omega_c^{APFC} = \frac{N_e \sqrt{2-1/Q^2}}{f} \operatorname{arccot} \left(\frac{(1-f^2)Q}{f} \right)$$

where $f = \omega / \omega_n^{tap}$

The frequency derivative of the arccot factor is $1/Q$ at the origin, so L’Hospital’s rule shows that this formula agrees with the DC result above.

Finding a good expression for the peak delay is hard, but at CF it’s just substitution of ω_c^{APFC} for ω in the above general expression:

$$T_{g(CF)}^{APFC} \omega_c^{APFC} = N_e \operatorname{arccot} \frac{1/Q - Q}{\sqrt{2-1/Q^2}}$$

$$= N_e \operatorname{arccos}(1/Q^2 - 1)$$

$$= N_e \left(\frac{\pi}{2} + \operatorname{arcsin}(1-1/Q^2) \right)$$

$$\approx N_e \left(\frac{\pi}{2} + (1-1/Q^2) \right) \text{ for } Q \text{ near } 1$$

The approximation shows that for Q near 1, the delay at CF is about a quarter cycle per stage (counting N_e stages). This value is a convenient point at which to numerically compare delay and dispersion between APFC and APGF:

$$\begin{aligned}
\text{APFC: } (T_{g(CF)}^{APFC} - T_{g(DC)}^{APFC})\omega_c^{APFC} &= N_e \left(\operatorname{arccot} \frac{1/Q - Q}{\sqrt{2 - 1/Q^2}} - \frac{\sqrt{2 - 1/Q^2}}{Q} \right) \\
&= N_e \left(\frac{\pi}{2} - 1 \right) \approx 0.571N_e \text{ at } Q=1 \\
\text{APGF: } (T_{g(CF)}^{APGF} - T_{g(DC)}^{APGF})\omega_c^{APFC} &= N \left(2Q\sqrt{1 - 1/(2Q^2)} - \sqrt{1 - 1/(2Q^2)}/Q \right) \\
&= N(\sqrt{2} - \sqrt{1/2}) \approx 0.707N \text{ at } Q=1
\end{aligned}$$

Early stages with high natural frequencies contribute delay about equally at DC and at the APFC's CF, so only a few stages near the tap contribute significantly to dispersion. Unlike the delay, which is higher in the APFC than in the APGF by a factor near 1.4, the dispersion is lower, by a factor typically between 0.4 (for Q near 3) and 0.8 (for Q near 1).

Matching the dispersion by increasing N_e and decreasing Q leads to an even worse exaggeration of the overall delay of the APFC relative to the APGF [F96]. Kates [K91] has developed an alternative filter cascade model motivated by trying to avoid this excessive delay feature of the APFC.

Comparative Responses

3 figures below, not yet referenced, compare structures, poles, transfun

Good Features of a Cascade Filterbank

xxx

propagation of distortion (but forward only)

implementation cost

natural coupling of gains

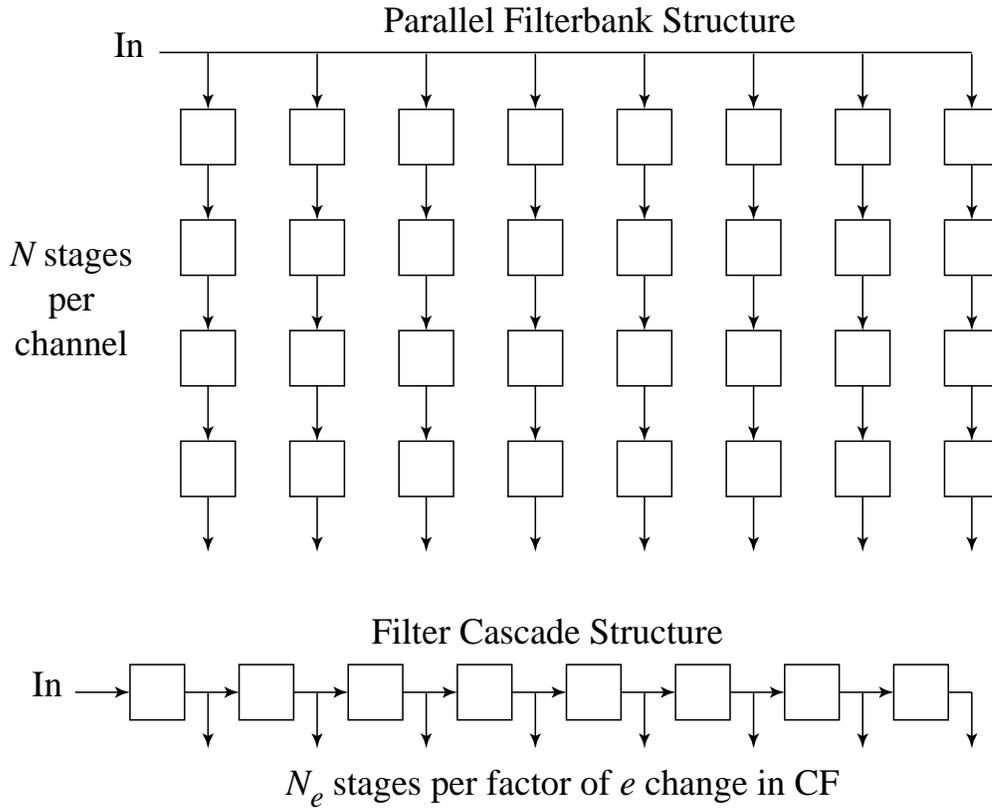


Figure A7.1. Comparison of structures that implement banks of APGF and APFC filters. The interconnected boxes represent second-order sections (SOS's) or two-pole filters. The comparison shows a potential advantage of a factor of N in complexity in favor of the cascade approach, but the higher order required by this approach will mitigate the difference.

