

Cascades of two-pole–two-zero asymmetric resonators are good models of peripheral auditory function

Richard F. Lyon^{a)}

Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043

(Received 28 February 2011; revised 10 October 2011; accepted 11 October 2011)

A cascade of two-pole–two-zero filter stages is a good model of the auditory periphery in two distinct ways. First, in the form of the *pole–zero filter cascade*, it acts as an auditory filter model that provides an excellent fit to data on human detection of tones in masking noise, with fewer fitting parameters than previously reported filter models such as the roex and gammachirp models. Second, when extended to the form of the *cascade of asymmetric resonators with fast-acting compression*, it serves as an efficient front-end filterbank for machine-hearing applications, including dynamic nonlinear effects such as fast wide-dynamic-range compression. In their underlying linear approximations, these filters are described by their poles and zeros, that is, by rational transfer functions, which makes them simple to implement in analog or digital domains. Other advantages in these models derive from the close connection of the filter-cascade architecture to wave propagation in the cochlea. These models also reflect the automatic-gain-control function of the auditory system and can maintain approximately constant impulse-response zero-crossing times as the level-dependent parameters change. © 2011 Acoustical Society of America.

[DOI: 10.1121/1.3658470]

PACS number(s): 43.66.Ba, 43.64.Bt, 43.66.Dc [CJP]

Pages: 3893–3904

I. INTRODUCTION

Over the last half century, many auditory filter models have been developed, analyzed, and applied to a variety of hearing-related problems. Linear filter models, as well as more realistic quasi-linear level-dependent models have been explored. Several lines of development, and several criteria that filter models might try to satisfy, have been reviewed with respect to their connections and applicability to psychoacoustic data, to physiological data, and to machine-hearing systems; the *pole–zero filter cascade* (PZFC) model structure achieves the specified properties better than other models do (Lyon *et al.*, 2010a).

Quasi-linear (level dependent) auditory filter models can be seen as belonging to three main families of filters: the rounded exponential (roex), the gammatone/gammachirp, and the filter cascade. In many cases, independent efforts led to somewhat similar results, without necessarily sharing a name or any other relationship; some of these have been discovered in retrospect, such as the early 1960s work by Jim Flanagan on gammatone, one-zero gammatone, and related pole–zero filter models of basilar membrane motion (Flanagan, 1960, 1962), long before the term gammatone was coined.

Transmission-line models of wave propagation on the basilar membrane go even further back, but the basis for approximating these systems as filter cascades was not made clear until Zweig *et al.* (1976) showed how to apply the Wentzel–Kramers–Brillouin (WKB) approximation in their 1976 “Cochlear Compromise” paper. They connected a 1D model of cochlear physics to a circuit model similar to the old transmission-line models of Wegel and Lane (1924),

Peterson and Bogert (1950), and Ranke (1950), but the method that they explained led via the WKB method to a wider class of filter-cascade models of the cochlea, “cascade filterbanks,” as opposed to conventional parallel filterbanks (Lyon, 1982, 1998). The reported approach is based on such cascades that relate to the wave mechanics but draws also on the gammatone line of development.

Models that incorporate nonlinearities, such as bandpass nonlinear (BPNL) and dual-resonance nonlinear (DRNL) models, are typically based on the gammatone or similar quasi-linear models. Nonlinear extensions can be arbitrarily complicated, but are often restricted to instantaneous nonlinearities in the signal path plus sometimes one or more level-dependent parameters. The cascade structure allows a straightforward way to incorporate both of these types of nonlinearity. The *cascade of asymmetric resonators with fast-acting compression* (CAR-FAC) extends the PZFC model with compressive cubic nonlinearities between resonator stages, as in BPNL models, plus an automatic gain control (AGC) feedback system to incorporate dynamic level dependence.

II. AUDITORY FILTER MODELS

The auditory filters considered here include both those motivated by psychoacoustic experiments, such as detection of tones in noise maskers, and those motivated by reproducing the observed mechanical response of the basilar membrane or neural response of the auditory nerve. These are not necessarily going to lead to the same models, but it is one thesis of this work that a single model can do a good job for both of these, and thereby provide a good basis for machine-hearing systems. Since there are several stages of neural processing between the cochlea and psychoacoustic perceptions, it would

^{a)}Author to whom correspondence should be addressed. Electronic mail: dicklyon@google.com

not be surprising if the best parameters were different between these types of models, but it seems likely that the linear and nonlinear filtering due to the cochlea plays a sufficient role in perception that one set of parameters may be adequate, at least for a range of machine-hearing applications.

Duifhuis (2004) recounts the history of cochlear models and divides them into two classes: (1) the transmission-line class and (2) the filterbank class. More specifically, he says, “The major difference is that models in class 1 take physical coupling between system elements into account, whereas in class 2 the channels are independent, and coupling is completely determined by the common input.” Filter cascades provide a natural model of coupling in the forward direction, and an AGC feedback network can model some coupling between channels in both directions, so these cascades can be viewed as a bridge between Duifhuis’s two classes: they do not support backward traveling waves as transmission lines do, but they do model the forward wave to efficiently implement filterbanks. The filter cascade is the strategy employed here for abstracting the transmission-line models into efficiently runnable filter models.

Auditory filters have traditionally been described by the power–frequency response (roex family) or by the impulse response (gammatone/gammachirp family). In electrical engineering, descriptions in terms of Laplace-domain poles and zeros is a more traditional approach to filter description and specification, with advantages in terms of analysis and implementation. Some filters, such as the cascade structures investigated in this work, do not have simple descriptions in terms of impulse responses or frequency responses but do have simple and natural descriptions in terms of poles and zeros (Lyon *et al.*, 2010a).

Several lines of auditory filter models, particularly those roex-family and gammachirp-family filters that have been fitted to human masking data, have been reviewed and assessed relative to models based on filter cascades by Lyon *et al.* (2010a).

A. Time-varying and nonlinear auditory filters

Although nonlinearities manifest themselves in various ways in hearing, there is still good value in quasi-linear models, that is, those models that can be described as linear filters but with parameters that depend on signal level. Such models will not reproduce effects such as distortion products and suppression but can still capture major masking effects and the large input–output compression associated with cochlear mechanisms and loudness perception (Bacon, 2004).

Linear filters can be parameterized in many ways and can be made quasi-linear, or signal-level-dependent, by letting some of the parameters depend on input level or output level or some other control level. The *compressive gammachirp* is one such level-parameterized filter, an approximation to the gammachirp using movable poles and zeros; two versions, parallel and cascade gammachirp models (PrIGC and CasGC) have been explored (Irimo and Patterson, 2001; Unoki *et al.*, 2006). The all-pole gammatone filter (APGF), one-zero gammatone filter (OZGF), all-pole filter cascade (APFC), and pole–zero filter cascade (PZFC) are similarly

given a compressive nonlinear response via movement of their poles (Lyon, 1997; Katsiamis *et al.*, 2007; Lyon *et al.*, 2010a).

Kim *et al.* (1973) introduced a model that incorporated ten cascaded stages of two-pole filters modified to have nonlinear damping terms in their differential equations. In the small-signal linear limit, their system is a 10th-order all-pole filter. It is close to an APGF, but the 10 stages have their natural frequencies decreasing at 3% per stage (over a total range of less than a half octave), so it is also a short piece of an APFC. The distributed nonlinearity was motivated by hydrodynamic wave propagation, so it resembles a nonlinear APFC in that respect, as well. At the time, with borrowed time on a PDP-12 minicomputer, ten stages with one output was all they could simulate. Motivated partly by interaction with Molnar, Lyon and Mead (1988) extended this system to a full multi-output APFC analog VLSI cochlea using nonlinear two-pole stages. Nonlinear distortion products that arise in such cascades are not modeled in quasi-linear auditory filter models such as the PZFC but can be included in dynamic models such as the CAR-FAC.

The filter-cascade family of auditory filter models is treated here, like other families, mainly in its quasi-linear version. But its architecture does provide a natural framework for incorporating nonlinear processes that interact with the traveling wave. A dynamic time-domain version of the PZFC model for processing sounds in machine-hearing applications can include instantaneous and fast-acting nonlinear effects in the cascaded filter stages. This application of the PZFC was introduced in a previous paper (Lyon *et al.*, 2010b). To avoid confusion between the quasi-linear auditory filter modeling application and the machine-hearing applications, the dynamic time-domain version of the PZFC is now referred to as the CAR-FAC.

