

A Pole–Zero Filter Cascade Provides Good Fits to Human Masking Data and to Basilar Membrane and Neural Data

Richard F. Lyon

Google, Inc.

Abstract. A cascade of two-pole–two-zero filters with level-dependent pole and zero dampings, with few parameters, can provide a good match to human psychophysical and physiological data. The model has been fitted to data on detection threshold for tones in notched-noise masking, including bandwidth and filter shape changes over a wide range of levels, and has been shown to provide better fits with fewer parameters compared to other auditory filter models such as gammatone filters. Originally motivated as an efficient machine implementation of auditory filtering related to the WKB analysis method of cochlear wave propagation, such filter cascades also provide good fits to mechanical basilar membrane data, and to auditory nerve data, including linear low-frequency tail response, level-dependent peak gain, sharp tuning curves, nonlinear compression curves, level-independent zero-crossing times in the impulse response, realistic instantaneous frequency glides, and appropriate level-dependent group delay even with minimum-phase response. As part of exploring different level-dependent parameterizations of such filter cascades, we have identified a simple sufficient condition for stable zero-crossing times, based on the shifting property of the Laplace transform: simply move all the s -domain poles and zeros by equal amounts in the real- s direction. Such pole-zero filter cascades are efficient front ends for machine hearing applications, such as music information retrieval, content identification, speech recognition, and sound indexing.

Keywords: Auditory filter model, filter cascade, automatic gain control

PACS: 43.64.Bt, , 43.66.Ba

INTRODUCTION

Filter cascades, such as the pole–zero filter cascade (PZFC, cascades of two-pole–two-zero stages), are a good basis for modeling the filtering due to the cochlear traveling-wave structure. Psychoacoustic data can be used to fit the parameters of such models to predict tone-in-masker thresholds, and the models can also be adjusted to match physiological data that imply stable zero-crossing times in the impulse responses at different levels. The results are not far from results with other rational-transfer-function models, such as one-zero gammatone filters (OZGF), but the cascade structure provides a better basis for efficient machine-hearing systems, and a better basis for incorporating nonlinear effects that propagate toward lower-CF places.

Two large datasets of human tone detection thresholds in the presence of notched-noise maskers, covering a range of frequency patterns and levels, with several subjects in each set, have previously been used to fit and compare different auditory filter models; we have used the same datasets. The first [1] used nine subjects and seven tone frequencies, with noises that were flat (white) within the noise bands; the second [4] 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 this work, we used only the mean thresholds across the subjects. Both datasets, totalling 1277 mean

TABLE 1. A PZFC model with 9 filter parameters (fit 530); the channel density is fixed at 2 per ERB 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	params
b_1	Zero bandwidth	Quadratic	3
B_2	Pole bandwidth	Quadratic	3
B_2^1	Pole BW level dependence	Constant	1
n_2	Channel density (channels per ERB)	2 (fixed)	0
f_{rat}	Ratio of zero freq. to pole freq.	Linear	2

detection threshold data points, can be accommodated together in fitting auditory filter parameters.

For the nonlinear optimization process, we follow Irino, Patterson, and Unoki [5, 10, 13] in using the Levenberg–Marquardt algorithm and the combined datasets [1, 4]. Each auditory filter model has its own parameters that need to be adjusted; in addition, there are several non-filter parameters to find in the search (detection threshold K and absolute threshold P_0 [10], treated as an effective *noise floor*).

In the filter fitting framework and MATLAB code provided by Unoki, we made several changes to get better fits, and to fit to a wider class of models, including using only the noise-only filter output levels in a feedback configuration to set the level-dependent parameters, and improving the search for best CF to make the fits converge more accurately.

In all cases, only a few parameters (one to three 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. We also adopted the strategy for simultaneously fitting at multiple probe frequencies [10]. Generally, we seek models that lead to a low rms error with few filter parameters.

The parameterization of one good-fitting PZFC, fit 530, is described in Tab. 1.

