1 Supplementary material

The asymmetric amplitude profile of Gammatones makes them suitable to model temporal masking in auditory filterbanks [**?**]. Yet, the introductory paper on Gammatone wavelets [**?**] does not provide a formula for deducing σ from the specification of a quality factor Q. In this appendix, we provide a rationale for choosing the topmost center frequency ξ of a Gammatone wavelet filter bank in a discrete-time setting. Then, we relate the bandwidth parameter σ to the choice of a quality factor Q.

Motivation

Time reversal of a real signal $\boldsymbol{x}(t)$ is equivalent to the complex conjugation of its Fourier transform $\widehat{\boldsymbol{x}}(\omega)$. As a consequence, the Fourier transform modulus $|\widehat{\boldsymbol{x}}(\omega)|$ is not only invariant to translation, but also invariant to time reversal. Yet, although invariance to translation is needed for classification, invariance to time reversal is an undesirable property. A simple way to break invariance to time reversal is to choose $\boldsymbol{\psi}(t)$ as an asymmetric wavelet instead of a Gabor symmetric wavelet.

The complex-valued Gammatone wavelet is a modification of the real-valued Gammatone auditory filter, originated in auditory physiology. The Gammatone auditory filter of center frequency ξ is defined as a gamma distribution of order $N \in \mathbb{N}^*$ and bandwidth σ modulated by a sine wave, that is,

$$t^{N-1} \exp(-2\pi\sigma t) \cos(2\pi\xi t).$$

For a fixed σ, the integer N controls the relative shape of the envelope, becoming less skewed as N increases. Psychoacoustical experiments have shown that, for $N = 4$, the Gammatone function provides a valid approximation of the basilar membrane response in the mammalian cochlea [**?**,**?**,**?**]. In particular, it is asymmetric both in the time domain and in the Fourier domain, which allows to reproduce the asymmetry of temporal masking as well as the asymmetry of spectral masking [**?**]. It is thus used in computational models for auditory physiology [**?**]. However, it does not comply with the Grossman-Morlet admissibility condition, because it has a non-negligible average. In addition, because the Gammatone auditory filter takes real values in the time domain, its Fourier transform satisfies Hermitian symmetry, which implies that it does not belong to the space H^2 of analytic functions. More generally, there are no real-valued functions in H^2 [**?**].

Related work

With the aim of building a pseudo-analytic admissible Gammatone wavelet, [**?**] have modified the definition of the Gammatone auditory filter, by replacing the real-valued sine wave $\cos(2\pi\xi t)$ by its analytic part $\exp(2\pi \mathrm{i}\xi t)$ and by taking the first derivative of the gamma distribution, thus ensuring null mean. The definition of the Gammatone wavelet becomes

$$\boldsymbol{\psi}(t) = \big(2\pi(\mathrm{i}-\sigma)t^{N-1} + (N-1)t^{N-2}\big)\exp(-2\pi\sigma t)\exp(2\pi\mathrm{i}\xi t)$$

in the time domain, and

$$\widehat{\boldsymbol{\psi}}(\omega) = \frac{\mathrm{i}\omega \times (N-1)!}{(\sigma + \mathrm{i}(\omega - \xi))^N}$$

in the Fourier domain. Besides its biological plausibility, the Gammatone wavelet enjoys a near-optimal time-frequency localization with respect to the Heisenberg uncertainty principle. Furthermore, this time-frequency localization tends to optimality as N approaches infinity, because the limit $N \rightarrow +\infty$ yields a Gabor wavelet [**?**]. Last but not least, the Gammatone wavelet transform of finite order N is causal, as opposed to the Morlet wavelet transform, which makes it better suited to real-time applications. From an evolutionary point of view, it has been argued that the Gammatone reaches a practical compromise between time-frequency localization and causality constraints [**?**].

Center frequency parameter

In order to preserve energy and allow for perfect reconstruction, the Gammatone wavelet filter bank must satisfy the inequalities

$$1 - \varepsilon \leq |\hat{\boldsymbol{\phi}}(\omega)| + \sum_{\gamma} |\hat{\boldsymbol{\psi}}(2^{\gamma}\omega)| + |\hat{\boldsymbol{\psi}}(-2^{\gamma}\omega)| \leq 1$$

for all frequencies ω in the audible range, where ε is a small margin [**?**]. Let ξ_{s} be the sampling rate of the audio signal $\boldsymbol{x}(t)$ at hand. Satisfying the equation above near the Nyquist frequency $\omega = \frac{1}{2}\xi_{\mathrm{s}}$ can be achieved by placing the log-frequency $\log_2 \xi$ of the first (topmost) wavelet in between the log-frequency $\log_2(\xi \times 2^{-1/Q})$ of the second wavelet and the log-frequency $\log_2(\xi_{\mathrm{s}} - \xi)$ of the mirror of the first wavelet. We obtain the equation

$$\log_2 \xi - \log_2(\xi \times 2^{-1/Q}) = \log_2(\tfrac{1}{2}\xi_{\mathrm{s}} - \xi) - \log_2 \xi,$$

of which we deduce the normalized identity

$$\frac{\xi}{\xi_{\mathrm{s}}} = \frac{1}{1 + 2^{1/Q}}.$$

For $Q = 1$, this yields a center frequency of $\xi = \frac{1}{3}\xi_{\mathrm{s}}$. For greater values of Q, the center frequency ξ tends towards the Nyquist frequency $\frac{1}{2}\xi_{\mathrm{s}}$.