The OZGF is treated here because it is a very simple gammatone-like abstraction of the quasi-linear PZFC, sharing many of its properties, including description in terms of level-dependent s -plane pole damping, a linear low-frequency tail and good asymmetric resonance shape that lead to good fits to masking data, and the ability to match physiological impulse responses. But as an approach to building machine-hearing systems, a parallel filterbank based on OZGF channels would not be nearly as computationally efficient as a cascade architecture, and would not have any natural relationship to traveling waves.

B. Level dependence via output-level feedback

In an AGC-based model, a feedback loop works to keep the output level from varying too much; the output level is fed back through parameters such that higher outputs lead to lower filter gains, resulting in a compressive input–output function. This scheme works well for auditory filter models that are parameterizing by their output level, as opposed to their input level. Rosen *et al.* (1998) have shown that the former provide better fits to masking data.

In the case of the PZFC auditory filter model, we control the damping of all the cascaded stages by the output level at one place, just as the other models are controlled by a single

output level. In the CAR-FAC, by contrast, all of the filter output levels interact in the AGC network to jointly control all of the damping parameters. Therefore, the PZFC filter model is not a perfect model of the CAR-FAC in action, but it works about like the other auditory filter models in this respect, with level control coming from a single filter's output.

C. Nonlinear frequency scales

A model for a single auditory filter channel is of limited use. The ear uses a large set, almost a continuum, of filter channels to analyze sounds into many parallel signals to send to the brain via the auditory nerve. For machine-hearing applications, a not-too-sparse set of channels is required. It is not clear what the sampling criterion should be for filterbanks, especially if the output is not being used just for a power measurement or just for signal reconstruction. About 50% overlap, relative to the equivalent rectangular bandwidth, will likely provide a more well-behaved representation of a sound than non-overlapping channels would.

Each equivalent rectangular bandwidth (ERB) at moderate levels (ERB_N), as estimated by psychophysical experiments, corresponds to about 0.89 mm on the BM (Glasberg and Moore, 1990; Moore, 1995), so that would be about 39 channels in 35 mm, without overlap, or 78 channels with 50% overlap. According to the Greenwood map, 0.89 mm is about a factor 0.88 in frequency from one channel to the next, in the upper octaves, or about 5.6 channels per octave. At 50% overlap, that is about 11 channels per octave. Machine hearing models typically use about 60 to 100 channels in total.

III. FILTER CASCADES

The structure of the filter cascades (whether all-pole, pole-zero, or other form) derives from a simple observation of how filter cascades can make good models of wave propagation in nonuniform systems such as the cochlea, starting with linear wave propagation and adding nonlinearity later.

A. How filter cascades work

The method known as WKB (or sometimes Liouville-Green) provides insight into wave propagation in nonuniform linear media such as the cochlea. The method says that if a wave is propagating from the input along one dimension, then the response from the input to any point can be found by composing the relative responses from each point to the next along that dimension, using local parameters as though the medium were uniform, with some correction gains, if needed, to enforce conservation of energy as the medium changes. The factors that depend only on local properties can be interpreted as filters arranged in cascade (Lyon, 1998):

$$H_n(\omega) \approx \prod_{j=1}^n \exp(-ik(\omega, x_j)\Delta x). \quad (1)$$

Here $H_n(\omega)$ is a net filter transfer function (of the type needed for a linear or quasi-linear auditory filter model) at place number n , and the individual factors in the product are

cascaded filter stages representing segments of length Δx of the wave propagation medium (in the case of the cochlea, from the base to any of a discrete set of places $x_n = n\Delta x$, x being distance along the basilar membrane). The wavenumber $k(\omega, x)$ is a function of both frequency and place, since the medium is nonuniform. In the case of the segmental approximation implied by the WKB method, $k(\omega, x_j)$ is the average value of k over segment number j ; that is, the segment is treated as if it were a short piece of a uniform medium.

The value of the function $k(\omega)$, a solution of the dispersion relation for the medium, is real for a lossless wave-propagation medium but can be complex to represent either dissipation or active amplification in the medium. Both positive and negative imaginary parts are needed to represent active gain followed by dissipation. The log magnitude gain of each cascaded stage is simply proportional to the imaginary part of k , while the phase delay is proportional to the real part.

Therefore, independent of the details and dimensionality of the underlying wave mechanics, the responses of the cochlea at a sequence of places are equivalent to the responses at the outputs of a sequence of cascaded filters. The WKB method constrains the design of those filters when the underlying physics is known. Alternatively, any design for a cascade of filters implies a corresponding approximate dispersion relation. The problem of designing practical runnable models than becomes the problem of finding simple rational transfer functions (poles and zeros) to approximate non-rational transfer functions of the form $\exp(-ik(\omega)\Delta x)$ for $k(\omega)$ resembling the actual mechanics of the cochlea. If the mechanics are not known well enough to lead to a good model, the alternative is to fit parameters for a simple stage transfer function, given whatever data are available.

Since $H_{n+1}(\omega)$ shares n factors, or filter stages, with $H_n(\omega)$, it is very efficient to process signals through an entire bank of filters concurrently; the computational cost per filter-bank output is just the cost of running a sound through a single simple stage filter.

Even for nonlinear and time-varying wave mechanics, one can reasonably assume that a nonlinear and time-varying filter cascade will be a useful structural analog and a fruitful modeling approach: modeling local behavior with local filters, shared over a bank of outputs.

B. Filter-cascade stages with zeros

The original model of Lyon (1982) incorporated pairs of both poles and zeros as anti-resonant notch filters in the filter cascade, motivated by the series-resonant circuits in the long-wave transmission-line model of Zweig *et al.* (1976). Lyon and Mead (1988) later focused on cascades of simpler two-pole stages, motivated by an analysis of a 2D short-wave model with pseudo-resonant behavior. With these all-pole cascades, it was hard to get a sharp enough high-side rolloff without excessive delay. Going back to the use of a zero pair at a frequency somewhat higher than the pole pair both gives a sharp cutoff and reduces the overall delay, as suggested by Lyon (1998). The PZFC therefore differs both

from the APFC of Lyon and Mead (1988) and Slaney and Lyon (1993) and from the early more complicated cascade-parallel pole-zero structure of Lyon (1982).

Both the APFC and the PZFC illustrate the fact that filter cascades can exhibit a very substantial group delay, even though they are minimum-phase filters. This delay corresponds to the wave propagation delay in the cochlea, and is associated with the steep high-frequency rolloff. The delay is adjustable in the filter models via the relative pole and zero positions.

Since a cochlea-like response arises from individual stages as simple as second-order filters, each described by a complex-conjugate pair of poles and a complex-conjugate pair of zeros in the s plane, that is the level of complexity chosen for the PZFC. If better data are found from cochlear mechanics, the stage model can be revised, perhaps to higher order, as needed.

C. The PZFC/CAR-FAC architecture

The cascaded filter stages used in the PZFC, and in its dynamic CAR-FAC extension, are second-order filters, each described by a complex-conjugate pair of poles and a complex-conjugate pair of zeroes in the s plane. The zeros are positioned slightly above the poles in frequency, leading to a peak in gain near the pole frequency, followed by a sharp gain drop at higher frequencies—an asymmetric resonator.

The initial (in quiet) positions of the poles and zeros are set for each stage, and level-dependence is achieved by modifying the pole damping in each stage in response to the filterbank's output levels. This modification of pole damping, or equivalently pole Q , corresponds to moving the pole along a circular trajectory in the s plane, as shown in Fig. 1, and thus the peak frequency of the resonance shifts a little as the gain and bandwidth of the resonance changes.