FITTED PSYCHOACOUSTIC FILTER SHAPES

We have fitted the all-pole filter cascade (APFC), the OZGF (including the special case with the zero at infinity, the all-pole gammatone filter, APGF, and the case with the zero at DC, the differentiated APGF or DAPGF), and the PZFC, and new feedback versions of the parallel and cascade compressive gammachirp (PrIGC and CasGC) models, to the data described above.

We have found that the APGF, OZGF, and PZFC can provide better fits 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—with output level controlling pole damping via feedback, are the best-fitting models. At 5 parameters, the OZGF with optimized zero location is best. With more parameters, the PZFC is best.

These experiments confirm the usefulness of the AGC feedback configuration [2, 8], where the filter’s own output is the signal whose level controls its parameters. The

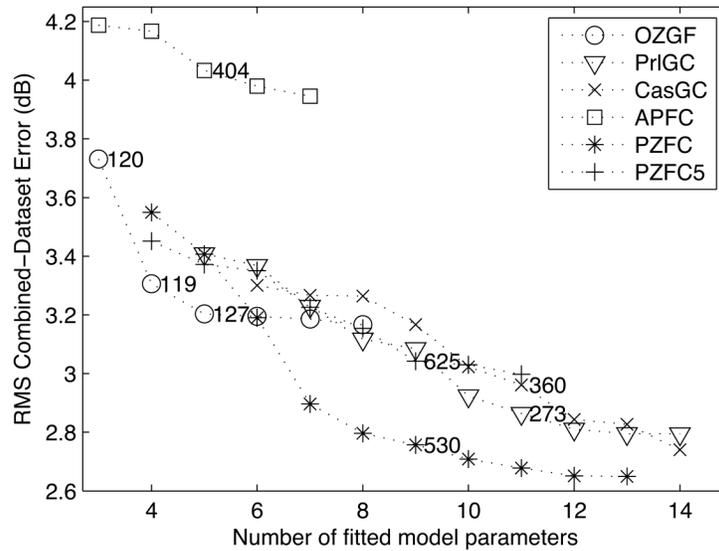


FIGURE 1. The rms error from fitting on the combined dataset. At each number of parameters, only the best result of each filter model is shown. The fits on the combined data suggest some winners and losers, but in their respective best cases, all of the different filter models generalize from the training to testing datasets nearly equally well (not shown). Fit 120 is an APGF, and fit 119 is a DAPGF, special cases of the OZGF with the zero at infinity and at DC; they provide fair fits with just 3 and 4 parameters. The PZFC5 model (zeros moving along with poles) is generally not quite as good as the original PZFC, but some cases such as fit 625 with 7 parameters are not bad.

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 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 our 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). For the frequency dependence, we use Glasberg and Moore’s formula for the equivalent rectangular bandwidth (ERB) as a function of frequency. Then we compute a pole bandwidth proportional to it, using a factor that may itself be frequency dependent and level dependent.

We predicted that “the DAPGF or OZGF will provide a significant benefit in applications that need a better model of level dependence or a better low-frequency tail behavior” [7]; this prediction is somewhat confirmed with respect to human masked-threshold data. As shown in Fig. 1, the best fits at each number of parameters are always OZGF or PZFC models.

With 6 and more parameters, the PZFC provides the best fits. The OZGF is “simplest,” but the connection of the PZFC to the underlying traveling wave mechanics makes it most “realistic” with not much additional complexity.