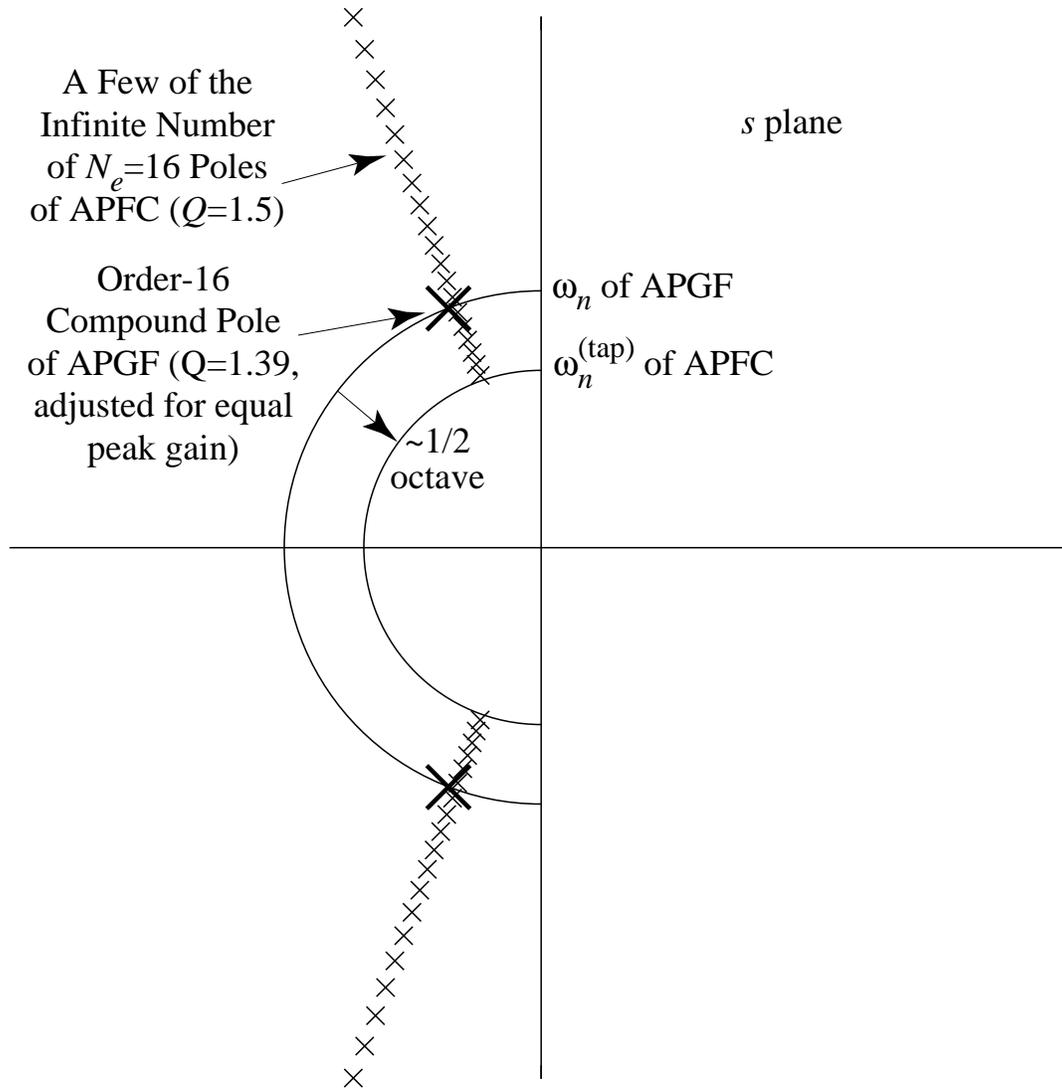


Figure A7.2. An s-plane pole plot for the APGF and APFC filters with equal peak gain, center frequency, and order parameters.