The initial pole positions are spaced proportional to nominal ERB as a function of frequency (from high to low in the cascade, to model wave propagation from base to apex), using the formula of Glasberg and Moore (1990). The zeros at each stage are placed at a frequency that is a

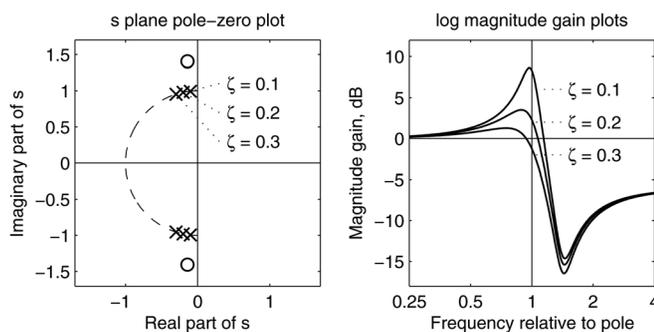


FIG. 1. Diagram of the motion of the poles of a PZFC or CAR-FAC stage in response to a gain-control feedback signal, and the effect on the resonator gain. The positions indicated by crosses in the s plane plot (left) correspond to pole damping ratios (ζ) of 0.1, 0.2, and 0.3, while the zero's damping ratio remains fixed at 0.1. Corresponding transfer function gains (right) of this asymmetric resonator stage do not change at low frequencies but vary by several decibels near the pole frequency. The fact that the stage gain comes back up after the dip has little effect in the transfer function of a cascade of such stages.

constant factor above the pole (typically about a half octave higher).

For the CAR-FAC, nonlinearity is incorporated by both a dynamic level-dependent positioning of the poles and an instantaneous cubic distortion at the output (between stages), like that between the bandpass filters in BPNL models. In the case of the PZFC, no instantaneous or dynamic nonlinearity is included, since the auditory filter framework used in fitting human masking data requires a quasi-linear filter model.

The PZFC filterbank architecture can be seen as intermediate between the all-pole filter cascade (Slaney and Lyon, 1993) on the one hand and cascade-parallel models (Lyon, 1982) on the other hand. As an auditory filter model with level dependence, the PZFC is quasi-linear but exhibits nonlinear compression. The compression exhibited by the dynamic CAR-FAC, on the other hand, includes both a fast-acting AGC part, similar to that of the “dynamic compressive gammachirp” (Irimo and Patterson, 2006), and an instantaneous part, from an odd-order nonlinearity similar to that in the “dual-resonance, nonlinear” (DRNL) model (Lopez-Poveda and Meddis, 2001) or the nonlinear model of Kim *et al.* (1973).

D. PZFC/CAR-FAC transfer functions

The complex transfer function of one stage of the linearized PZFC is a rational function of the Laplace transform variable s , of second order in both numerator and denominator, corresponding to a pair of zeros (roots of the numerator) and a pair of poles (roots of the denominator):

$$H(s) = \frac{s^2/\omega_z^2 + 2\zeta_z s/\omega_z + 1}{s^2/\omega_p^2 + 2\zeta_p s/\omega_p + 1}, \quad (2)$$

where ω_p and ω_z are the natural frequencies and ζ_p and ζ_z are the damping ratios of the poles and zeros, respectively.

Figure 2 shows the transfer function gain of all the outputs of the filter cascade, in the case of silence, and as adapted to a vowel sound at moderate level.

E. CAR-FAC implementation

In auditory-model-based machine-hearing applications of these filters, the first processing step, the dynamic cochlear model, is the CAR-FAC based on the PZFC auditory filter model plus a coupled AGC loop (Lyon *et al.*, 2010b), as illustrated in Fig. 3. It produces a bank of bandpass-filtered, compressed, half-wave rectified, output signals that represent the response of the inner hair cells along the length of the cochlea. The CAR-FAC can be viewed as approximating the auditory nerve's instantaneous firing rate as a function of cochlear place, modeling both the frequency filtering and the compressive or AGC characteristics of the human cochlea (Lyon, 1990); it currently models the inner hair cell as a simple half-wave rectifier rather than a better model with depletion and smoothing.

The filters are implemented as discrete-time approximations at sample rate f_s (22 050 Hz, for example) by mapping

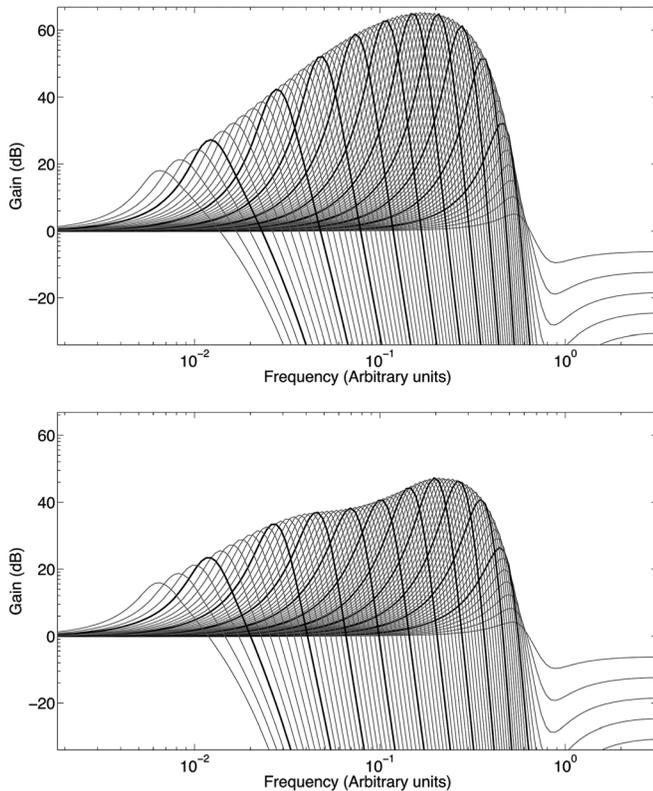


FIG. 2. Adaptation of the overall filterbank response at each output tap. (Top) The initial response of the filterbank before adaptation. (Bottom) The response after adaptation to a human/a/vowel of 0.6 s duration. The plots show that the adaptation affects the peak gains (the upper envelope of the filter curves shown), while the tails, behaving linearly, remain fixed.

the poles and zeros from the s plane to the z plane using $z = \exp(s/f_s)$ as is conventional in the simple “pole-zero mapping” or “matched Z-transform” method of digital filter design (Yang, 2009).

The CAR-FAC poles are modified dynamically by feedback from a spatial/temporal loop filter, or smoothing network, thereby making an AGC system. The smoothing network takes the half-wave-rectified outputs of all channels, applies smoothing in both the time and place dimensions, and uses both local and more global averages of the filterbank response (that is, a mixture of different time scales and space scales of smoothing) to proportionately increase the

pole damping of each stage. This coupled AGC smoothing network descends from one first described by Lyon (1982); in that work, the loop filter directly controlled a post-filterbank gain rather than a pole damping as it does in more recent versions.

IV. FITTING FILTERS TO MASKING DATA

A. Human notched-noise masking data

A *notched noise* consists of two frequency bands of noise with a quiet frequency band (the notch) between them, as shown in Fig. 4. Such noises have been used as maskers in tone-detection experiments, to get at the filtering that the auditory system does, since the 1950s (Webster *et al.*, 1952); the method became more important in the 1970s (Patterson, 1976; Patterson and Nimmo-Smith, 1980), after it became clear that listeners were employing an “off-frequency listening” strategy to detect masked tones. That is, listeners would effectively choose to pay attention to a filter channel with best signal-to-noise (SNR) (or tone-to-masker) ratio, rather than to the channel with the filter’s peak frequency matched to the probe tone. Experiments with *asymmetric notched noise*, that is, using probe tones placed off-center in the notches, provided a way to better assess the effects of different parts of the auditory filter shape.

A number of teams have repeated and extended experiments on human detection of tones in asymmetric notched-noise maskers (Lutfi and Patterson, 1984; Glasberg *et al.*, 1984; Moore *et al.*, 1990; Rosen *et al.*, 1998; Baker *et al.*, 1998). Others provided increasingly sophisticated analyses to derive auditory filter shapes that would predict the experimental data (Patterson and Moore, 1986; Moore and Glasberg, 1987; Glasberg and Moore, 1990; Rosen and Baker, 1994; Irino and Patterson, 2001; Patterson *et al.*, 2003; Unoki *et al.*, 2006). Their data and methods are used and extended in this paper to provide parameter fits for the OZGF and PZFC and related filter models.

