ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

A biomimetic basis for the perception of natural sounds

H. Ammari and B. Davies

A biomimetic basis for the perception of natural sounds

Habib Ammari*   Bryn Davies*

Abstract

Arrays of subwavelength resonators can mimic the biomechanical properties of the cochlea, at the same scale. We derive, from first principles, a modal time-domain expansion for the scattered pressure field due to such a structure and propose that these modes should form the basis of a signal processing architecture. We investigate the properties of such an approach and show that higher-order gammatone filters appear by cascading. Further, we propose an approach for extracting meaningful global properties from the coefficients, tailored to the statistical properties of so-called natural sounds.

1 Introduction

1.1 Biomimetic signal processing

Humans are exceptionally good at recognising different sound sources in their environment and there have been many attempts at designing artificial approaches that can replicate this feat. The human auditory system first amplifies and filters sounds biomechanically in the peripheral auditory system before processing the transduced neural signals in the central auditory signal. With a view to trying to mimic this world-beating system, we consider using an artificial routine with a similar two-step architecture: physical filtering followed by additional processing stages.

There has been much attention paid to designing biomimetic structures that replicate the biomechanical properties of the cochlea [1–7]. At the heart of any such structure are graded material parameters, so as to replicate the spatial frequency separation of the cochlea. In particular, a size-graded array of subwavelength resonators can be designed to have similar dimensions to the cochlea and respond to an appropriate range of audible frequencies [1]. An acoustic subwavelength resonator is a cavity with material parameters that are greatly different from the background medium [3]. Bubbly structures of this kind can be constructed, for example, by injecting air bubbles into silicone-based polymers [8, 9].

A graded array of resonators effectively behaves as a distributed system of band-pass filters [10]. The choice of kernel filter for auditory processing has been widely explored. Popular options include windowed Fourier modes [11, 12], wavelets [13–17] and learned basis functions [18]. In particular, gammatone filters (Fourier modes windowed by gamma distributions) have been shown to approximate auditory filters well and, thanks also to their relative simplicity, are used widely in modelling auditory function as a result [10, 19–21]. We will prove that, at leading order, an array of N subwavelength resonators behaves as an array of N gammatone filters.

The human auditory system is known to be adapted to the structure of the most important inputs and exhibits greatly enhanced neural responses to natural and behaviourally significant sounds such as animal and human vocalisations and environmental sounds [22]. It has been observed that such sounds, often known collectively as *natural sounds*, display certain statistical properties [22–25]. By design, most music also falls into this class; music satisfying these properties sounds "much more pleasing" [23]. Thus, it is clear that the human auditory system is able to account for global properties of a sound and that a biomimetic processing architecture needs to replicate this. Many attempts have been made to extract these. We propose using the parameters of the observed statistical distributions as meaningful and tractable examples of global properties to be used in artificial representations of auditory signals.

---
*Department of Mathematics, ETH Zurich, Rämistrasse 101, CH-8092 Zürich, Switzerland.
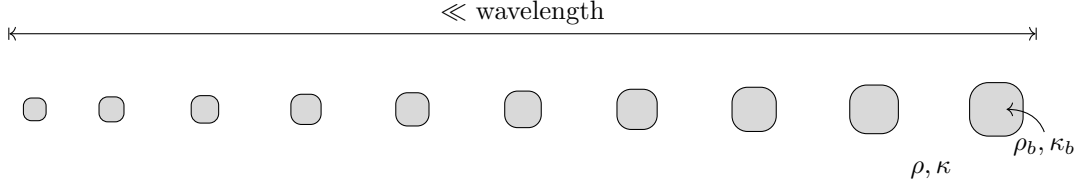 ✉ habib.ammari@math.ethz.ch
  ✉ bryn.davies@sam.math.ethz.ch

Figure 1: A graded array of subwavelength resonators.

1.2 Main Contributions

In Section 3 we derive results that describe the resonant properties of a system of N resonators in three dimensions and prove a modal decomposition in the time domain. This formula takes the form of N spatial eigenmodes with first-order gammatone coefficients. Further to this, we show in Section 4 that a cascade of these filters equates to filtering with higher-order gammatones and that extracting information from temporal averages is stable to deformations. Finally, in Section 5.1 we focus our attention on the class of natural sounds, which we define as sounds satisfying certain (widely observed) statistical and spectral properties. Using these properties, we propose a parametric coding approach that extracts the global properties of a sound.

2 Boundary integral operators

2.1 Problem setting

We are interested in studying wave propagation in a homogeneous background medium with $N \in \mathbb{N}$ disjoint bounded inclusions, which we will label as $D_1, D_2, \ldots, D_N \subset \mathbb{R}^3$. We will assume that the boundaries are all Lipschitz continuous and will write $D = D_1 \cup \cdots \cup D_N$. In order to replicate the spatial frequency separation of the cochlea, we are interested in the case where the array has a size gradient, meaning each resonator is slightly larger than the previous, as depicted in Figure 1.

We will denote the density and bulk modulus of the material within the bounded regions D by ρ_b and κ_b, respectively. The corresponding parameters for the background medium are ρ and κ. The wave speeds in D and $\mathbb{R}^3 \setminus \overline{D}$ are given by

$$v_b = \sqrt{\frac{\kappa_b}{\rho_b}}, \qquad v = \sqrt{\frac{\kappa}{\rho}}.$$

We also define the dimensionless contrast parameter

$$\delta = \frac{\rho_b}{\rho}.$$

We will assume that $\delta \ll 1$, meanwhile $v_b = O(1)$, $v = O(1)$ and $v_b/v = O(1)$.

2.2 Boundary integral operators

The Helmholtz single layer potential associated to the domain D and wavenumber $k \in \mathbb{C}$ is defined, for some density function $\varphi \in L^2(\partial D)$, as

$$\mathcal{S}_D^k[\varphi](x) := \int_{\partial D} G(x - y, k)\varphi(k) \, \mathrm{d}\sigma(y), \qquad x \in \mathbb{R}^3,$$

where G is the Helmholtz Green's function, given by

$$G(x, k) := -\frac{1}{4\pi|x|} e^{\mathrm{i}k|x|}.$$

The Neumann-Poincaré operator associated to D and $k \in \mathbb{C}$ is defined as

$$\mathcal{K}_D^{k,*}[\varphi](x) := \int_{\partial D} \frac{\partial G(x - y, k)}{\partial \nu(x)} \varphi(y) \, \mathrm{d}\sigma(y), \qquad x \in \partial D,$$

where $\partial/\partial\nu$ denotes the outward normal derivative on ∂D. These two integral operators are related by the conditions

$$\frac{\partial}{\partial\nu}\mathcal{S}_D^k[\varphi]\big|_{\pm}(x) = \left(\pm\frac{1}{2}I + \mathcal{K}_D^{k,*}\right)[\varphi](x), \qquad x \in \partial D, \tag{2.1}$$

where the subscripts $+$ and $-$ denote evaluation from outside and inside the boundary ∂D, respectively, and I is the identity operator on $L^2(\partial D)$.