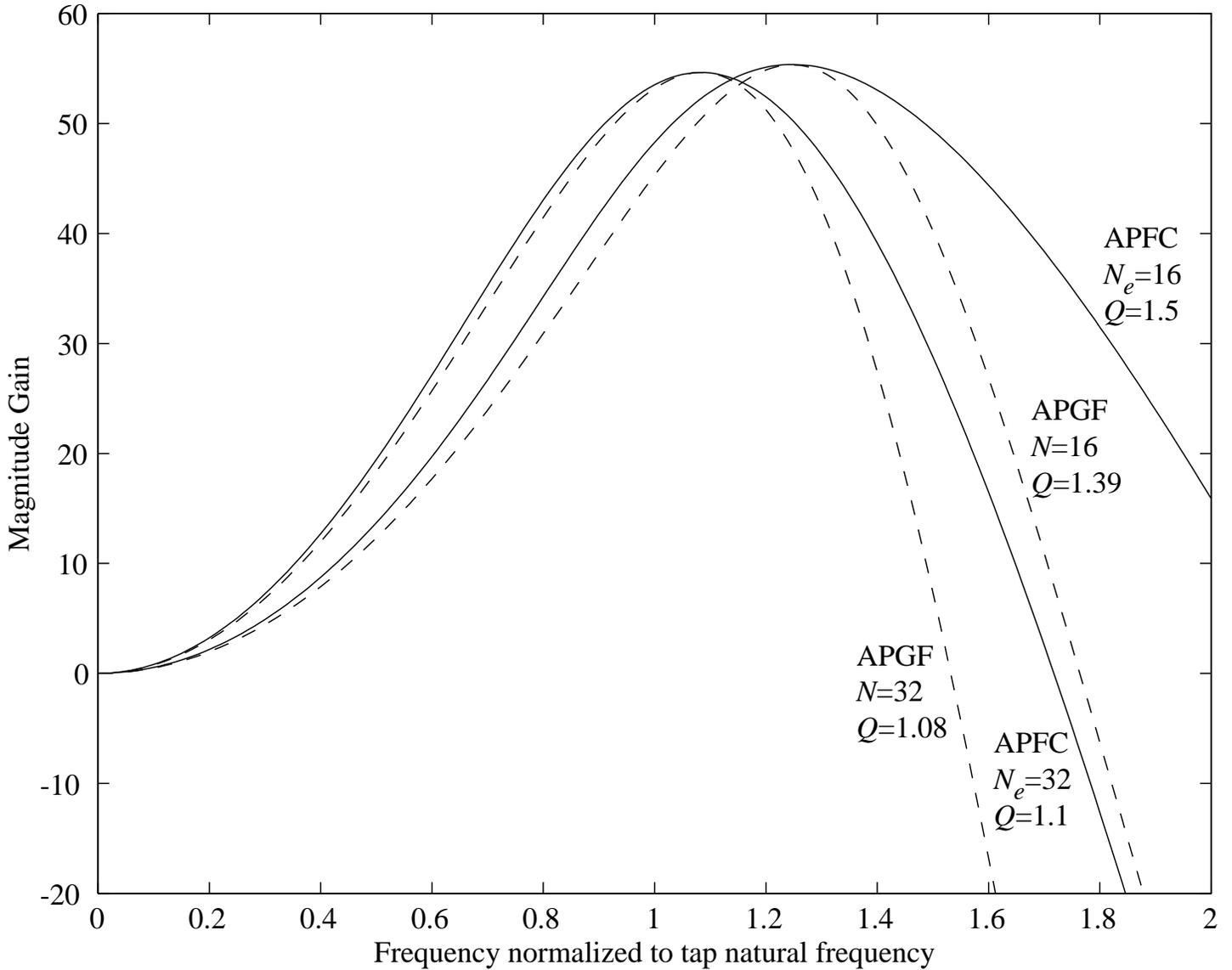


Figure A7.3. Comparison of the shapes of the APGF (dashed) and APFC (solid) filters. The APFC is not as asymmetric as the APGF. For $N_e = 16$ (parameters corresponding to the pole plot of Figure 20), the filter is actually less steep on the high side than on the low side—i.e., the asymmetry is backward.

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
Apple Computer, Inc.
Cupertino, CA 95014 USA

email: lyon@apple.com
fon: (408) 974-4245
fax: (408) 974-8414

Appendix 8—Parametric Nonlinearity and Gain Control

Throughout this paper we have tried to tie the APGF and related auditory filters to what is known about the mechanical filtering in the cochlea, including nonlinear effects such as the variation of gain with level. The notion of automatic gain control (AGC), in which the output power of a system is fed back to control the gain of the system, provides a powerful framework for analyzing such a compressive nonlinearity.

AGC is an important but often neglected aspect of cochlea models [L90]. The effect of such a nonlinearity in making iso-intensity curves look quite different from iso-response curves (tuning curves) was first pointed out by Zwislocki [Z68]. Observations of this phenomenon, such as de Boer's [B69] that revcor-derived filters appear to be less sharp than tuning curves, continued to puzzle the field for many years, despite Zwislocki's explanations. The effect has been best documented and illustrated by Ruggero [R92] at the level of basilar membrane motion.

We have discussed the notion of using the peak gain as an independent parameter to characterize auditory filters, and letting the peak gain, or another parameter that controls it, depend on level. There is an open issue in the literature [RB94, MG87] as to whether auditory filter shape (and here we would include gain) depends on the power at the input of the filter or at the output, or more generally on the input power at frequencies near the passband of the filter. It would be easier to get a handle on this issue if models were built, using an AGC framework, in which the low-frequency and center-frequency gains were constrained to roughly match mechanical data, under control of feedback from a level-sensing mechanism at the input or the output or elsewhere. This approach mixes together the psychoacoustic and mechanical aspects of filtering, and relies on an assumption that they are very closely related, if not completely equivalent, in an attempt to get around the limitations of each type of experimental data.

In an auditory filterbank with AGC, it is not necessary that the parameter controlling the gain is either the input power or the output power for each filter. Rather, the “neighborhood” concept is easily implemented, such that energy near the passband of a filter can reduce the gain of that filter, even though it is somewhat outside the passband, through a mechanism resembling lateral inhibition. The simplest way that this inhibition might arise is through the coupling of parameters in a filter cascade; i. e., reducing the gain of any stage by feeding back from a detector at its output tap will reduce the gain to later taps as well, with a diminishing effect over distance. This and other techniques for lateral spread of gain reduction need to be parameterized and fitted to the experimental data, to help produce a better answer to the input vs. output argument.

As a start on designing an auditory filterbank with AGC, consider an APFC or APGF-bank stimulated by a constant level of noise per bandwidth, or per bark, roughly. Then each stage, or each output tap, should have about the same gain, which should be a function of the noise level. This scheme avoids having to decide on a spreading function in space or in time, and allows a global level-dependent Q —just for this analysis, since spatially varying gain is important in a real cochlea.

The best data on compressive input/output power relations are measured with tones, however, so it’s hard to do the analysis we want. So let’s pretend the tone data near CF resembles what happens for noise, and ignore CF shift with level for now. Measurements of cochlear mechanics show from 2:1 to 5:1 compression (stapes dB change : basilar membrane dB change) over most of the range of normal hearing levels [e.g., R71, RRR86, RRR92, ND96].

Figure A8.1 shows an example of how feedback of the output power of a filter to its θ parameter might control the gain vs. level nonlinearity. For this plot, we first specify a linear feedback relationship between output r.m.s. amplitude (square root of smoothed power measurement) and the filter’s θ parameter, using a maximum θ at zero output corresponding to a 60 dB maximum peak gain. For each output level we compute θ and H_{\max} , then infer the corresponding input, assuming a sinusoid at CF. Linear 4:1 compression over a wide range of input levels is not designed into this process, but is an approximate result of the very expansive relationship between θ and H_{\max} . Both the APGF and the APFC show about the same result.