Two large datasets, covering a range of frequency patterns and levels, with several subjects in each set, have been used to fit and compare different auditory filter models; the same datasets are used in the present study. The first (Baker *et al.*, 1998) used nine subjects and seven tone

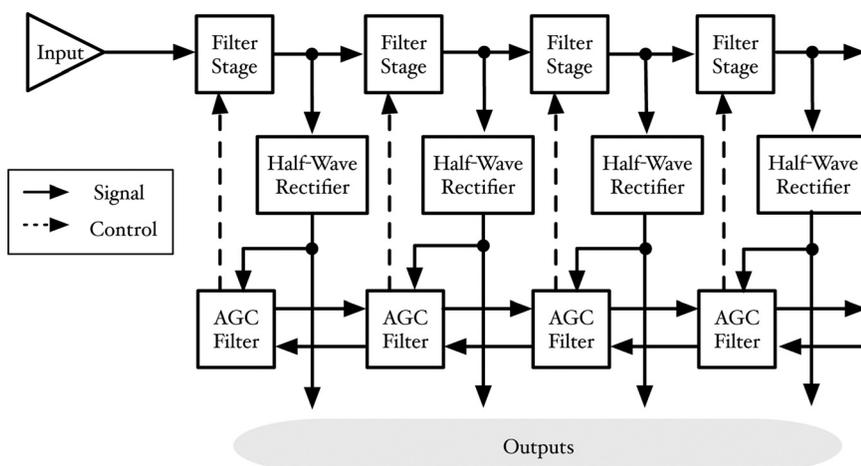


FIG. 3. Schematic of the CAR-FAC design. The cascaded filter stages (upper row) have variable peak gains, which are controlled by their damping ratios, set by feedback from the coupled AGC filters (lower row). The “control” signals can be fast-acting in response to an onset, but usually vary slowly. In the case of quasi-linear PZFC filter models, the control values are static but level-dependent.

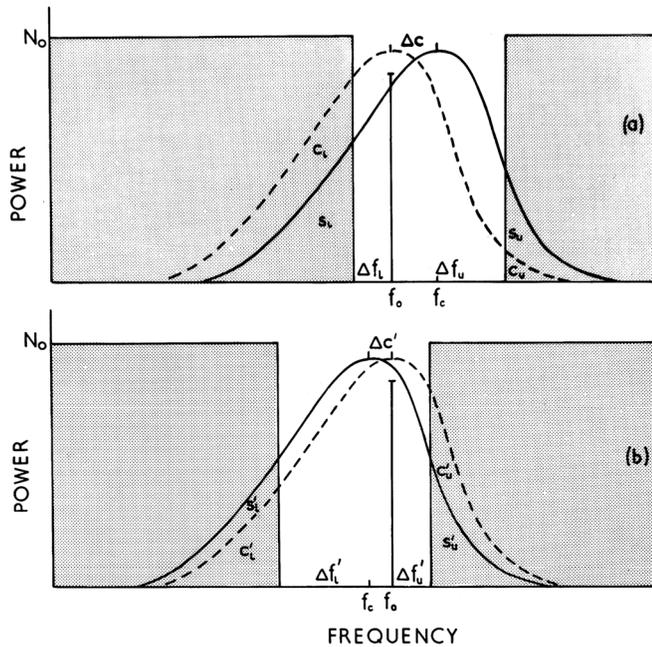


FIG. 4. The “asymmetric notched noise” masking paradigm, and data from human listeners, were introduced with this figure that explains the significant shifts between the filter with best SNR and the filter with CF at the probe-tone frequency (Patterson and Nimmo-Smith, 1980). In each example, the filter with best probe-tone-to-masking-noise ratio in its output (solid curve) is near the filter with highest probe-tone output power (dashed curve, filter with peak at probe-tone frequency f_o) but shifted in the direction that reduces the noise power output (generally toward a point slightly to the right of the center of the notch).

frequencies, with noises that were flat (*white*) within the noise bands; the second (Glasberg and Moore, 2000) used four subjects and five tone frequencies, with a *uniformly exciting* noise, that is, spectrally shaped to provide approximately equal excitation per critical band. For most, including the present, filter fitting studies, only the mean thresholds across the subjects within each group were used. Both datasets, totaling 1277 mean detection threshold data points, can be accommodated together in fitting auditory filter parameters.

B. Nonlinear filter fitting approach

Here *fitting* refers to the process of finding the best values of the parameters of auditory filter models; *best* means that the model’s predicted tone detection thresholds are as good as possible, that is, that the sum of squared errors, between the human data and the model prediction, is minimized. This is a basic *least squares* optimization problem, but since the system (predictions as a function of parameters) is nonlinear, it takes a more complicated search to find the optimum.

For the nonlinear optimization process, the methods of Irino and Patterson (1997), Patterson *et al.* (2003), Unoki *et al.* (2006) are followed, using the Levenberg–Marquardt algorithm and the combined datasets (Baker *et al.*, 1998; Glasberg and Moore, 2000); none of this work would have been possible without the generous help of all of these authors, and their code and data.

Each auditory filter model has its own parameters that need to be adjusted; in addition, there are three non-model parameters that are fitted in every case.

- (1) The center frequency of the filter: for each set of filter parameters, the filter’s CF dimension is searched to optimize the SNR at the filter output.
- (2) The noise floor: a parameter P_0 that represents an internal noise power (added to any other noise that is present) is needed to model the approach of masked threshold to absolute threshold at low levels.
- (3) The detection threshold criterion: a parameter K represents the output SNR at which the model predicts detection of the probe tone.

In the filter fitting framework and MATLAB code provided by Unoki, several changes have been made to get better fits, and to fit to a wider class of models.

- (1) Level-dependent parameters depend on the output level of a filter (sometimes a linear “passive” filter) with noise-only input, as opposed to noise-plus-probe level; using the latter was found to provide an unfair extra clue to predicting the probe level.
- (2) Optionally, the level-dependent filter model itself can be used as the level-detection filter, in a feedback configuration, necessitating an inner search over filter output level for each set of parameters being evaluated in the search.
- (3) The nonlinear fit search integrates optimization of P_0 , but for each set of parameters being searched, K is quickly computed linearly (in dB space).
- (4) P_0 is redefined as an input-referred noise level, so that filters with variable gain will behave right; it had previously been used as a noise level added at the filter output after SNR optimization.
- (5) The search for best CF was made via nearly continuous, rather than discrete choices, so that the system being optimized would be differentiable in all parameters; this change helped the search converge to a better optimum, compared to published results on the combined dataset (Unoki *et al.*, 2006).

Figure 5 shows the structure of the filter model configurations considered in this and prior work. For the present study, the PrIGC and CasGC models are modified to be feedback versions by taking the level detector input from the final output instead of using a feed-forward connection from a “passive” filter. The passive filter is still used as part of the PrIGC and CasGC model structures, but in the feedback configuration the passive filter’s output is no longer what controls the level-dependent parameters.

In all cases, only a few parameters (one or two in each model) were allowed to depend on level, and those only with a dependence that is linear in the filter output level in dB. The model parameters are optionally frequency-dependent, to support fitting one model at multiple probe frequencies (Patterson *et al.*, 2003); extra parameters optionally let the filter parameters, and P_0 and K , be linear or quadratic functions of the probe frequency (on an auditory ERB-rate frequency scale). In counting model parameters (“filter coefficients”), the parameters that allow frequency dependence are also

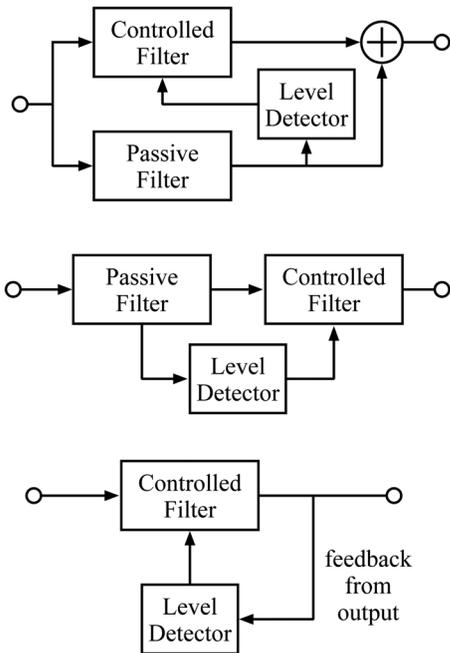


FIG. 5. Parallel (top), cascade (middle), and feedback (bottom) structures for level-dependent auditory filter models. The PrIGC and CasGC models originally used the upper and middle structures as a way to achieve a controllable gain near the tip while keeping a stable low-frequency tail. In the case of the PrIGC model, following an older parallel roex structure, the adder is actually adding power levels (Unoki *et al.*, 2006), not signals, so this model structure does not correspond to an actual filter.

counted, but the 6 parameters (the “nonfilter coefficients”) that allow P0 and K to be quadratic functions of frequency are not counted. Generally, models that lead to a low rms error with few filter parameters are preferred; another useful criterion is the ability of a model fitted to one dataset to predict results of another; that is, to *generalize* across different conditions and subjects.