2.3 Asymptotic properties

The single layer potential and the Neumann–Poincaré operator both have helpful asymptotic expansions as $k \to 0$ (see *e.g.* [26]). In particular, we have that

$$\mathcal{S}_D^k[\varphi] = \mathcal{S}_D[\varphi] + k\mathcal{S}_{D,1}[\varphi] + O(k^2), \tag{2.2}$$

where $\mathcal{S}_D := \mathcal{S}_D^0$ (*i.e.* the Laplace single layer potential) and

$$\mathcal{S}_{D,1}[\varphi](x) := \frac{1}{4\pi\mathrm{i}} \int_{\partial D} \varphi(y)\,\mathrm{d}\sigma(y).$$

One crucial property to note is that $\mathcal{S}_D$ is invertible. Similarly,

$$\mathcal{K}_D^{k,*}[\varphi] = \mathcal{K}_D^*[\varphi] + k\mathcal{K}_{D,1}[\varphi] + k^2\mathcal{K}_{D,2}[\varphi] + k^3\mathcal{K}_{D,3}[\varphi] + O(k^4), \tag{2.3}$$

where $\mathcal{K}_D^* := \mathcal{K}_D^{0,*}$, $\mathcal{K}_{D,1} = 0$,

$$\mathcal{K}_{D,2}[\varphi](x) := \frac{1}{8\pi} \int_{\partial D} \frac{(x-y)\cdot\nu(x)}{|x-y|}\varphi(y)\,\mathrm{d}\sigma(y) \quad \text{and} \quad \mathcal{K}_{D,3}[\varphi](x) := \frac{\mathrm{i}}{12\pi} \int_{\partial D} (x-y)\cdot\nu(x)\varphi(y)\,\mathrm{d}\sigma(y).$$

Several of the operators in the expansion (2.3) can be simplified when integrated over all or part of the boundary ∂D. As proved in *e.g.* [27, Lemma 2.1], it holds that for any $\varphi \in L^2(\partial D)$ and $i = 1, \ldots, N$,

$$\begin{aligned}
&\int_{\partial D_i} \left(-\frac{1}{2}I + \mathcal{K}_D^*\right)[\varphi]\,\mathrm{d}\sigma = 0, \qquad \int_{\partial D_i} \left(\frac{1}{2}I + \mathcal{K}_D^*\right)[\varphi]\,\mathrm{d}\sigma = \int_{\partial D_i} \varphi\,\mathrm{d}\sigma, \\
&\int_{\partial D_i} \mathcal{K}_{D,2}[\varphi]\,\mathrm{d}\sigma = -\int_{D_i} \mathcal{S}_D[\varphi]\,\mathrm{d}x \quad \text{and} \quad \int_{\partial D_i} \mathcal{K}_{D,3}[\varphi]\,\mathrm{d}\sigma = \frac{\mathrm{i}|D_i|}{4\pi}\int_{\partial D} \varphi\,\mathrm{d}\sigma.
\end{aligned} \tag{2.4}$$

3 Subwavelength scattering decompositions

3.1 Scattering of pure tones

Suppose, first, that the incoming signal is a plane wave parallel to the x_1-axis with angular frequency ω, given by $A\cos(kx_1 - \omega t)$ where $k = \omega/v$. Then, the scattered pressure field is given by $\mathrm{Re}(u(x,\omega)e^{-\mathrm{i}\omega t})$ where u satisfies the Helmholtz equation

$$\begin{cases} (\Delta + k^2)\,u = 0 & \text{in } \mathbb{R}^3 \setminus \overline{D}, \\ (\Delta + k_b^2)\,u = 0 & \text{in } D, \end{cases} \tag{3.1}$$

where $k = \omega/v$ and $k_b = \omega/v_b$, along with the transmission conditions

$$\begin{cases} u_+ - u_- = 0 & \text{on } \partial D, \\ \frac{1}{\rho}\frac{\partial u}{\partial\nu_x}\big|_+ - \frac{1}{\rho_b}\frac{\partial u}{\partial\nu_x}\big|_- = 0 & \text{on } \partial D, \end{cases} \tag{3.2}$$

and the Sommerfeld radiation condition in the far field, which ensures that energy radiates outwards [26], given by

$$\left(\frac{\partial}{\partial|x|} - ik\right)(u - u^{in}) = o(|x|^{-1}) \quad \text{as } |x| \to \infty, \tag{3.3}$$

where, in this case, $u^{in}(x,\omega) = Ae^{\mathrm{i}kx_1}$.

Definition 3.1 (Resonant frequency)**.** *We define a* resonant frequency *to be $\omega = \omega(\delta)$ such that there exists a non-trivial solution to* (3.1) *which satisfies* (3.2) *and* (3.3) *when $u^{in} = 0$.*

Definition 3.2 (Subwavelength resonant frequency)**.** *We define a* subwavelength resonant frequency *to be a resonant frequency ω that depends continuously on δ and is such that $\omega(\delta) \to 0$ as $\delta \to 0$.*

Lemma 3.3. *A system of N subwavelength resonators exhibits N subwavelength resonant frequencies with positive real parts, up to multiplicity.*

Proof. This was proved in [3] and follows from the theory of Gohberg and Sigal [26, 28]. □

3.1.1 Capacitance matrix analysis

Our approach to solving the resonance problem is to study the *(weighted) capacitance matrix*, which offers a rigorous discrete approximation to the differential problem. We will see that the eigenstates of this $N \times N$-matrix characterise, at leading order in δ, the resonant modes of the system.

In order to introduce the notion of capacitance, we define the functions ψ_j, for $j = 1, \dots, N$, as

$$\psi_j := \mathcal{S}_D^{-1}[\chi_{\partial D_j}],$$

where $\chi_A : \mathbb{R}^3 \to \{0, 1\}$ is used to denote the characteristic function of a set $A \subset \mathbb{R}^3$. The capacitance coefficients C_{ij}, for $i, j = 1, \dots, N$, are then defined as

$$C_{ij} := -\int_{\partial D_i} \psi_j \, d\sigma.$$

We will need two objects involving the capacitance coefficients. Firstly, the weighted capacitance matrix $C^{\mathrm{vol}} = (C_{ij}^{\mathrm{vol}})$, given by

$$C_{ij}^{\mathrm{vol}} := \frac{1}{|D_i|} C_{ij},$$

which has been weighted to account for the different sized resonators (see *e.g.* [27, 29, 30] for other variants in slightly different settings). Secondly, we will want the capacitance sums contained in the matrix $C^{\mathrm{sum}} = (C_{ij}^{\mathrm{sum}})$, given by

$$C^{\mathrm{sum}} := JC,$$

where $C = (C_{ij})$ is the matrix of capacitance coefficients and J is the $N \times N$ matrix of ones (*i.e.* $J_{ij} = 1$ for all $i, j = 1, \dots, N$).