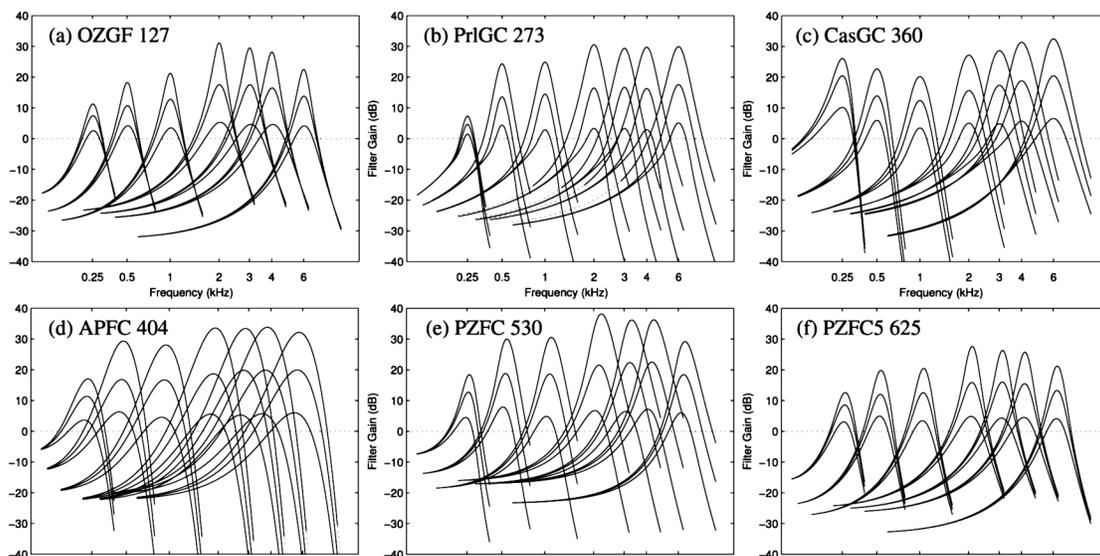


FIGURE 2. Auditory filter gain plots for a selected representative of each of six model types. The frequency axes are on the ERB-rate scale. In each case, the curves represent filter gain when the tone detection thresholds are 30 dB (highest curves), 50 dB, and 70 dB (lowest curves). The curve spacing is related to the input–output compression: curves close together, as at 250 Hz, correspond to a nearly linear response, while curve tips 15 dB apart represent a 4:1 compressive response (15 dB gain decrease per 20 dB level increase). The effective rectangular bandwidths range from approximately the nominal ERB to more than twice that.

However, we must temper this interpretation, in light of the possibility of *overfitting* that is a common issue in machine learning and other modeling paradigms. We investigated this possibility by *training* the models on just one data set (the one from Baker et al.), and then *testing* on the other (Glasberg & Moore), to see how well the retrained model generalizes from the training set to the test set. The models that generalize well are very often not the ones with the lowest fitting error on the combined dataset. The OZGF and PZFC5 with 4 to 8 parameters yield the best generalization to the G&M data at frequencies below 4000 Hz, with PZFC close behind; but at 4000 Hz the gammachirps do best at 6 and more parameters (it has previously been shown that fits to the G&M data behave very differently at the five different probe frequencies [10]). These results suggest that the PZFC5 has no net disadvantage relative to the PZFC, but otherwise do not tell us which model is best.

IMPULSE RESPONSES FROM PHYSIOLOGICAL DATA

In neural experiments, impulse responses are estimated as revcor functions. We want filter models whose impulse responses resemble the neural revcor data, or corresponding mechanical data. Data from mechanical and neural experiments [3, 11, 12] 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, but like those of the models considered here), 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 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 basic gammachirp (and its special case, the gammatone), this stability 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 Irino and Patterson [6] (but in the case of the CasGC, the level-dependent stage has its poles and zeros moving orthogonal to that direction). 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.

Similarly, for the APFC, 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.