C. Fitted psychoacoustic filter shapes

Parameter fits were done for several filter models in this study; the model types are displayed in Table I for easy reference. All are feedback configurations, including the feedback versions of PrIGC and CasGC described above. Other models included are the simplified gammatone-family types (OZGF and its special cases, the APGF and the differentiated APGF or DAPGF) and the two filter-cascade types, APFC and

TABLE I. Acronyms for the different auditory filter models discussed are tabulated here for reference; they are ordered from simplest to most complex, or number of fitted parameters required, roughly.

Acronym	Definition
APGF	All-pole gammatone filter
DAPGF	Differentiated APGF
OZGF	One-zero gammatone filter
APFC	All-pole filter cascade
PZFC	Pole-zero filter cascade
PZFC5	PZFC with movable zeros
PrIGC	Parallel gammachirp
CasGC	Cascade gammachirp

PZFC. Parameters were fitted using the datasets described above, which had previously been used with a range of roex and gammachirp models without feedback. By using feedback control of parameters, all of the models easily achieve a compressive input-output relationship, thereby avoiding the need for other constraints that had previously been used to ensure sensible level dependence (Patterson *et al.*, 2003).

Concerning the ability to fit the data by optimizing a large number of parameters, Rosen *et al.* (1998) had conjectured, “...models with similar goodness-of-fit lead to filter shapes that are very similar. Therefore it is not particularly important which model is chosen from the ‘better-fitting’ ones. The relatively large number of good-fitting filter shapes is also an indication that the roex(p, w, t) shape may be too flexible. There are likely to be other adequate functional forms with fewer controlling parameters (e.g., Irino and Patterson, 1997 [gammachirp]; Lyon, 1996 [all-pole gammatone]).”

It has already been shown that the gammachirp can provide better fits with fewer parameters than the roex (Unoki *et al.*, 2006). The current work finds that the APGF, OZGF, and PZFC can provide better fits with fewer parameters than the various roex filters, and also better and/or with fewer parameters than the gammachirp versions. At the lowest numbers of parameters, two extremes of the OZGF—the APGF with 3 parameters and DAPGF with 4 parameters—are the best-fitting models. At 5 parameters, the OZGF with optimized zero location fits best. With 6 or more parameters, the PZFC fits best. If it is “not particularly important which model is chosen,” then it is probably a good idea to use models that are easy to run efficiently and that connect well to traveling waves.

These experiments confirm that a filter architecture that gives a natural coupling of gain, bandwidth, and shape to level-dependent parameters provides a parsimonious model with no loss of realism (relative to these datasets at least). At the same time, this architecture provides the stable low-frequency tail similar to that which had been added by developing compound structures (parallel or cascade) for the level-dependent roex and gammachirp models.

These experiments also confirm the value of the AGC-like form of feedback shown in Fig. 5 (bottom) (Lyon, 1990; Carney, 1993), where the filter’s own output is the signal whose level controls its parameters. The filter models based on feedback from the output always provided better fits with fewer parameters than the models with forward control from the input noise spectrum. In the typical alternative to using the filter’s own output to control its parameters, others (Zhang *et al.*, 2001; Unoki *et al.*, 2006; Rosen and Baker, 1994; Tan and Carney, 2003) have used a *control-path* filter whose output controls the parameters of the signal path. This approach can be easier to implement, as it is a feed-forward computation, but the idea of a separate control-path filter is hard to reconcile with the structure of the auditory system.

In the PZFC model, the zero frequency is a parameterized ratio times the pole frequency (the ratio that maps pole frequency to zero frequency can optionally be allowed to vary linearly or quadratically with pole frequency, using the available fitting parameters).

The pole bandwidth is computed proportional to the ERB, using factors such that the b_2 parameter (which may itself be frequency dependent) is the nominal bandwidth relative to the ERB when the order is 4:

$$BW_p = \frac{1}{2} \sqrt{n_2} b_2 ERB_w, \quad (3)$$

where the “order” parameter n_2 is the gammatone order, or the channels-per-ERB of the PZFC. The bandwidth factor b_2 depends geometrically on level (linearly in the dB or log-bandwidth domain) according to

$$\log_{10}(b_2) = \log_{10}(B_2) + B_2^1 \left(\frac{P_p - 60}{20 B_2 n_2} \right), \quad (4)$$

where P_p is the output power of the filter, on a dB scale (P_p is typically 60 to 100 dB for the filter gains and input levels used, and corresponds to the input level in dB SPL amplified by the level-dependent filter transfer function). The B_2 parameter is the nominal bandwidth (relative to the ERB) at an output power of 60 dB. Other factors scale the level dependence parameter B_2^1 to a convenient value; the inclusion of B_2 in the denominator in the scaling means that there will be less level dependence in high-relative-bandwidth channels, when B_2 is frequency dependent. This formula is an example of what are called structural parameters embedded in the model; such parameters have not been counted in comparing the model complexities.

Fits with linear instead of geometric pole bandwidth variation have also been tried; also with and without the B_2 in the denominator of the level dependence. The model described works best, by a small margin, so in that sense these structural parameters have been fitted.

Similar optimizations have been done in the construction and parameterization of the other models that were previously published; such decisions are not explicitly accounted for in the parameter counts.

An example of a parameterization of the PZFC model, with 9 fitted parameters, is shown in Table II.

D. PZFC and OZGF provide good fits with few parameters

Katsiamis *et al.* (2007) predicted that “the DAPGF or OZGF will provide a significant benefit in applications that need a better model of level dependence or a better

TABLE II. A PZFC model with 9 filter parameters (fit 530); the channel density is fixed at 2 and not counted. The pole damping b_2 is computed from the CF-dependent B_2 as modified by the output power level (in dB) times B_2^1 . In this version of the model, the zeros do not move with level.

Name	Function	f dependence	#
b_1	Zero bandwidth	Quadratic	3
B_2	Pole bandwidth	Quadratic	3
B_2^1	Pole BW level dependence	Constant	1
n_2	Channels per ERB	2 (fixed)	0
f_{rat}	Ratio of zero freq. to pole freq.	Linear	2

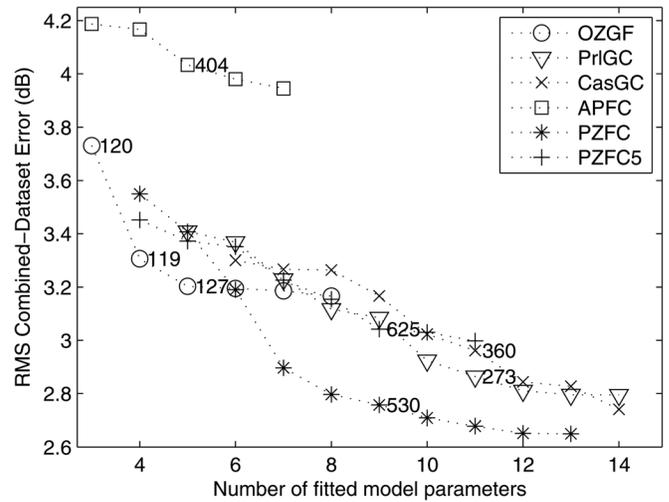


FIG. 6. Threshold-prediction rms errors for various filter models, versus number of fitted parameters, on the combined dataset. The fit numbers are for reference only; different filter models are identified by different symbols, as shown in the legend. For each model type, only the fit with lowest error at each number of parameters is shown; the errors are monotonically decreasing, since adding a free parameter never increases the error. The PZFC5 variants (+), such as fit 625, are the PZFC modified to have the zeros move with level, parallel with the poles, as opposed to the original PZFC (*) for which the zeros are fixed.

low-frequency tail behavior”; this prediction is somewhat confirmed with respect to human masked-threshold data. As shown in Fig. 6, the best fits at each number of parameters are always OZGF or PZFC models.