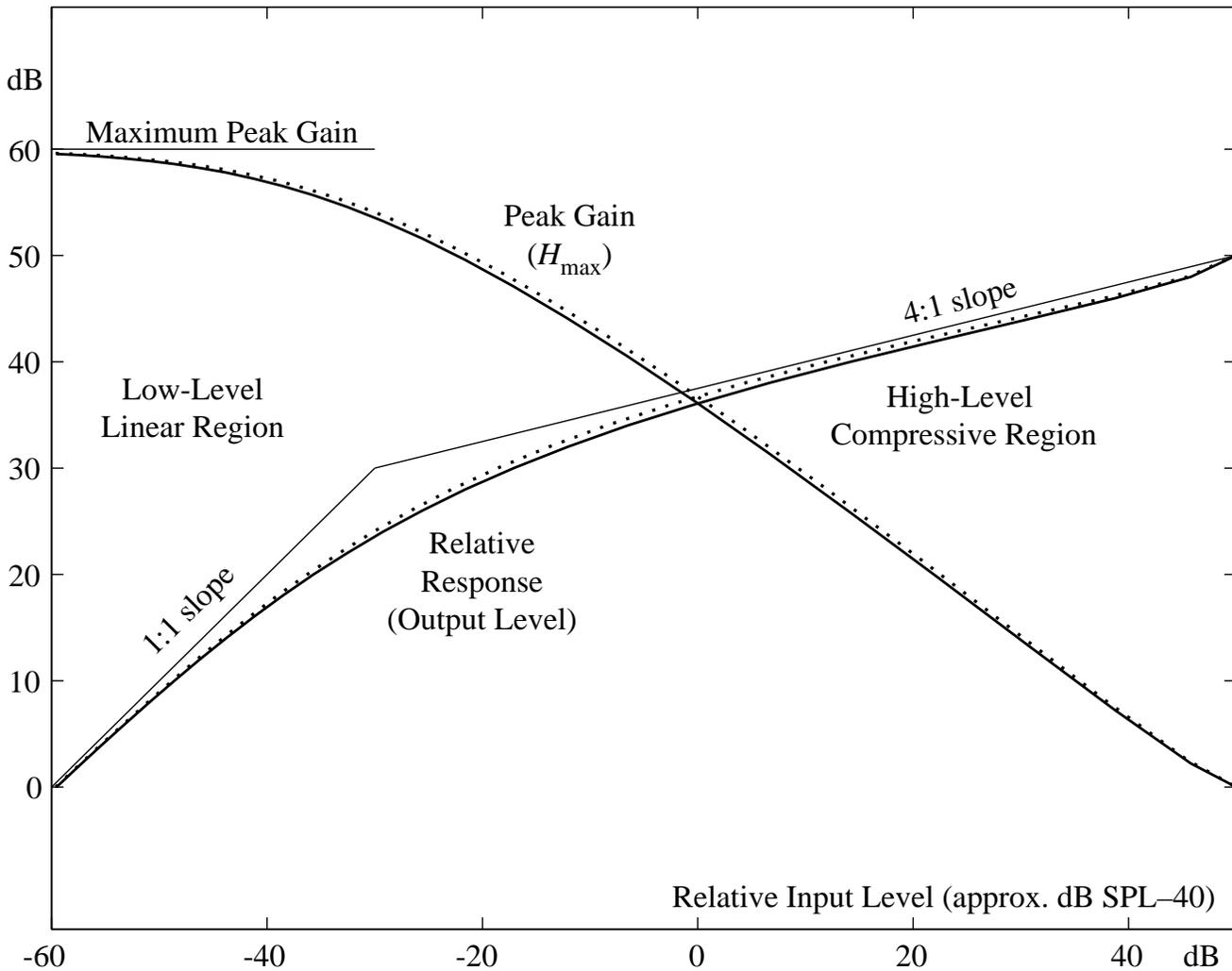


Figure A8.1. Level-dependent filter gains (downward sloping curves) and input-output curves (upward sloping curves) using linear feedback of output amplitude to reduce θ from its maximum value. The low-level limit is established by a maximum θ value. The high-level limit is a peak gain of 0 dB. At low levels, the system behaves linearly (below -50 dB relative input level), while over a wide range, the system behaves almost as a 4:1 compressor, as shown by the slope of the reference line at high levels. Solid curves are for APGF and dotted curves for APFC, each with its own maximum θ value corresponding to a peak gain of 60 dB.

— DRAFT of 5 Feb. 96 — do not distribute

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon

Apple Computer, Inc.