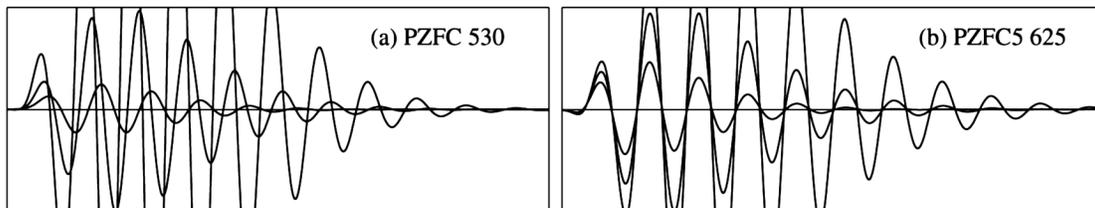


FIGURE 3. The impulse responses for the 1 kHz channel of two versions of the PZFC, at three tone threshold levels. The large (off-scale) curves are for the noise level that leads to 30 dB SPL tone threshold, the medium (full-scale) curves for 50 dB, and the small curves for 70 dB. The PZFC5 variant is designed to have more stable zero-crossing times; the difference is apparent in the plots.

In the filter cascade models, we assume that poles and zeros of the different stages move in a coordinated way, 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. 3. The first fitted PZFC model, in which the zeros are fixed and the poles move, does not achieve stable 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 fitted shapes are very close to the APGF or OZGF fitting results, as shown in Fig. 2.

In our machine hearing work to date, we have not needed stable zero crossing times, since we have not been doing binaural ITD extraction or other operations that might depend on it [9].

CONCLUSION

Modeling cochlear wave propagation as a filter cascade has given rise to the PZFC, 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 cascade structure can also provide a good basis for modeling distortion products that propagate to their own lower-CF place, and for modeling suppression via instantaneous compression and cross-channel-coupled automatic gain control. Future work should tie down the parameters to make these effects match experimental data.

ACKNOWLEDGMENTS

None of this work would have been possible without the generous help, code, and data, of Patterson, Irino, Unoki, Baker, Rosen, Darling, Glasberg, and Moore.

REFERENCES

- [1] Baker RJ, Rosen S, Darling AM (1998) An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable. In: Palmer AR, Rees A, Summerfield AQ, Meddis R (eds) *Psychophysical and Physiological Adv Hearing*, Whurr, pp. 81–88
- [2] Carney LH (1993) A model for the responses of low-frequency auditory-nerve fibers in cat. *J Acoust Soc Am* 93:401–417
- [3] Carney LH, McDuffy MJ, Shekhter I (1999) Frequency glides in the impulse responses of auditory-nerve fibers. *J Acoust Soc Am* 105:2384–2391
- [4] Glasberg BR, Moore BCJ (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *J Acoust Soc Am* 108:2318–2328
- [5] Irino T, Patterson RD (1997) A time-domain, level-dependent auditory filter: The gammachirp. *J Acoust Soc Am* 101:412–419
- [6] Irino T, Patterson RD (2001) A compressive gammachirp auditory filter for both physiological and psychophysical data. *J Acoust Soc Am* 109:2008–2022
- [7] Katsiamis AG, Drakakis EM, Lyon RF (2007) Practical gammatone-like filters for auditory processing. *EURASIP J Audio, Speech, and Music Processing* 2007
- [8] Lyon RF (1990) Automatic gain control in cochlear mechanics. In: Dallos P, et al (eds) *The Mechanics and Biophysics of Hearing*, Springer-Verlag, pp. 395–420
- [9] Lyon RF, Rehn M, Bengio S, Walters TC, Chechik G (2010) Sound retrieval and ranking using sparse auditory representations. *Neural computation* 22:2390–2416
- [10] Patterson RD, Unoki M, Irino T (2003) Extending the domain of center frequencies for the compressive gammachirp auditory filter. *J Acoust Soc Am* 114:1529–1542
- [11] Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. *Physiol Rev* 81:1305–1352
- [12] Shera CA (2001) Intensity-invariance of fine time structure in basilar-membrane click responses: implications for cochlear mechanics. *J Acoust Soc Am* 110:332–348
- [13] Unoki M, Irino T, Glasberg B, Moore BCJ, Patterson RD (2006) Comparison of the roex and gammachirp filters as representations of the auditory filter. *J Acoust Soc Am* 120:1474–1492