We define the functions S_n^ω, for $n = 1 \dots, N$, as

$$S_n^\omega(x) := \begin{cases} \mathcal{S}_D^k[\psi_n](x), & x \in \mathbb{R}^3 \setminus \overline{D}, \\ \mathcal{S}_D^{k_b}[\psi_n](x), & x \in D. \end{cases}$$

Lemma 3.4. *The solution to the scattering problem* (3.1) *can be written, for $x \in \mathbb{R}^3$, as*

$$u(x) - Ae^{ikx_1} = \sum_{n=1}^N q_n S_n^\omega(x) - \mathcal{S}_D^k \left[\mathcal{S}_D^{-1}[Ae^{ikx_1}] \right](x) + O(\omega),$$

for constants q_n which satisfy, up to an error of order $O(\delta\omega + \omega^3)$, the problem

$$\left(\omega^2 I - v_b^2 \delta \, C^{\mathrm{vol}} \right) \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix} = v_b^2 \delta \begin{pmatrix} \frac{1}{|D_1|} \int_{\partial D_1} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, d\sigma \\ \vdots \\ \frac{1}{|D_N|} \int_{\partial D_N} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, d\sigma \end{pmatrix}.$$

Proof. The solutions can be represented as

$$u(x) = \begin{cases} Ae^{ikx_1} + \mathcal{S}_D^k[\psi](x), & x \in \mathbb{R}^3 \setminus \overline{D}, \\ \mathcal{S}_D^{k_b}[\phi](x), & x \in D, \end{cases} \tag{3.4}$$

4

for some surface potentials $(\phi, \psi) \in L^2(\partial D) \times L^2(\partial D)$, which must be chosen so that u satisfies the transmission conditions across ∂D. Using (2.1), we see that in order to satisfy the transmission conditions on ∂D, the densities ϕ and ψ must satisfy

$$\mathcal{S}_D^{k_b}[\phi](x) - \mathcal{S}_D^k[\psi](x) = Ae^{ikx_1}, \qquad x \in \partial D,$$

$$\left(-\frac{1}{2}I + \mathcal{K}_D^{k_b,*}\right)[\phi](x) - \delta\left(\frac{1}{2}I + \mathcal{K}_D^{k,*}\right)[\psi](x) = \delta\frac{\partial}{\partial\nu}(Ae^{ikx_1}), \qquad x \in \partial D.$$

Using the asymptotic expansions (2.2) and (2.3), we can see that

$$\psi = \phi - \mathcal{S}_D^{-1}[Ae^{ikx_1}] + O(\omega),$$

and, further, that

$$\left(-\frac{1}{2}I + \mathcal{K}_D^* + \frac{\omega^2}{v_b^2}\mathcal{K}_{D,2} - \delta\left(\frac{1}{2}I + \mathcal{K}_D^*\right)\right)[\phi] = -\delta\left(\frac{1}{2}I + \mathcal{K}_D^*\right)\mathcal{S}_D^{-1}[Ae^{ikx_1}] + O(\delta\omega + \omega^3). \qquad (3.5)$$

Then, integrating (3.5) over ∂D_i, for $1 \le i \le N$, and using the properties (2.4) gives us that

$$-\omega^2 \int_{D_i} \mathcal{S}_D[\phi]\,dx - v_b^2\delta \int_{\partial D_i} \phi\,d\sigma = -v_b^2\delta \int_{\partial D_i} \mathcal{S}_D^{-1}[Ae^{ikx_1}]\,d\sigma + O(\delta\omega + \omega^3).$$

At leading order, (3.5) says that $\left(-\frac{1}{2}I + \mathcal{K}_D^*\right)[\phi] = 0$ so, in light of the fact that $\{\psi_1, \ldots, \psi_N\}$ forms a basis for $\ker\left(-\frac{1}{2}I + \mathcal{K}_D^*\right)$, the solution can be written as

$$\phi = \sum_{n=1}^N q_n\psi_n + O(\omega^2 + \delta), \qquad (3.6)$$

for constants $q_1, \ldots, q_N = O(1)$. Making this substitution we reach, up to an error of order $O(\delta\omega + \omega^3)$, the problem

$$\left(-\omega^2 I_N + v_b^2\delta C^{\mathrm{vol}}\right)\begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix} = -v_b^2\delta\begin{pmatrix} \frac{1}{|D_1|}\int_{\partial D_1}\mathcal{S}_D^{-1}[Ae^{ikx_1}]\,d\sigma \\ \vdots \\ \frac{1}{|D_N|}\int_{\partial D_N}\mathcal{S}_D^{-1}[Ae^{ikx_1}]\,d\sigma \end{pmatrix}. \qquad (3.7)$$

The result now follows by combining the above. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 3.5. *As $\delta \to 0$, the subwavelength resonant frequencies satisfy the asymptotic formula*

$$\omega_n^{\pm} = \pm\sqrt{v_b^2\lambda_n\delta} - i\tau_n\delta + O(\delta^{3/2}),$$

for $n = 1, \ldots, N$, where λ_n are the eigenvalues of the weighted capacitance matrix C^{vol} and τ_n are real numbers that depend on D, v and v_b.

Proof. If $u^{in} = 0$, we find from Lemma 3.4 that there is a non-zero solution $q_1, \ldots, q_N$ to the eigenvalue problem (3.4) when $\omega^2/v_b^2\delta$ is an eigenvalue of C^{vol}, at leading order.

To find the imaginary part, we adopt the ansatz

$$\omega_n^{\pm} = \pm\sqrt{v_b^2\lambda_n\delta} - i\tau_n\delta + O(\delta^{3/2}). \qquad (3.8)$$

Using the expansions (2.2) and (2.3) with the representation (3.4) we have that

$$\psi = \phi + \frac{k_b - k}{4\pi i}\left(\sum_{n=1}^N \psi_n\right)\int_{\partial D}\phi\,d\sigma + O(\omega^2),$$

and, hence, that

$$\left(-\frac{1}{2}I + \mathcal{K}_D^* + k_b^2\mathcal{K}_{D,2} + k_b^3\mathcal{K}_{D,3} - \delta\left(\frac{1}{2}I + \mathcal{K}_D^*\right)\right)[\phi] - \frac{\delta(k_b - k)}{4\pi i}\left(\sum_{n=1}^N \psi_n\right)\int_{\partial D}\phi\,d\sigma = O(\delta\omega^2 + \omega^4).$$
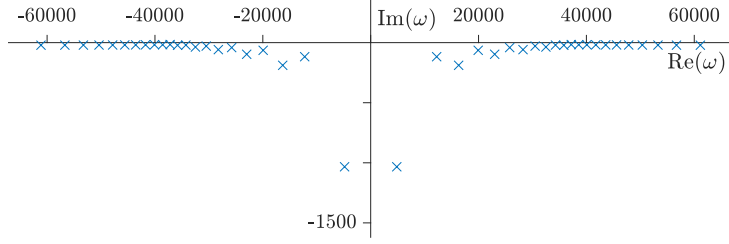
5

Figure 2: The resonant frequencies $\{\omega_n^+, \omega_n^- : n = 1, \ldots, N\}$ in the complex plane, for an array of 22 subwavelength resonators.