Cupertino, CA 95014 USA

email: lyon@apple.com

fon: (408) 974-4245

fax: (408) 974-8414

Appendix 9—Comparing GTF and RoEx Filter Shapes

2 figures below, not yet referenced, compare filter shapes

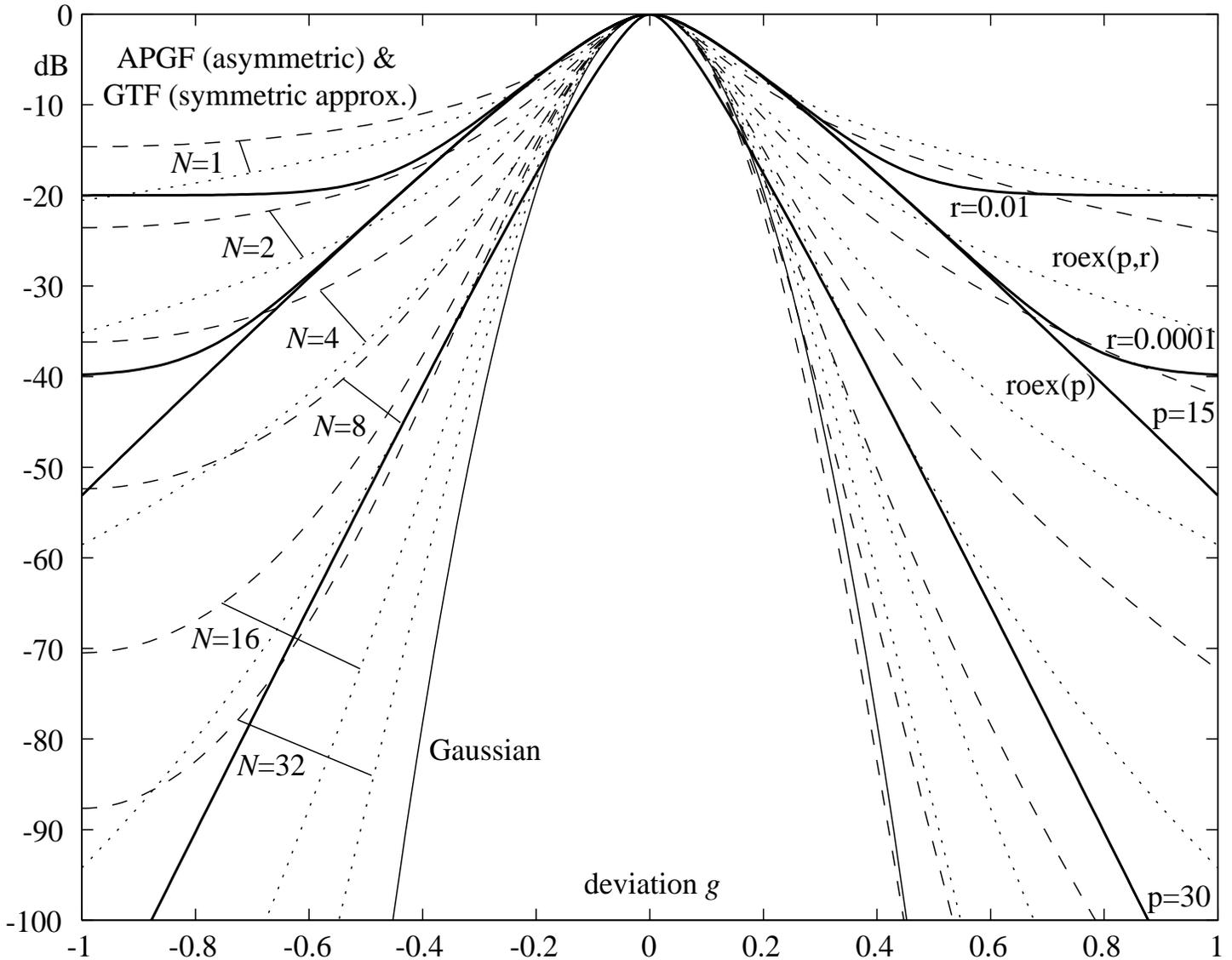


Figure A9.1. A comparison of the shapes of different functional forms used for describing auditory filters: GTF (dotted), APGF (dashed), Gaussian (light solid) and rounded exponential (roex) filters (heavy solid). With one exception, the various forms are matched in terms of low-order Taylor expansions about the peak; i.e., unity gain, zero slope, and equal curvatures at zero deviation from the center frequency. For the roex family, a p value of only 15 is used; in fitting auditory filters, a higher value (20–30) is typically found, bringing the skirts more into line with the other filter forms, but resulting in a fairly sharp curvature near the peak. Except for the narrow peak, the $p=30$ roex curve comes close to matching the $N=8$ GTF curve over a wide dynamic range. The Gaussian filter generally has the steepest skirts. For very large deviations, the roex(p) filter is steeper than any of the gammatone filter family, including the high- N APGF that is steeper than the Gaussian for small deviations. Fitting data on the skirts and tails with the roex instead of a GTF or APGF probably

results in underestimation of the auditory filter's 3-dB bandwidth and equivalent rectangular bandwidth.

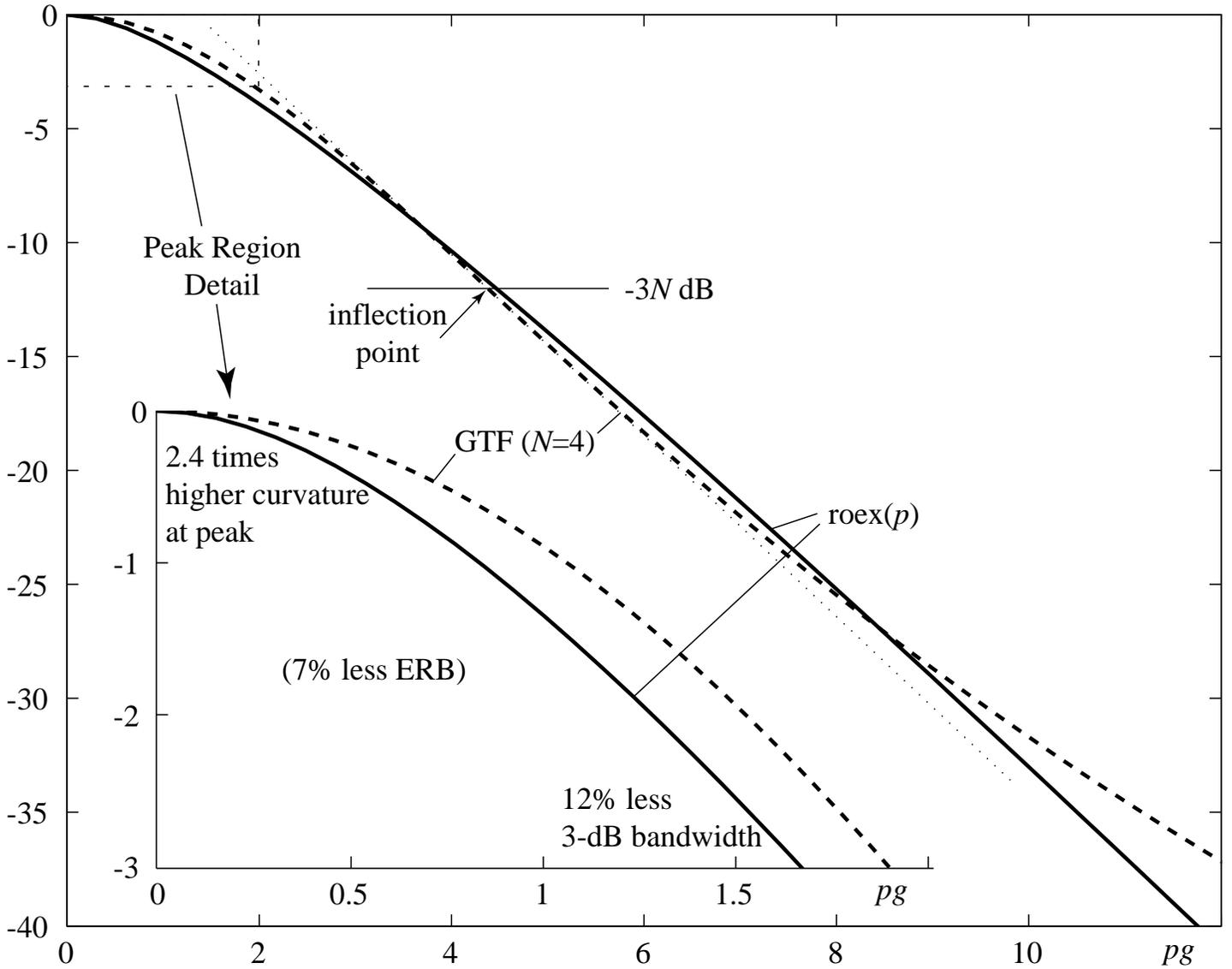


Figure A9.2. Detailed comparison of the fit between the $roex(p)$ and the order-4 GTF, showing the narrower peak of the $roex(p)$ when the skirts are matched. The deviation is normalized by p to emphasize the fact that the shape is independent of the bandwidth; in the case of the GTF, the symmetric approximation is used, which holds best for large p (small $\alpha = b/\omega_r$, normalized bandwidth parameter). Curvatures at the peak would be matched at $p^2/2$ for $p = \sqrt{2N}/\alpha$; the fit shown is obtained when $p = 1.55\sqrt{2N}/\alpha$. The dotted line is the tangent at the inflection point of the GTF, which occurs at $-3N$ dB for a GTF of any order. At higher orders, the peak sharpness and bandwidth discrepancies get worse. For $N=2$ the fit is very good with essentially

no ERB discrepancy, but the inflection point is at only -6 dB, so the fit does not extend over a wide dynamic range.