When the OZGF is specialized to an APGF or DAPGF (no zero, or zero at DC, respectively), the zero-position parameter is not counted; the model with only 3 parameters (fit 120) is an APGF model, with only a linear dependence of bandwidth on frequency; at 4 parameters, a quadratic frequency dependence is added, and the DAPGF (fit 119) is best. At 5 parameters, the zero is added to make a full OZGF, fit 127; at 6 parameters, nothing helps much.

With more parameters (7 to 13), the PZFC provides the best fits. The gammachirp models typically need 3 to 5 more parameters to fit the data as well. These results suggest that the OZGF is “simplest” but that the connection of the PZFC to the underlying traveling wave mechanics makes it most “realistic” with not much additional complexity. Since the PZFC is also the one that has the lowest computational cost when used for a filterbank (with the possible exception of the APFC), it is a good base for the CAR-FAC used in machine-hearing applications.

The implication that one or another filter model is really the “best” should be evaluated with a dose of skepticism, in light of the possibility of *over-fitting* that is a common issue in machine learning. This possibility was investigated by *training* the models on just one dataset [the one from Baker *et al.* (1998)], and then *testing* on the other (Glasberg and Moore, 2000), to see how well the retrained model generalizes from the training set to the test set. The models that generalize well are often not the ones with the lowest fitting error on the combined dataset.

As previously observed by Patterson *et al.* (2003), the difference between the datasets from the two labs is larger

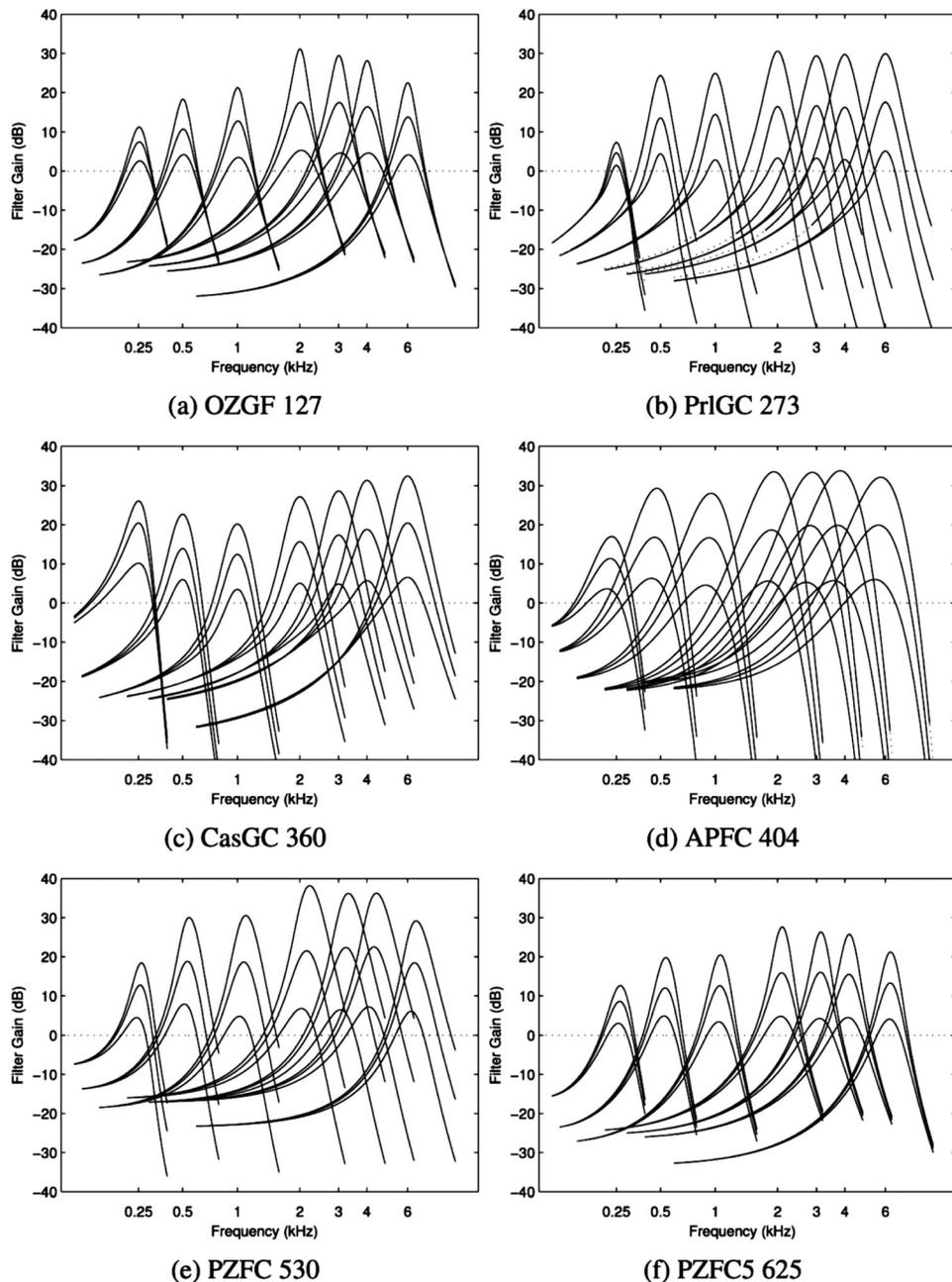


FIG. 7. Auditory filter gain plots for the best of each of six model types. The frequency axes are on the ERB-rate scale. In each case, the curves represent filter gain when the tone detection thresholds are 30 dB (highest curves), 50 dB, and 70 dB (lowest curves). The curve spacing is related to the input-output compression: curves close together, as at 250 Hz, correspond to a response that is only slightly compressive, while curve tips 15 dB apart represent a 4:1 compressive response. The model ERBs range from approximately the nominal ERB to more than twice that.

than the typical differences between models, with the Glasberg and Moore data showing low level dependence at some frequencies, and high at others, compared to the more regular Baker *et al.* (1998) data. In the present experiment, the OZGF and PZFC5 with 4 to 8 parameters yield the best generalization to the Glasberg and Moore data at frequencies below 4000 Hz, with PZFC close behind; but at 4000 Hz the gammachirps do best at 6 and more parameters. These results suggest that the PZFC5 has no net disadvantage relative to the PZFC, but otherwise do not tell us which model is best.

The filter shapes for a representative model of each type, in the range that generalizes not too poorly, are plotted in Fig. 7. The shape details show the different “personalities” of the various model types in trying to fit the data.

The OZGF with only 5 parameters (fit 127) illustrates the point that a simple model using one cluster of movable

poles and one fixed zero is a fairly good fit to the data. As shown in Fig. 8, the shapes of the OZGF’s simplest special cases with even fewer parameters (with the one zero moved to zero or to infinity) are generally similar to the best OZGF fit found, except in the low-frequency tail, and still fit fairly well, since moving the poles still gives a realistic level-dependent coupling of shape, bandwidth, and peak gain. This behavior is inherited by the filter cascades, but a few more parameters are needed to describe the placement of the zeros in the PZFC.