We then substitute the decomposition (3.6) and integrate over ∂D_i, for $i = 1, \ldots, N$, to find that

$$\left(-\frac{\omega^2}{v_b^2} I - \frac{\omega^3}{v_b^3} \frac{\mathrm{i}}{4\pi} C^{\mathrm{sum}} + \delta C^{\mathrm{vol}} + \delta\omega \left(\frac{1}{v_b} - \frac{1}{v} \right) \frac{\mathrm{i}}{4\pi} C^{\mathrm{vol}} C^{\mathrm{sum}} \right) \underline{q} = O(\delta\omega^2 + \omega^4).$$

Then, using the ansatz (3.8) for ω_n and setting $\underline{q} = \underline{v}_n$ (the eigenvector corresponding to λ_n) we reach that

$$\left(\tau_n I - \frac{v_b \lambda_n}{8\pi} C^{\mathrm{sum}} + \left(1 - \frac{v_b}{v} \right) \frac{v_b}{8\pi} C^{\mathrm{vol}} C^{\mathrm{sum}} \right) \underline{v}_n = \underline{0}. \tag{3.9}$$

$\square$

Remark 3.6. *The resonant frequencies will have negative imaginary parts, due to the loss of energy (e.g. to the far field), thus $\tau_n \geq 0$ for all $n = 1, \ldots, N$.*

Remark 3.7. *Note that in some cases $\tau_n = 0$ for some n, meaning the imaginary parts exhibit higher-order behaviour in δ. For example, the second (dipole) frequency for a pair of identical resonators is known to be $O(\delta^2)$ [27].*

Remark 3.8. *The numerical simulations presented in this work were all carried out on an array of 22 cylindrical resonators. We approximate the problem by studying the two-dimensional cross section using the multipole expansion method, as in [1].*

Remark 3.9. *The array of 22 resonators that simulated in this work measures 35 mm, has material parameters corresponding to air-filled resonators surrounded by water and has subwavelength resonant frequencies within the range 500 Hz – 10 kHz. Thus, this structure has similar dimensions to the human cochlea, is made from realistic materials and experiences subwavelength resonance in response to frequencies that are audible to humans.*

It is more illustrative to rephrase Lemma 3.4 in terms of basis functions that are associated with the resonant frequencies. Denote by $\underline{v}_n = (v_{1,n}, \ldots, v_{N,n})$ the eigenvector of C^{vol} with eigenvalue λ_n. Then, we have a modal decomposition with coefficients that depend on the matrix $V = (v_{i,j})$, provided the system is such that V is invertible.

Remark 3.10. *The invertibility of V is a subtle issue and depends only on the geometry of the inclusions $D = D_1 \cup \cdots \cup D_N$. In the case that the resonators are all identical, V is invertible since C^{vol} is symmetric. If the size gradient is not too drastic, we expect V to also be invertible (this is supported by our numerical analysis, which typically simulates an array of resonators where each is approximately 1.05 times the size of the previous).*

Lemma 3.11. *Suppose the resonator's geometry is such that the matrix of eigenvectors V is invertible. Then if $\omega = O(\sqrt{\delta})$ the solution to the scattering problem (3.1) can be written, for $x \in \mathbb{R}^3$, as*

$$u(x) - Ae^{ikx_1} = \sum_{n=1}^{N} a_n u_n(x) - \mathcal{S}_D \left[\mathcal{S}_D^{-1}[Ae^{ikx_1}] \right](x) + O(\omega),$$

for constants given by

$$a_n = \frac{-A\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)},$$

where $\nu_n = \sum_{j=1}^{N} [V^{-1}]_{n,j}$, i.e. the sum of the nth row of V^{-1}.

6

Proof. In light of (3.6), we define the functions

$$u_n(x) = \sum_{i=1}^{N} v_{i,n} \, \mathcal{S}_D[\psi_i](x), \tag{3.10}$$

for $n = 1, \dots, N$. Then, by diagonalising C^{vol} with the change of basis matrix V, we see that the solution to the scattering problem (3.1) can be written as

$$u - u^{in} = \sum_{n=1}^{N} a_n u_n - \mathcal{S}_D \left[\mathcal{S}_D^{-1}[Ae^{ikx_1}] \right] + O(\omega),$$

for constants a_n given, at leading order, by

$$V \begin{pmatrix} \omega^2 - v_b^2 \delta \lambda_1 & & \\ & \ddots & \\ & & \omega^2 - v_b^2 \delta \lambda_N \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = v_b^2 \delta \begin{pmatrix} \frac{1}{|D_1|} \int_{\partial D_1} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, \mathrm{d}\sigma \\ \vdots \\ \frac{1}{|D_N|} \int_{\partial D_N} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, \mathrm{d}\sigma \end{pmatrix} + O(\omega^3).$$

Now, $\omega^2 - v_b^2 \delta \lambda_n = (\omega - \omega_n^+)(\omega - \omega_n^-) + O(\omega^3)$ so we have that up to an error of order $O(\omega^3)$

$$\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = v_b^2 \delta \begin{pmatrix} (\omega - \omega_1^+)^{-1}(\omega - \omega_1^-)^{-1} & & \\ & \ddots & \\ & & (\omega - \omega_N^+)^{-1}(\omega - \omega_N^-)^{-1} \end{pmatrix} V^{-1} \begin{pmatrix} \frac{1}{|D_1|} \int_{\partial D_1} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, \mathrm{d}\sigma \\ \vdots \\ \frac{1}{|D_N|} \int_{\partial D_N} \mathcal{S}_D^{-1}[Ae^{ikx_1}] \, \mathrm{d}\sigma \end{pmatrix}.$$

In order to simplify this further, we use the fact that $e^{ikx_1} = 1 + ikx_1 + \cdots = 1 + O(\omega)$ to see that

$$\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \frac{-\operatorname{Re}(\omega_1^+)^2}{(\omega - \omega_1^+)(\omega - \omega_1^-)} & & \\ & \ddots & \\ & & \frac{-\operatorname{Re}(\omega_N^+)^2}{(\omega - \omega_N^+)(\omega - \omega_N^-)} \end{pmatrix} V^{-1} \begin{pmatrix} A \\ \vdots \\ A \end{pmatrix} + O(\omega).$$