— DRAFT of 29 Jan. 96 — do not distribute

The All-Pole Gammatone Filter and Auditory Models

Richard F. Lyon
 Apple Computer, Inc.
 Cupertino, CA 95014 USA

email: lyon@apple.com
 fon: (408) 974-4245
 fax: (408) 974-8414

References

(by first letter and year, still under construction)

[A78] Altes, R. A. (1978). “The Fourier-Mellin transform and mammalian hearing,” *J. Acoust. Soc. Am.* 63, 174–183.

[AJ80] Aertsen, A., and Johannesma, P. (1980). “Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli.” *Biol. Cybern.* 38, 223–234.

[AS72] Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions*, Dover Publications, New York.

[AS89] Assman, P. F. & Summerfield, A. Q. (1989) “Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency,” *J. Acoust. Soc. Am.* 85, 327–338.

[B69] Boer, E. de (1969) “Encoding of Frequency Information in the Discharge Pattern of Auditory Nerve Fibers,” *Intern. Audiol.* 8, 547–556.

[B73] Boer, E. de (1973) “On the principle of specific coding,” *J. Dyn. Syst. Meas. Control* 95G, 265–273. xxx check this ref where revcor is as sharp as FTC

[B75] Boer, E. de (1975) “Synthetic whole-nerve action potentials for the cat,” *J. Acoust. Soc. Am.* 58, 1030–1045.

[BJ78] Boer, E. de and Jongh, H. R. de (1978) “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” *J. Acoust. Soc. Am.* 63, 115–135.

- [BK90] Boer, E. de and Kruidenier, C. (1990) "On ringing limits of the auditory periphery," *Biol. Cybernetics* 63, 433–442.
- [BC95] Brown, G. J., and Cooke, M. P. (1995) "Temporal synchronisation in a neural oscillator model of primitive auditory stream segregation," *IJCAI Workshop on Computational Auditory Scene Analysis*, Montreal, Aug. 1995.
- [BN96] Boer, E. de, and Nuttall, A. L. (1996) confidential draft xxx
- [C91] Cooke, M. P. (1991) "Modelling Auditory Processing and Organisation," Ph.D. thesis, CS Dept., Univ. of Sheffield (also Cambridge University Press, 1993).
- [C93] Carney, L. H. (1993) "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* 93, 401–417.
- [C96] Culling, J. (1996) "Digital signal processing software for teaching and research in psychoacoustics under UNIX and X-windows" *Behavioural Research Methods Instruments and Computers* (in press) (URL <ftp://ftp.ihr.mrc.ac.uk/pub/johncu/wave.tar.Z>).
- [CD96] Chau, W., and Duda, R. O. (1996) "Combined Monaural and Binaural Localization of Sound Sources," *29th Asilomar Conference on Signals, Systems, and Computers*, (in press), IEEE Computer Society Press, Los Alamitos, CA.
- [CY88] Carney, L. H., and Yin, T. C. T. (1988). "Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model," *J. Neurophysiology* 60, 1653–1677.
- [D91] Darling, A. M. (1991). "Properties and Implementation of the GammaTone Filter: A Tutorial," in *Speech Hearing and Language (UCL Work in Progress)*, 5, 43–61, University College London, Department of Phonetics and Linguistics.
- [DP91] de Vries, B., and Principe, J. C., (1991) "A Theory for Neural Networks with Time Delays," in *Neural Information Processing Systems (NIPS-90)*, R. Lippmann, J. Moody, and D. Touretsky (eds.), 162–168, San Mateo: Morgan Kaufmann.
- [DP92] de Vries, B., and Principe, J. C., (1992) "The Gamma Model—A New Neural Model for Temporal Processing," *Neural Networks* 5, pp. 565–576.

- [DW94] Dijk, P. van, and Wit, H. P. (1994) “Speculations on the relation between emission generation and hearing mechanisms in frogs,” in *Advances in Hearing Research*, G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (eds.), pp. 105–115.
- [E89] Evans, E. F. (1989) “Cochlear Filtering: A View Seen through the Temporal Discharge Patterns of Single Cochlear Nerve Fibres,” in *Cochlear Mechanisms—Structure, Function, and Models*, J. P. Wilson and D. T. Kemp (eds.), 241–250, Plenum Press, New York.
- [F60] Flanagan, J. L. (1960). “Models for approximating basilar membrane displacement,” *Bell Sys. Tech. J.* 39, 1163–1191.
- [F62a] Flanagan, J. L. (1962). “Models for approximating basilar membrane displacement—Part II. Effects of middle-ear transmission and some relations between subjective and physiological behavior”. *Bell Sys. Tech. J.* 41, 959–1009.
- [F62b] Flanagan, J. L. (1962) “Computer Simulation of Basilar Membrane Motion,” Fourth Intl. Congress on Acoustics, Copenhagen, Aug. 1962, paper H26.
- [F65] Flanagan, J. L. (1965) *Speech Analysis, Synthesis, and Perception*, pp. 93–94 & p. 126, Springer, Berlin (or 2nd edition, 1972, pp. 110–112 & p. 149).
- [F95] Feinstein, D. (1996). “xxx” Caltech CNS memo (still in preparation).
- [G74] Grashuis, J. L. (1974) “The pre-event stimulus ensemble: an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons,” Ph.D. thesis, Laboratory of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands.
- [GM90] Glasberg, B. R., and Moore, B. C. J. (1990) “Derivation of auditory filter shapes from notched-noise data,” *Hearing Research* 47, 103–138.
- [GMPN84] Glasberg, B. R., Moore, B. C. J., Patterson, R. D., and Nimmo-Smith, I. (1984) “Dynamic range and asymmetry of the auditory filter,” *J. Acoust. Soc. Am.* 76, 419–427.
- [H80] Hall, J. L. (1980) “Cochlear models: Evidence in support of mechanical nonlinearity and a second filter (A review),” *Hearing Research* 2, 455–464.
- [HC95] Hogg, R. V. and Allen, A. T. (1995) *Introduction to Mathematical Statistics*, fifth edition, Prentice Hall, Englewood Cliffs NJ.