V. IMPULSE RESPONSES AND PHYSIOLOGICAL DATA

From auditory-nerve data, one estimates impulse responses—really first-order Volterra kernels—by the process of *reverse correlation*: every time the neuron fires an action potential in response to a noise, a piece of the noise

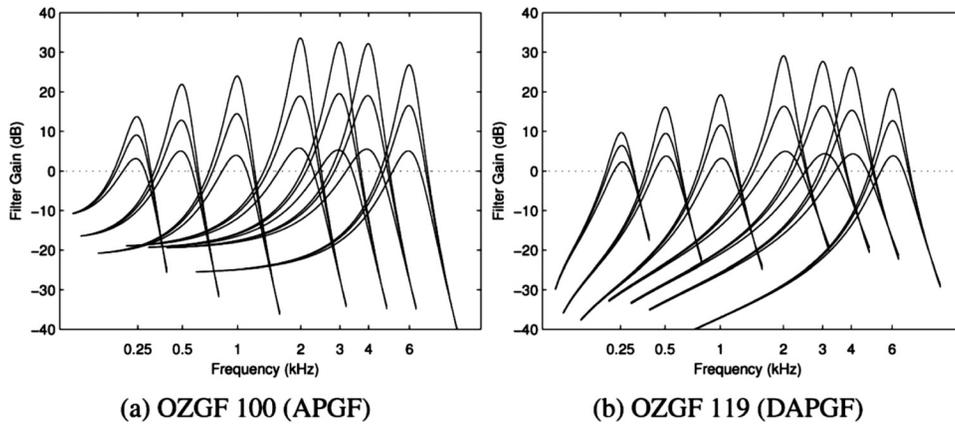


FIG. 8. The two degenerate cases of the OZGF, the APGF (left) and the DAPGF (right), provide good fits with only 4 parameters (quadratic bandwidth, and a bandwidth-level-dependence coefficient). They differ from the better-fitting OZGFs (the ones with more parameters) in the low-frequency tails, especially in the differentiated case (the DAPGF, which has a zero at DC).

waveform that led up to it is added to a waveform accumulation buffer. The shape of the sum in the buffer (divided by the number of segments added) approaches the effective time-reversed impulse response of the cochlea at the point innervated by the neuron, as described by [de Boer \(1976\)](#) and [de Boer and de Jongh \(1978\)](#). These correlation-derived impulse responses are called *revcor functions*.

Filter models whose impulse responses closely resemble the neural *revcor* data, or corresponding mechanical data, are thus physiologically supported. Indeed, the gammatone model was introduced as a simple approximation to *revcor* functions measured in cats ([Johannesma, 1972](#)).

Data from mechanical and neural experiments ([Carney et al., 1999](#); [Robles and Ruggero, 2001](#); [Shera, 2001](#)) show that the zero-crossing times, or local phases, of the filter's output in response to impulses are variably spaced, unlike the zero-crossings of the gammatone, and do not change much with signal level. This observation puts an important constraint on how the auditory filter model should behave as its level-dependent parameters are varied.

In the case of the gammatone, gammachirp, and APGF models, the zero-crossing times of the impulse responses remain exactly fixed as the exponential decay time parameter is varied; this variation corresponds to moving the poles of filters horizontally (varying real part) in the s plane. In the case of gammachirp (and its special case, the gammatone), this stability of zero crossings is apparent from the time-domain description in which a decay-time-dependent envelope multiplies a fixed oscillating term that determines the zero crossings, as has been pointed out by [Iriño and Patterson \(2001\)](#) when they fitted gammachirp filters to both human masking data and cat auditory nerve impulse responses:

$$h_{\text{GCF}}(t) = t^{N-1} \exp(-bt) \cos(\omega_r t + c \log(t)). \quad (5)$$

In the case of the APGF, a similar relationship is apparent when the impulse response is written in a similar way, which involves a Bessel function in place of the sinusoid:

$$h_{\text{APGF}}(t) = t^N \exp(-bt) j_{N-1}(\omega_r t), \quad (6)$$

where j_{N-1} is a spherical Bessel function.

[Shera \(2001\)](#) has also shown that this direction of pole motion in basilar-membrane-impedance models leads to nearly fixed zero-crossing locations.

For the gammatone, APGC, OZGF, PZFC, and other filters representable as rational transfer functions, the zero crossings are exactly fixed if the poles and zeros are all moved horizontally in the s plane by equal amounts. This observation follows from the shifting property of the Laplace transform, which says that shifting the Laplace transform by d corresponds to multiplying the impulse response by $\exp(dt)$. For real d , corresponding to horizontal movement, this change of envelope will not affect the zero crossings; it corresponds to adjusting the real b in the factor $\exp(-bt)$ in the above equations. Of course, if d is too big, moving one or more poles into the right half of the s plane, then b is negative and $\exp(-bt)$ will increase without bound; nevertheless, the zero-crossing times will not change.

In some systems, it may be more natural to vary the damping, or pole Q , leaving the poles' natural frequencies fixed, in which case the poles move along a circle in the s plane, centered at the origin and of radius equal to the natural frequency ω_n (in a simple harmonic oscillator, natural frequency is determined by the mass and spring constant, independent of the damping). This is what the reported CAR-FAC implementation used ([Lyon et al., 2010b](#)). For the filter model fitting, it makes no difference, since the optimal CF is selected for each data point. When damping is low, horizontal motion is nearly tangent to the circle, so these directions are not so different; but they may be different enough to make a testable difference in how well a model matches the observed zero-crossing stability. Moving the zeros by different amounts from the poles can approximately compensate for the effect of moving along non-horizontal trajectories, at least in the early part of the impulse response. In the long-time limit, the decaying impulse response will ring at the ringing frequency of the pole with the longest time constant (that is, later zero-crossing intervals will be determined by the imaginary part of the pole with real part closest to zero).

In the filter-cascade models, the poles and zeros of the different stages move in a coordinated way based on the level parameter, but in amounts proportional to their frequencies, so the shifting property does not exactly apply. Nevertheless, reasonable choices of pole and zero motion directions and amounts lead to stable zero crossings, as illustrated in [Fig. 9](#). The first fitted PZFC model, in which the zeros are fixed and the poles move, does not achieve stable

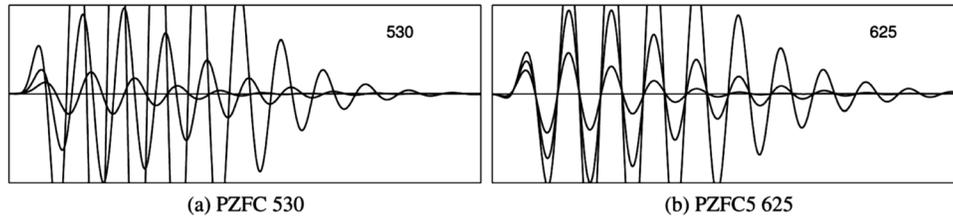


FIG. 9. The impulse responses for the 1 kHz channel of two versions of the PZFC, at three tone threshold levels. The large (off-scale) curves are for the noise level that leads to 30 dB SPL tone threshold, the medium (full-scale) curves for 50 dB, and the small curves for 70 dB. The PZFC5 variant is designed to have stable zero-crossing times; the difference is apparent in the plots.

zero crossings—the zeros need to move about as much as the poles do. In a modified model called PZFC5, the bandwidths of the zeros change in proportion to the bandwidth of the poles, at each stage, with the constant of proportionality being a fitted parameter that is optimized at about 1.14; the resulting fits to the masking data are not quite as good as the original PZFC is. In such a cascade, the zeros stay close to the poles of an earlier stage, approximately canceling out most of the effects of the cascade except for a few uncanceled poles in stages just basal to the place under consideration; the net filter is close to an all-pole model, and the fitting results are very close to the APGF or OZGF fitting results, as shown in Fig. 7. Zero-crossing stability is not enforced, but the free parameter that determines how much the zeros move happens to give stable zero crossings in typical fits. Other ways of coupling the zero motion to the pole motion were not as good, in terms of zero-crossing stability.

Impulse responses and instantaneous frequency analysis from revcor and cochlear mechanics experiments also show a “glide” or “chirp” in the response of the cochlea, with an upward glide at high CF and a downward glide at low CF (Tan and Carney, 2003); this glide corresponds to the unequally spaced zero crossings mentioned above. In general, if the filters are minimum phase, the glide direction will be determined by the frequency response gain asymmetry; filters with a sharp high side will have an upward glide; that is, the initial cycles of the response will be at lower frequencies than the later cycles. To the extent that filter models get the asymmetry right, they will get the glide about right, as long as they are minimum phase, which most of the considered models are (gammatone and gammachirp filters are not quite, but their complex versions, all-pole versions, and approximations are).

Another important aspect of impulse-response data is the group delay; again, delay is determined by the amplitude response in the case of minimum-phase models. The gammatone filters have their delay tied to filter shape and bandwidth, with higher orders providing more delay along with a somewhat different overall shape. The PZFC allows, by adjustment of the zeros, considerable room to tune the delay, to more or less than the delay of the typical order-4 gammatone-family filters. The PZFC has the property that as the cascaded segments are more finely divided, the overall shape and delay can be kept fixed by letting the zeros move closer to the poles, whereas all-pole cascades will generally have too much delay as they move to higher orders.