$\qquad$ $\square$

3.2 Modal decompositions of signals

Consider, now, the scattering of a more general signal, $s : [0, T] \to \mathbb{R}$, whose frequency support is wider than a single frequency and whose Fourier transform exists. Again, we assume that the radiation is incident parallel to the x_1-axis. Consider the Fourier transform of the incoming pressure wave, given for $\omega \in \mathbb{C}$, $x \in \mathbb{R}^3$ by

$$u^{in}(x, \omega) = \int_{-\infty}^{\infty} s(x_1/v - t) e^{i\omega t} \, \mathrm{d}t$$
$$= e^{i\omega x_1/v} \hat{s}(\omega) = \hat{s}(\omega) + O(\omega),$$

where $\hat{s}(\omega) := \int_{-\infty}^{\infty} s(-u) e^{i\omega u} \, \mathrm{d}u$. The resulting pressure field satisfies the Helmholtz equation (3.1) along with the conditions (3.2) and (3.3).

Working in the frequency domain, the scattered acoustic pressure field u in response to the Fourier transformed signal $\hat{s}$ can be decomposed in the spirit of Lemma 3.11. We write that, for $x \in \partial D$, the solution to the scattering problem is given by

$$u(x, \omega) = \sum_{n=1}^{N} \frac{-\hat{s}(\omega) \nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} u_n(x) + r(x, \omega), \tag{3.11}$$

for some remainder r. We are interested in signals whose energy is mostly concentrated within the subwavelength regime. In particular, we want that

$$\sup_{x \in \mathbb{R}^3} \int_{-\infty}^{\infty} |r(x, \omega)| \, \mathrm{d}\omega = O(\delta). \tag{3.12}$$

7

Remark 3.12. *Note that the condition (3.12) is satisfied e.g. by a pure tone within the subwavelength regime, since if $\omega = O(\sqrt{\delta})$ then Lemma 3.11 gives us that $\sup_x |r| = O(\omega)$.*

Now, we wish to apply the inverse Fourier transform to (3.11) to obtain a time-domain decomposition of the scattered field. From now on we will simplify the notation for the resonant frequencies by assuming we can write that $\omega_n^+ = \omega_n \in \mathbb{C}$ and $\omega_n^- = -\operatorname{Re}(\omega_n) + i\operatorname{Im}(\omega_n)$ (which, by Theorem 3.5, is known to hold at least at leading order in δ).

Theorem 3.13 (Time-domain modal expansion). *For $\delta > 0$ and a signal s which is subwavelength in the sense of the condition (3.12), it holds that the scattered pressure field $p(x,t)$ satisfies, for $x \in \partial D$, $t \in \mathbb{R}$,*

$$p(x,t) = \sum_{n=1}^{N} a_n[s](t) u_n(x) + O(\delta),$$

where the coefficients are given by $a_n[s](t) = (s * h_n)(t)$ for kernels defined as

$$h_n(t) = \begin{cases} 0, & t < 0, \\ c_n e^{\operatorname{Im}(\omega_n)t} \sin(\operatorname{Re}(\omega_n)t), & t \geq 0, \end{cases} \tag{3.13}$$

for $c_n = \nu_n \operatorname{Re}(\omega_n)$.

Proof. Applying the inverse Fourier transform to the modal expansion (3.11) yields

$$p(x,t) = \sum_{n=1}^{N} a_n[s](t) u_n(x) + O(\delta),$$

where, for $n = 1, \ldots, N$, the coefficients are given by

$$a_n[s](t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{-\hat{s}(\omega) \nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega = (s * h_n)(t),$$

where $*$ denotes convolution and the kernels h_n are defined for $n = 1, \ldots, N$ by

$$h_n(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega. \tag{3.14}$$

We can use complex integration to evaluate the integral in (3.14). For $R > 0$, let $\mathcal{C}_R^{\pm}$ be the semicircular arc of radius R in the upper (+) and lower (−) half-plane and let $\mathcal{C}^{\pm}$ be the closed contour $\mathcal{C}^{\pm} = \mathcal{C}_R^{\pm} \cup [-R, R]$. Then, we have that

$$h_n(t) = \frac{1}{2\pi} \oint_{\mathcal{C}^{\pm}} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega - \frac{1}{2\pi} \int_{\mathcal{C}_R^{\pm}} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega.$$

The integral around $\mathcal{C}^{\pm}$ is easy to evaluate using the residue theorem, since it has simple poles at $\omega_n^{\pm}$. We will make the choice of + or − so that the integral along $\mathcal{C}_R^{\pm}$ converges to zero as $R \to \infty$. For large R we have a bound of the form

$$\left| \int_{\mathcal{C}_R^{\pm}} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega \right| \leq C_n R^{-1} \sup_{\omega \in \mathcal{C}_R^{\pm}} e^{\operatorname{Im}(\omega)t}, \tag{3.15}$$

for a positive constant C_n.

Suppose first that $t < 0$. Then we choose to integrate over $\mathcal{C}_R^+$ in the upper complex plane so that (3.15) converges to zero as $R \to \infty$. Thus, we have that

$$h_n(t) = \frac{1}{2\pi} \oint_{\mathcal{C}^+} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega = 0, \qquad t < 0,$$

since the integrand is holomorphic in the upper half plane. Conversely, if $t \geq 0$ then we should choose to integrate over $\mathcal{C}_R^-$ in order for (3.15) to disappear. Then, we see that

$$h_n(t) = \frac{1}{2\pi} \oint_{\mathcal{C}^-} \frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t} \, \mathrm{d}\omega$$

$$= i \operatorname{Res}\left(\frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t}, \omega_n^+ \right) + i \operatorname{Res}\left(\frac{-\nu_n \operatorname{Re}(\omega_n^+)^2}{(\omega - \omega_n^+)(\omega - \omega_n^-)} e^{-i\omega t}, \omega_n^- \right), \qquad t \geq 0.$$
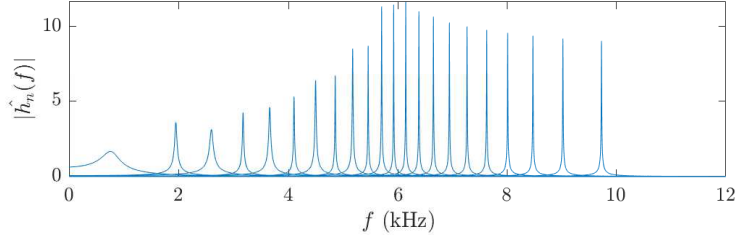
8

Figure 3: The frequency support of the band-pass filters h_n. Shown here for the case of 22 resonators.