[H88] Holdsworth, J., Nimmo-Smith, I., Patterson, R. D., and Rice, P. (1988) “Implementing a Gamma Tone Filter Bank,” Annex C in “SVOS Final Report—Part A: The Auditory Filterbank,” MRC Applied Psychology Unit, Cambridge, England.

[I95a] Irino, T. (1995) “An Asymmetric Extension of the Gammatone Filter Function,” abstract, Short Papers Meeting on Experimental Studies on Hearing and Deafness, British Society of Audiology, Univ. Lab. or Psych., Oxford, Sept. 1995.

[I95b] Irino, T. (1995) “An Optimal Auditory Filter,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, Oct. 1995.

[J72] Johannesma, P. I. M. (1972). “The pre-response stimulus ensemble of neurons in the cochlear nucleus”. Proc. of the Symposium on Hearing Theory, IPO, Eindhoven, The Netherlands.

[K84] Khanna, S. M. (1984) “Inner Ear Function Based on the Mechanical Tuning of the Hair Cells,” in Hearing Science, C. Berlin (ed.), pp. 213–240, College-Hill Press, San Diego.

[K91] Kates, J. M. (1991) “A Time-Domain Digital Cochlear Model,” IEEE Trans. Signal Processing 39, 2573–2592.

[K61] Kelly, J. L., Vyssotsky, V. A., and Lochbaum, C. (1961) “A block diagram compiler,” Bell Sys. Tech. J. 40, 669–676.

[KMP73] Kim, D. O., Molnar, C. E., Pfeiffer, R. R. (1973) “A system of nonlinear differential equations modeling basilar-membrane motion,” J. Acoust. Soc. Am. 54, 1517–1529.

[L90] Lyon, R. F. (1990) “Automatic Gain Control in Cochlear Mechanics,” The Mechanics and Biophysics of Hearing, P. Dallos et al. (eds.), 395–401, Springer-Verlag.

[LM88] Lyon, R. F. and Mead, C. A. (1988) “An Analog Electronic Cochlea,” IEEE Trans. ASSP. 36, 1119–1134.

[LM89a] Lyon, R. F. and Mead, C. A. (1989) “Electronic Cochlea,” Ch. 16 in: Analog VLSI and Neural Systems, C. Mead, Addison-Wesley.

[LM89b] Lyon, R. F. and Mead, C. A. (1989) "Cochlear Hydrodynamics Demystified," Caltech Comp. Sci. Dept. report Caltech-CS-TR-88-4.

[MN79] Møller, A. R. and Nilsson, H. G. (1979) "Inner Ear Impulse Response and Basilar Membrane Modelling," *Acustica* 41, 258–262.

[MG87] Moore, B. C. J. and Glasberg, B. R. (1987) "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hearing Research* 28, 209–225.

[MH91] Meddis, R. and Hewitt, M. J. (1991) "Virtual pitch and phase-sensitivity studied using a computer model of the auditory periphery: I pitch identification," *J. Acoust. Soc. Am.* 89, 2866–2882.

[MH92] Meddis, R. and Hewitt, M. J. (1992) "Modelling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* 91, 233–245.

[MHS90] Meddis, R., Hewitt, M. J., and Shackleton, T. M. (1990) "Non-linearity in a computational model of the response of the basilar membrane," in *The Mechanics and Biophysics of Hearing*, P. Dallos et al. (eds.), 403–410, Springer-Verlag.

[MHS92] Meddis, R., Hewitt, M. J., and Shackleton, T. (1992) "An Anatomical/Physiological Approach to Auditory Selective Attention," in *Auditory Physiology and Perception*, Y. Cazals, L. Demany, and K. Horner (eds.), 495–504, Pergamon Press, Oxford.

[ND96] Nuttall, A. L., and Dolan, D. F. (1996) "Steady-state sinusoidal responses of the basilar membrane in guinea pig," *J. Acoust. Soc. Am.* (in press).

[O94] O'Mard, L. P. (1995) "LUTEar Core Routines Library 1.9.0," (URL <http://info.lut.ac.uk/departments/hu/groups/speechlab/speechlabhome.html>).

[OS92] Oliveira e Silva, T., Guedes de Oliveira, P., Principe, J. C., and De Vries, B. (1992) "Generalized feedforward filters with complex poles," *Neural Networks for Signal Processing II. Proceedings of the IEEE-SP Workshop*, pp. 503–510, IEEE Press.

[P74] Patterson, R. D. (1974) "Auditory filter shape," *J. Acoust. Soc. Am.* 55, 802–809.

- [P76] Patterson, R. D. (1976) "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* 59, 640–654.
- [PM86] Patterson, R. D., and Moore, B. C. J. (1986) "Auditory Filters and Excitation Patterns as Representations of Frequency Resolution," in *Frequency Selectivity in Hearing*, B. C. J. Moore (ed.), 123–177, Academic Press.
- [PN80] Patterson, R. D., and Nimmo-Smith, I. (1980) "Off-frequency listening and auditory filter asymmetry," *J. Acoust. Soc. Am.* 67, 229–245.
- [PNWM82] Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982) "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* 67, 1788–1803.
- [Pi88] Pickles, J. O. (1988) *An Introduction to the Physiology of Hearing*, 2nd edition (see p. 81 & p. 97), Academic Press, London.
- [Pr88] Prijs, V. F. (1988) "Dynamic tuning properties in the guinea pig," in *Basis Issues in Hearing*, H. Duifhuis, J. W. Horst, and H. P. Wit (eds.), pp. 196–203, Academic Press, London.
- [P88] Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988) *Spiral VOS Final Report—Part A: The Auditory Filterbank*, MRC Applied Psychology Unit, Cambridge, England.
- [P91] Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill, New York.
- [P94] Patterson, R. D. (1994). "The sound of a sinusoid: Spectral models". *J. Acoust. Soc. Am.* 96, 1409–1418.
- [PH89] Patterson, R. D. & Hirahara, T. (1989). "HMM speech recognition using DFT and auditory spectrograms," ATR HIP Technical Report 10.23.89.
- [PAH92] Patterson, R. D., Allerhand, M. H. and Holdsworth, J. (1992) "Auditory representations of speech sounds," in *Visual representations of speech signals*, Eds. Martin Cooke, Steve Beet, and Malcolm Crawford, John Wiley & Sons, Chichester. 307–314.