VI. CONCLUSION

Modeling cochlear wave propagation as a filter cascade has given rise to the PZFC filter model, which provides better fits to human masked-threshold data than any other known auditory filter models. The model is easily modified to have approximately level-independent zero-crossing times as seen in auditory nerve physiology. These two good fits do not appear to be achieved simultaneously, as they require different treatment of the positions of the zeros in the cascaded filter stages, but the generalization experiments suggest that the PZFC5 with stable zero crossings is at least an excellent compromise.

The PZFC leads to the CAR-FAC time-domain implementation that can incorporate both dynamic level dependence and instantaneous nonlinearities. The connection of the cascade architecture to the traveling-wave nature of the cochlea gives the CAR-FAC at least the potential to model cochlear nonlinearities fairly accurately.

If for no other reason than its computational efficiency (only second-order per channel) the PZFC/CAR-FAC is the architecture of choice for processing sounds in machine-hearing applications. But since it also provides excellent fits, with few parameters, to human psychophysical data, and also connects well with cochlear hydromechanics including nonlinear level-dependent phenomena such as dynamic amplitude compression, zero-crossing stability, and cubic distortion tones, it is a model that may be useful in forming bridges between the various facets of hearing research.

- Bacon, S. P. (2004). “Overview of auditory compression,” in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Ray, and A. N. Popper (Springer-Verlag, New York), pp. 1–17.
- Baker, R. J., Rosen, S., and Darling, A. M. (1998). “An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable,” in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 81–88.
- Carney, L. H. (1993). “A model for the responses of low-frequency auditory-nerve fibers in cat,” *J. Acoust. Soc. Am.* **93**, 401–417.
- Carney, L. H., McDuffy, M. J., and Shekhter, I. (1999). “Frequency glides in the impulse responses of auditory-nerve fibers,” *J. Acoust. Soc. Am.* **105**, 2384–2391.
- de Boer, E. (1976). “Cross-correlation function of a bandpass nonlinear network,” *Proc. IEEE* **64**, 1443–1444.
- de Boer, E., and de Jongh, H. R. (1978). “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” *J. Acoust. Soc. Am.* **63**, 115–135.
- Duifhuis, H. (2004). “Comment on ‘An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity’ [J. Acoust. Soc. Am. **114**, 2112–2117],” *J. Acoust. Soc. Am.* **115**, 1889–1890.

- Flanagan, J. L. (1960). "Models for approximating basilar membrane displacement," *Bell Sys. Tech. J.* **39**, 1163–1191.
- 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.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched noise data," *Hearing Res.* **47**, 103–138.
- Glasberg, B. R., and Moore, B. C. J. (2000). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," *J. Acoust. Soc. Am.* **108**, 2318–2328.
- 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.
- Irino, T., and Patterson, R. D. (1997). "A time-domain, level-dependent auditory filter: The gammachirp," *J. Acoust. Soc. Am.* **101**, 412–419.
- Irino, T., and Patterson, R. D. (2001). "A compressive gammachirp auditory filter for both physiological and psychophysical data," *J. Acoust. Soc. Am.* **109**, 2008–2022.
- Irino, T., and Patterson, R. D. (2006). "A dynamic compressive gammachirp auditory filterbank," *IEEE Trans. Audio Speech Language Process.* **14**, 2222–2232.
- Johannesma, P. I. M. (1972). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Proc. IPO Symposium on Hearing Theory*, edited by B. L. Cardozo (IPO, Eindhoven), pp. 58–69.
- Katsiamis, A. G., Drakakis, E. M., and Lyon, R. F. (2007). "Practical gammatone-like filters for auditory processing," *EURASIP J. Audio Speech Music Process.* **2007**, 63685.
- Kim, D. O., Molnar, C. E., and Pfeiffer, R. R. (1973). "A system of nonlinear differential equations modeling basilar-membrane motion," *J. Acoust. Soc. Am.* **54**, 1517–1529.
- Lopez-Poveda, E. A., and Meddis, R. (2001). "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am.* **110**, 3107–3118.
- Lutfi, R. A., and Patterson, R. D. (1984). "On the growth of masking asymmetry with stimulus intensity," *J. Acoust. Soc. Am.* **76**, 739–745.
- Lyon, R. F. (1982). "A computational model of filtering, detection, and compression in the cochlea," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1282–1285.
- Lyon, R. F. (1990). "Automatic gain control in cochlear mechanics," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 395–420.
- Lyon, R. F. (1997). "All-pole models of auditory filtering," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific Publishing, Singapore), pp. 205–211.
- Lyon, R. F. (1998). "Filter cascades as analogs of the cochlea," in *Neuromorphic Systems Engineering: Neural Networks in Silicon*, edited by T. S. Lande (Kluwer Academic Publishers, Norwell, MA), pp. 3–18.
- Lyon, R. F., Katsiamis, A. G., and Drakakis, E. M. (2010a). "History and future of auditory filter models," in *IEEE International Conference on Circuits and Systems*, pp. 3809–3812.
- Lyon, R. F., and Mead, C. (1988). "An analog electronic cochlea," *IEEE Trans. Acoust. Speech Signal Process.* **36**, 1119–1134.
- Lyon, R. F., Rehn, M., Bengio, S., Walters, T. C., and Chechik, G. (2010b). "Sound retrieval and ranking using sparse auditory representations," *Neural Comp.* **22**, 2390–2416.
- Moore, B. C. J. (1995). "Frequency analysis and masking," in *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego), pp. 161–205.
- 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 Res.* **28**, 209–225.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1990). "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.* **88**, 132–140.
- Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* **59**, 640–654.
- Patterson, R. D., and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic Press, London), pp. 123–177.
- Patterson, R. D., and Nimmo-Smith, I. (1980). "Off-frequency listening and auditory-filter asymmetry," *J. Acoust. Soc. Am.* **67**, 229–245.
- Patterson, R. D., Unoki, M., and Irino, T. (2003). "Extending the domain of center frequencies for the compressive gammachirp auditory filter," *J. Acoust. Soc. Am.* **114**, 1529–1542.
- Peterson, L. C., and Bogert, B. P. (1950). "A dynamical theory of the cochlea," *J. Acoust. Soc. Am.* **22**, 369–381.
- Ranke, O. F. (1950). "Theory of operation of the cochlea: A contribution to the hydrodynamics of the cochlea," *J. Acoust. Soc. Am.* **22**, 772–777.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**, 1305–1352.
- Rosen, S., and Baker, R. J. (1994). "Characterising auditory filter non-linearity," *Hearing Res.* **73**, 231–243.
- Rosen, S., Baker, R. J., and Darling, A. (1998). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," *J. Acoust. Soc. Am.* **103**, 2539–2550.
- Shera, C. A. (2001). "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," *J. Acoust. Soc. Am.* **110**, 332–348.
- Slaney, M., and Lyon, R. F. (1993). "On the importance of time—A temporal representation of sound," in *Visual Representations of Speech Signals*, edited by M. Cooke, S. Beet, and M. Crawford (John Wiley and Sons, Sussex), pp. 95–116.
- Tan, Q., and Carney, L. (2003). "A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," *J. Acoust. Soc. Am.* **114**, 2007–2020.
- Unoki, M., Irino, T., Glasberg, B., Moore, B. C. J., and Patterson, R. D. (2006). "Comparison of the roex and gammachirp filters as representations of the auditory filter," *J. Acoust. Soc. Am.* **120**, 1474–1492.
- Webster, J. C., Miller, P. H., Thompson, P. O., and Davenport, E. W. (1952). "The masking and pitch shift of pure tones near abrupt changes in a thermal noise spectrum," *J. Acoust. Soc. Am.* **24**, 147–152.
- Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**, 266–285.
- Yang, W. Y. (2009). "Continuous-time systems and discrete-time systems," in *Signals and Systems with MATLAB* (Springer-Verlag, Berlin), pp. 292–293.
- Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001). "A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression," *J. Acoust. Soc. Am.* **109**, 648–670.
- Zweig, G., Lipes, R., and Pierce, J. R. (1976). "The cochlear compromise," *J. Acoust. Soc. Am.* **59**, 975–982.