Using the notation $\omega_n^+ = \omega_n$, $\omega_n^- = -\operatorname{Re}(\omega_n) + \mathrm{i}\operatorname{Im}(\omega_n)$ we can simplify the expressions for the residues at the two simple poles to reach the result. □

Remark 3.14. *The fact that $h_n(t) = 0$ for $t < 0$ ensures the causality of the modal expansion in Theorem 3.13.*

Remark 3.15. *The asymmetry of the eigenmodes $u_n(x)$ means that the decomposition from Theorem 3.13 replicates the cochlea's famous travelling wave behaviour. That is, in response to an incoming wave the position of maximum amplitude moves from left to right in the array, see [3] for details.*

4 Subwavelength scattering transforms

In Section 3 we showed that when a subwavelength (*i.e.* audible) sound is scattered by a cochlea-mimetic array of resonators the resulting pressure field is described by a model decomposition. This decomposition takes the form of convolutions with the basis functions h_n from Theorem 3.13. Since $\operatorname{Im}(\omega_n) < 0$, each h_n is a windowed oscillatory mode that acts as a band pass filter centred at $\operatorname{Re}(\omega_n)$. We wish to explore the extent to which these decompositions reveal useful properties of the sound and can be used as a basis for signal processing applications.

In order to reveal richer properties of the sound, a common approach is to use the filters h_n in a convolutional neural network. That is, a repeating cascade of alternating convolutions with h_n and some activation function Θ:

$$
\begin{aligned}
a_{n_1}^{(1)}[s](t) &= \Theta\left(s * h_{n_1}\right)(t), \\
a_{n_1,n_2}^{(2)}[s](t) &= \Theta\left(a_{n_1}^{(1)}[s] * h_{n_2}\right)(t), \\
&\;\;\vdots \\
a_{n_1,\ldots,n_k}^{(k)}[s](t) &= \Theta\left(a_{n_1,\ldots,n_{k-1}}^{(k-1)}[s] * h_{n_k}\right)(t),
\end{aligned}
\tag{4.1}
$$

where, in each case, the indices are such that $(n_1, n_2, \ldots, n_k) \in \{1, \ldots, N\}^k$. We will use the notation $P_k = (n_1, \ldots, n_k)$ from now on, and refer to the vector P_k as the *path* of $a_{P_k}^{(k)}$.

4.1 Example: identity activation

As an expository example, we consider the case where $\Theta : \mathbb{R} \to \mathbb{R}$ is the identity $Id(x) = x$. In this case, for any depth k we have that $a_{P_k}^{(k)}[s] = s * h_{P_k}^{(k)}$ for some function $h_{P_k}^{(k)}$ which is the convolution of k functions of the form (3.13), indexed by the path P_k. This simplification means that a more detailed mathematical analysis is possible.

The basis functions $h_{P_k}^{(k)}$ take specific forms. In particular, the diagonal terms contain *gammatones*. A gammatone is a sinusoidal mode windowed by a gamma distribution:

$$
g(t; m, \omega, \phi) = t^{m-1} e^{\operatorname{Im}(\omega)t} \cos(\operatorname{Re}(\omega)t - \phi), \qquad t \geq 0,
$$

for some order $m \in \mathbb{N}^+$ and constants $\omega \in \{z \in \mathbb{C} : \operatorname{Im}(z) < 0\}$, $\phi \in \mathbb{R}$. Gammatones have been widely used to model auditory filters [10]. We notice that $h_n(t) = c_n g(t; 1, \omega_n, \pi/2)$ and that higher order gammatones emerge at deeper levels in the cascade (4.1).
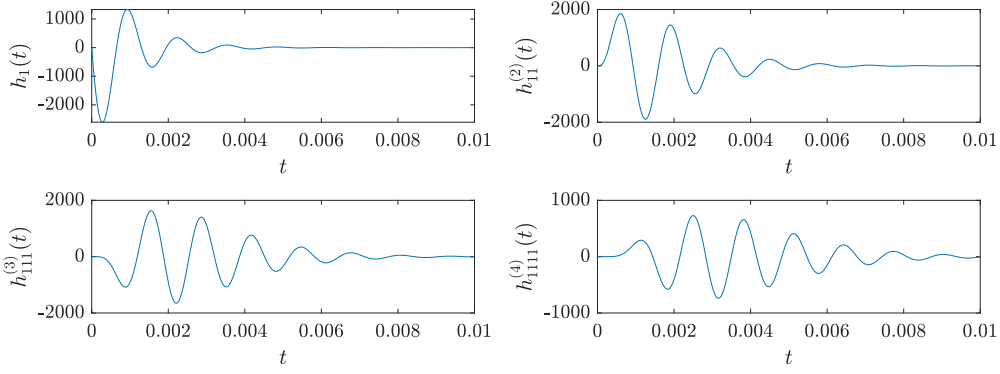
Figure 4: The emergence of gammatones at successively deeper layers in the cascade, shown for the first subwavelength resonant frequency in the case of 22 resonators.

Lemma 4.1 (The emergence of higher-order gammatones)**.** *For $k \in \mathbb{N}^+$ and $n \in \{1, \ldots, N\}$, there exist non-negative constants $C_m^{n,k}$, $m = 1, \ldots, k$, such that*

$$h_{n,\ldots,n}^{(k)}(t) = (c_n)^k \sum_{m=1}^{k} C_m^{n,k} g(t; m, \omega_n, m\tfrac{\pi}{2}).$$

In particular, $C_k^{n,k} \neq 0$.

Proof. Let us write $G_n^m(t) := g(t; m, \omega_n, m\tfrac{\pi}{2})$, for the sake of brevity. Firstly, it holds that $h_n(t) = c_n G_n^1(t)$. Furthermore, we have that

$$(G_n^1 * G_n^1)(t) = \frac{1}{2} G_n^2(t) + \frac{1}{2\,\mathrm{Re}(\omega_n)} G_n^1(t),$$

as well as, for $m \geq 3$, the recursion relation

$$(G_n^{m-1} * G_n^1)(t) = \frac{1}{2(m-1)} G_n^m(t) + \frac{m-2}{2\,\mathrm{Re}(\omega_n)}(G_n^{m-2} * G_n^1)(t).$$

The result follows by repeatedly applying this formula. In particular, we find that

$$C_k^{n,k} = \frac{1}{2^{k-1}(k-1)!} > 0.$$

$\square$

Remark 4.2. *While the gammatones appeared here through the cascade of filters, gammatones also arise directly from resonator scattering if higher-order resonators are used: resonators that exhibit higher-order singularities in the frequency domain [10,31]. It was recently shown that if sources of energy gain and loss are introduced to an array of coupled subwavelength resonators then such higher-order resonant modes can exist [30].*

Remark 4.3. *Since the imaginary part of the lowest frequency is much larger than the others (see Figure 2), h_1 acts somewhat as a low-pass filter (see Figure 3).*

For any depth $k \in \mathbb{N}$ and path $P_k \in \{1, \ldots, N\}^k$ it holds that $h_{P_k}^{(k)} \in L^\infty(\mathbb{R})$ meaning that if $s \in L^1(\mathbb{R})$ then $a_{P_k}^{(k)}[s] \in L^\infty(\mathbb{R})$. If, moreover, s is compactly supported then the decay properties of $h_{P_k}^{(k)}$ mean that $a_{P_k}^{(k)}[s] \in L^p(\mathbb{R})$ for any $p \in [1, \infty]$. Further, we have the following lemmas which characterise the continuity and stability of $s \mapsto a_{P_k}^{(k)}[s]$.