- [PHA92] Patterson, R. D., Holdsworth, J. and Allerhand M. (1992). "Auditory Models as preprocessors for speech recognition," in *The Auditory Processing of Speech: From the auditory periphery to words*, M. E. H. Schouten (ed.), 67–83, Mouton de Gruyter, Berlin.
- [PAA94] Patterson, R. D., Anderson, T., and Allerhand, M. (1994). "The auditory image model as a preprocessor for spoken language," in *Proc. Third ICSLP, Yokohama, Japan*, 1395–1398.
- [P92] Patterson, R. D., Robinson, K., Holdsworth, J. W., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, Y. Cazals, L. Demany, and K. Horner (eds.), 429–446, Pergamon Press, Oxford.
- [PAG95] Patterson, R. D., Allerhand, M., and Giguere, C. (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* 98, pp. 1890–1894 (URL <http://www.mrc-apu.cam.ac.uk/aim/>).
- [PWJJ94] Palmer, A. R., Winter, I. M., Jiang, D., and James, N. (1994) "Across-frequency integration by neurones in the ventral cochlear nucleus," in *Advances in Hearing Research*, G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (eds.), pp. 250–263.
- [PDO93] Principe, J. C., de Vries, B., and Oliveira, P. G. de, (1993) "The Gamma Filter—A New Class of Adaptive IIR Filters with Restricted Feedback," *IEEE Trans. Sig. Proc.* 41, 649–656.
- [R71] Rhode, W. S. (1971) "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* 49, 1218–1231.
- [RS85] Rhode, W. S. and Smith, P. H. (1985) "Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers," *Hearing Research* 18, 159–168.
- [RHAB71], Rose, J. E., Hind, J. E., Anderson, D. J., and Brugge, J. F. (1971) "Some Effects of Stimulus Intensity on Response of Auditory Nerve Fibers in the Squirrel Monkey," *J. Neurophysiology* 34, 685–699.
- [R90] Robinson, A., Holdsworth, J., Patterson, R. D., and Fallside, F. (1990). "A Comparison of Preprocessors for the Cambridge Recurrent Error Propagation Network Speech Recognition System," *ICSLP, Kobe, Japan*, 1033–1036.

- [RRR86] Robles, L., Ruggero, M. A., and Rich, N. C. (1986) “Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves and response phases,” *J. Acoust. Soc. Am.* 80, 1364–1374.
- [RRR92] Ruggero, M. A., Rich, N. C., and Recio, A. (1992) “Basilar membrane responses to clicks,” in *Auditory Physiology and Perception*, Y. Cazals, L. Demany, and K. Horner (eds.), 85–91, Pergamon Press, Oxford.
- [R92] Ruggero, M. A. (1992) “Responses to sound of the basilar membrane of the mammalian cochlea,” *Current Opinion in Neurobiology* 2, pp. 449–456.
- [RB94] Rosen, S., and Baker, R. J. (1994) “Characterising auditory filter nonlinearity,” *Hearing Research* 73, 231–243.
- [S85] Schofield, D. (1985). “Visualisations of speech based on a model of the peripheral auditory system”. National Physical Laboratory Report DITC 62, Middlesex, UK.
- [SG89] Stokkum, I. H. M. van, and Gielen, C. C. A. M. (1989), “A model for the peripheral auditory nervous system of the grassfrog,” *Hearing Research* 41, pp. 71–85.
- [S93] Slaney, M. (1993). “An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank,” Apple Technical Report #35 (URL <ftp://ftp.apple.com/pub/malcolm/Gammatone.math>).
- [S94] Slaney, M. (1994). “Auditory Model Toolbox,” Apple Technical Report #45. (URL <ftp://ftp.apple.com/pub/malcolm/AuditoryToolbox.sea.hqx>).
- [SMH92] Shackleton, T. M., Meddis, R. and Hewitt, M. J. (1992) “Across frequency integration in a model of lateralisation,” *J. Acoust. Soc. Am.* 91, 2276–2279.
- [SM96] Sarpeshkar, R., Mahajan, S., and Lyon, R. F. (1996) tech report in progress
- [SWK96] Solbach, L., Wöhrmann, R. and Kliewer, J. (1996) “The complex-valued Continuous Wavelet transform as a preprocessor for auditory scene analysis,” Working Notes of the Workshop on Computational Auditory Scene Analysis at the International Joint Conference on Artificial Intelligence, Montreal, Canada (in press—preprint available via <http://goofy.ti6.tu-harburg.de/research/>).

[W91] Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer*, 2nd ed., Addison-Wesley.

[W92] Watts L., Kerns, D., Lyon, R., and Mead, C. (1992) "Improved Implementation of the Silicon Cochlea," *IEEE J. Solid State Circuits* 27, 692–700.

[WBW88] Wilson, J. P., Baker, R. J., and Whitehead, M. L. (1988) "Level dependence of frequency tuning in human ears," in *Basis Issues in Hearing*, H. Duifhuis, J. W. Horst, and H. P. Wit (eds.), pp. 80–87, Academic Press, London.

[WS96] Wöhrmann, R. and Solbach, L. (1996) "The AnnaLisa Software," Technische Universität Hamburg-Harburg (URL <http://www.ti6.tu-harburg.de/~rolf/AnnaLisa.html>).

[Y94] Yost, W. A. (1994) *Fundamentals of Hearing*, 3rd edition (see p. 119), Academic Press, San Diego.

[YCG83] Young, E. D., Costalupes, J. A., and Gibson, D. J. (1983) "Representation of Acoustic Stimuli in the Presence of Background Sounds: Adaptation in the Auditory Nerve and Cochlear Nucleus," (see General Discussion question by Harrison) in *Hearing—Physiological Bases and Psychophysics*, R. Klinke and R. Hartmann (eds.), 119–127 Springer-Verlag, Berlin.

[Z68] Zwislocki, J. J. (1968) "Frequency Analysis in Hearing," Special Report LSC-S-4 (presented at the joint 1968 meeting of the Committee on Vision and Hearing, Bioacoustics and Biomechanics of the National Research Council), Laboratory of Sensory Communication, Syracuse University, Syracuse, New York (see also Zwislocki's written comments at the end of [B69])