Bandwidth parameter

The quality factor Q of the Gammatone wavelet is defined as the ratio between the center frequency ξ of the wavelet $\hat{\boldsymbol{\psi}}(\omega)$ and its bandwidth B in the Fourier domain. This bandwidth is given by the difference between the two solutions ω of the following equation:

$$\frac{|\hat{\boldsymbol{\psi}}(\omega)|}{|\hat{\boldsymbol{\psi}}(\xi)|} = \frac{\omega}{\xi} \times \left(1 + \frac{(\omega - \xi)^2}{\sigma^2}\right)^{-N/2} = r,$$

where the magnitude cutoff $r < 1$ is most often set to $\sqrt{\frac{1}{2}}$. Let $\Delta\omega = \omega - \xi$. Since $\Delta\omega \ll \xi$, we may linearize the exponentiation to $(-\frac{N}{2})$ by the first-order Taylor expansion

$$\left(1 + \frac{(\Delta\omega)^2}{\sigma^2}\right)^{\frac{N}{2}} \approx 1 + \frac{N}{2}\left(\frac{\Delta\omega}{\sigma}\right)^2.$$

Furthermore, we may rewrite the ratio $\frac{\omega}{\xi}$ as $\frac{\xi+\Delta\omega}{\xi} = (1 + \frac{\Delta\omega}{\xi})$. This leads to a quadratic equation of the variable $\Delta\omega$:

$$1 + \frac{\Delta\omega}{\xi} = r \times \left(1 + \frac{N}{2}\left(\frac{\Delta\omega}{\sigma}\right)^2\right),$$

which rewrites as

$$(\Delta\omega)^2 - \frac{2\sigma^2}{Nr\xi}(\Delta\omega) - \frac{2\sigma^2}{N}\frac{1-r}{r} = 0.$$

The discriminant of the above equation is

$$D = \frac{4\sigma^2}{Nr}\left(\frac{\sigma^2}{Nr\xi^2} + 2(1-r)\right),$$

which is a positive number because $r < 1$. The bandwidth B of $\hat{\boldsymbol{\psi}}$ is given by the difference between the two solutions of the quadratic equation, that is:

$$B = \sqrt{D} = \frac{2\sigma}{\sqrt{Nr}}\sqrt{\frac{\sigma^2}{Nr\xi^2} + 2(1-r)}.$$

Now, let us express the parameter σ as a function of some required bandwidth B at some cutoff threshold r. After having raised the above to its square and rearranged the terms, we obtain another quadratic equation, yet of the variable σ^2:

$$\frac{4}{N^2r^2\xi^2}(\sigma^2)^2 + \frac{8(1-r)}{Nr}\sigma^2 - B^2 = 0.$$

We multiply the equation by $\frac{1}{8}N^2r^2\xi^2 \neq 0$, thus yielding

$$\frac{1}{2}(\sigma^2)^2 + Nr(1-r)\xi^2\sigma^2 - \frac{N^2r^2\xi^2B^2}{8} = 0.$$

The discriminant of the above equation is equal to:

$$D' = N^2 r^2 \xi^2 \left(\xi^2 (1-r)^2 + \tfrac{1}{4} B^2 \right).$$

Because this discriminant is strictly positive, the equation has two roots. Yet, only the smallest of the two satisfies $\sigma^2 < \xi^2$. Therefore, we have

$$\sigma^2 = Nr(1-r) \left(\sqrt{1 + \left(\frac{B}{2\xi(1-r)} \right)^2} - 1 \right) \xi^2.$$

If the filter bank has to be approximately orthogonal, we typically set B to

$$B = \left(\sqrt[Q]{2} - \frac{1}{\sqrt[Q]{2}} \right) \times \xi.$$

We conclude with the following normalized closed form for σ:

$$\frac{\sigma}{\xi} = \sqrt{Nr(1-r)} \left(\sqrt{1 + \frac{\left(\sqrt[Q]{2} - \frac{1}{\sqrt[Q]{2}} \right)^2}{4(1-r)^2}} - 1 \right).$$

References

1. Andén, J., Mallat, S.: Deep scattering spectrum. IEEE Transactions on Signal Processing **62**(16), 4114–4128 (2014)
2. Cohen, L.: Time-frequency analysis. Prentice Hall (1995)
3. Flanagan, J.: Models for approximating basilar membrane displacement. Journal of the Acoustical Society of America **32**(7), 937–937 (1960)
4. Grossmann, A., Morlet, J.: Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM Journal on Mathematical Analysis **15**(4), 723–736 (1984)
5. Lyon, R., Katsiamis, A., Drakakis, E.: History and future of auditory filter models. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCS) (2010)
6. Patterson, R.: Auditory filter shapes derived with noise stimuli. Journal of the Acoustical Society of America **59**(3), 640–654 (1976)
7. Pressnitzer, D., Gnansia, D.: Real-time auditory models. In: Proceedings of the International Computer Music Conference (ICMC) (2005)
8. Venkitaraman, A., Adiga, A., Seelamantula, C.S.: Auditory-motivated Gammatone wavelet transform. Signal Processing **94**, 608–619 (2014)
9. Zwicker, E., Fastl, H.: Psychoacoustics: Facts and models, vol. 22. Springer Science & Business Media (2013)