Lemma 4.4 (Continuity of representation)**.** *Consider the network coefficients given by (4.1) with Θ being the identity. Given $k_{max} \in \mathbb{N}^+$, there exists a positive constant C_1 such that for any depth $k = 1, \ldots, k_{max}$, any path $P_k \in \{1, \ldots, N\}^k$ and any signal $s \in L^1(\mathbb{R})$ it holds that*

$$\|a_{P_k}^{(k)}[s_1] - a_{P_k}^{(k)}[s_2]\|_\infty \leq C_1 \|s_1 - s_2\|_1.$$

10

Proof. It holds that

$$C_1 := \sup_{k \in \{1,\dots,k_{max}\}} \sup_{P_k \in \{1,\dots,N\}^k} \sup_{x \in \mathbb{R}} (1-c) \left| h_{P_k}^{(k)}(x) \right| < \infty.$$

Then, the result follows from the fact that

$$\left| a_{P_k}^{(k)}[s_1](t) - a_{P_k}^{(k)}[s_2](t) \right| \le \int_{-\infty}^{\infty} |s_1(u) - s_2(u)| \left| h_{P_k}^{(k)}(t-u) \right| \, \mathrm{d}u.$$

$\square$

Remark 4.5. *The continuity property proved in Lemma 4.4 implies, in particular, that the representation of a signal s is stable with respect to additive noise.*

Lemma 4.6 (Pointwise stability to time warping). *Consider the network coefficients given by (4.1) with Θ being the identity. For $\tau \in C^0(\mathbb{R}; \mathbb{R})$, let T_τ be the associated time warping operator, given by $T_\tau f(t) = f(t + \tau(t))$. Then, given $k_{max} \in \mathbb{N}^+$ there exists a positive constant C_2 such that for any depth $k = 1, \dots, k_{max}$, any path $P_k \in \{1, \dots, N\}^k$ and any signal $s \in L^1(\mathbb{R})$ it holds that*

$$\left\| a_{P_k}^{(k)}[s] - a_{P_k}^{(k)}[T_\tau s] \right\|_\infty \le C_2 \|s\|_1 \|\tau\|_\infty.$$

Proof. Let $(h_{P_k}^{(k)})'$ denote the first derivative of $h_{P_k}^{(k)}$ (which is zero on $(-\infty, 0)$ and does not exist at 0). Then, we see that

$$C_1 := \sup_{k \in \{1,\dots,k_{max}\}} \sup_{P_k \in \{1,\dots,N\}^k} \sup_{x \in (0,\infty)} \left| (h_{P_k}^{(k)})'(x) \right| < \infty,$$

and, by the mean value theorem, that for $t \in \mathbb{R}$

$$\left| h_{P_k}^{(k)}(t - \tau(t)) - h_{P_k}^{(k)}(t) \right| \le C_1 |\tau(t)|.$$

Thus, we see that for any $t \in \mathbb{R}$

$$|a_{P_k}^{(k)}[s] - a_{P_k}^{(k)}[T_\tau s]| \le \int_{-\infty}^{\infty} |s(t-u)| \left| h_{P_k}^{(k)}(u) - h_{P_k}^{(k)}(u - \tau(u)) \right| \, \mathrm{d}u,$$

$$\le C_1 \|\tau(u)\|_\infty \int_{-\infty}^{\infty} |s(t-u)| \, \mathrm{d}u.$$

$\square$

A common approach to extracting information from the coefficients (4.1) is to use their temporal averages. A particular advantage of such an approach is that it gives outputs that are invariant to translation and time-dilation (*cf.* the scattering transform [32,33]). Let $\langle a_{P_k}^{(k)}[s] \rangle_{(t_1,t_2)}$ denote the average of $a_{P_k}^{(k)}[s](t)$ over the interval (t_1, t_2), given by

$$\langle a_{P_k}^{(k)}[s] \rangle_{(t_1,t_2)} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} a_{P_k}^{(k)}[s](t) \, \mathrm{d}t. \tag{4.2}$$

Lemma 4.7 (Stability of averages to time warping). *Consider the network coefficients given by (4.1) with Θ being the identity. For $\tau \in C^1(\mathbb{R}; \mathbb{R})$, let T_τ be the associated time warping operator, given by $T_\tau f(t) = f(t + \tau(t))$. Suppose that τ is such that $\|\tau'\|_\infty < \frac{1}{2}$. Then, given $k_{max} \in \mathbb{N}^+$ there exists a positive constant C_2 such that for any depth $k = 1, \dots, k_{max}$, any path $P_k \in \{1, \dots, N\}^k$ and any signal $s \in L^1(\mathbb{R})$ it holds that*

$$\left| \langle a_{P_k}^{(k)}[s] \rangle_{(t_1,t_2)} - \langle a_{P_k}^{(k)}[T_\tau s] \rangle_{(t_1,t_2)} \right| \le C_2 \|s\|_1 \left(\frac{2}{t_2 - t_1} \|\tau\|_\infty + \|\tau'\|_\infty \right).$$
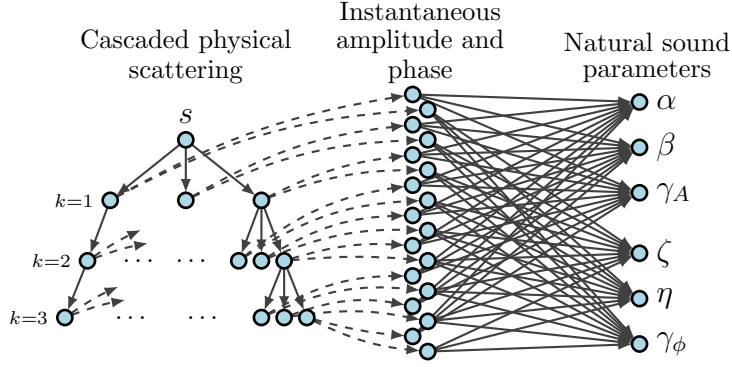
Figure 5: The architecture considered in this work cascades the physically-derived subwavelength scattering, extracts the instantaneous amplitude and phase before, finally, estimating the parameters of the associated natural sound distributions.

Proof. Since $\|\tau'\|_\infty \leq c < 1$, $\varphi(t) = t - \tau(t)$ is invertible and $\|\varphi'\|_\infty \geq 1 - c$,

$$\int_{t_1}^{t_2} \left(h_{P_k}^{(k)}(t - \tau(t)) - h_{P_k}^{(k)}(t) \right) \, \mathrm{d}t = \int_{\varphi(t_1)}^{\varphi(t_2)} h_{P_k}^{(k)}(t) \frac{1}{\varphi'(\varphi^{-1}(t))} \, \mathrm{d}t - \int_{t_1}^{t_2} h_{P_k}^{(k)}(t) \, \mathrm{d}t$$

$$= \int_{I_1 - I_2} h_{P_k}^{(k)}(t) \frac{1}{\varphi'(\varphi^{-1}(t))} \, \mathrm{d}t + \int_{t_1}^{t_2} h_{P_k}^{(k)}(t) \frac{\tau'(\varphi^{-1}(t))}{\varphi'(\varphi^{-1}(t))} \, \mathrm{d}t,$$

for some intervals $I_1, I_2 \subset \mathbb{R}$, each of which has length bounded by $\|\tau\|_\infty$. Now, define the constant

$$C_2 := \sup_{k \in \{1, \ldots, k_{max}\}} \sup_{P_k \in \{1, \ldots, N\}^k} \sup_{x \in (0, \infty)} (1 - c) \left| h_{P_k}^{(k)}(x) \right| < \infty.$$

Finally, we can compute that

$$\langle a_{P_k}^{(k)}[s] \rangle_{(t_1, t_2)} - \langle a_{P_k}^{(k)}[T_\tau s] \rangle_{(t_1, t_2)} = \frac{1}{t_2 - t_1} \int_{-\infty}^{\infty} s(u) \int_{t_1}^{t_2} \left(h_{P_k}^{(k)}(t - u - \tau(t)) - h_{P_k}^{(k)}(t - u) \right) \, \mathrm{d}t \, \mathrm{d}u$$

$$= \frac{1}{t_2 - t_1} \int_{-\infty}^{\infty} s(u) \left(\int_{I_1 - I_2} h_{P_k}^{(k)}(t - u) \frac{1}{\varphi'(\varphi^{-1}(t - u))} \, \mathrm{d}t + \int_{t_1}^{t_2} h_{P_k}^{(k)}(t - u) \frac{\tau'(\varphi^{-1}(t - u))}{\varphi'(\varphi^{-1}(t - u))} \, \mathrm{d}t \right) \, \mathrm{d}u,$$

meaning that

$$\left| \langle a_{P_k}^{(k)}[s] \rangle_{(t_1, t_2)} - \langle a_{P_k}^{(k)}[T_\tau s] \rangle_{(t_1, t_2)} \right| \leq \frac{1}{t_2 - t_1} \|s\|_1 \left[2\|\tau\|_\infty C_2 + (t_2 - t_1) C_2 \|\tau'\|_\infty \right].$$

$\square$

Remark 4.8. *Lemma 4.7 shows that temporal averages are approximately invariant to translations if the length of the window is large relative to the size of the translation (i.e. if $t_2 - t_1 \gg \|\tau\|_\infty$).*

5 Representation of natural sounds

Extracting meaning from the representation of a sound, beyond elementary statements about which tones are more prolific at different stages, is difficult. In this section, we propose a novel approach tailored to the class of natural sounds which exploits their observed statistical properties.

5.1 Properties of natural sounds

Let us briefly summarise what has been observed about the low-order statistics of natural sounds [22–25]. For a sound $s(t)$, let $a_\omega(t)$ be the component at frequency ω (obtained *e.g.* through the application of a band-pass filter centred at ω). Then we can write that

$$a_\omega(t) = A_\omega(t) \cos(\omega t + \phi_\omega(t)),$$

where $A_\omega(t) \geq 0$ and $\phi_\omega(t)$ are the instantaneous amplitude and phase, respectively. We view $A_\omega(t)$ and $\phi_\omega(t)$ as stochastic processes and wish to understand their statistics.

It has been widely observed that several properties of the frequency components of natural sounds vary according to the inverse of the frequency. In particular, it is well known that the power spectrum (the square of the Fourier transform) of the amplitude satisfies a relationship of the form

$$S_{A_\omega}(f) = |\hat{A}_\omega(f)|^2 \propto \frac{1}{f^\gamma}, \qquad 0 < f < f_{max}, \tag{5.1}$$

for a positive parameter γ (which often lies in a neighbourhood of 1) and some maximum frequency f_{max}. Further, this property is independent of the frequency band that is studied [24].

Consider the log-amplitude, $\log_{10} A_\omega(t)$. It has been observed that for a variety of natural sounds (including speech, animal vocalisations, music and environmental sounds) the log-amplitude is locally stationary. Suppose we normalise the log-amplitude so that it has zero mean and unit variance, giving a quantity that is invariant to amplitude scaling. Then, the normalised log-amplitude averaged over some time interval $[t_1, t_2]$ has a distribution of the form [25]

$$p_A(x) = \beta \exp(\beta x - \alpha - e^{\beta x - \alpha}), \tag{5.2}$$

where α and β are real-valued parameters and $\beta > 0$. Further, this property is scale invariant in the sense that it is true irrespective of the scale over which the temporal average is taken. It also known that the curves for different frequency bands fall on top of one another, meaning that p_A does not depend on ω (the frequency band).

Further, the power spectrum S_{ϕ_ω} of the instantaneous phase also satisfies a $1/f$-type relationship, of the same form as (5.1). On the other hand, the instantaneous phase is non-stationary (even locally), making it difficult to describe through the above methods. A more tractable quantity is the instantaneous frequency (IF), defined as

$$\lambda_\omega = \frac{\mathrm{d}\phi_\omega}{\mathrm{d}t}.$$

It has been observed that $\lambda_\omega(t)$ is locally stationary for natural sounds and the temporal mean of its modulus satisfies a distribution p_λ of the form [24]

$$p_\lambda(x) \propto (\zeta^2 + x^2)^{-\eta/2}, \tag{5.3}$$

for positive parameters ζ and $\eta > 1$.

5.2 Representation algorithm

For a given natural sound, we wish to find the parameters that characterise its global properties, according to (5.1)–(5.3). Given a signal s we first compute the convolution with the band-pass filter h_n to yield the spectral component at the frequency $\mathrm{Re}(\omega_n)$, given by

$$a_n[s](t) = A_n(t) \cos(\mathrm{Re}(\omega_n)t + \phi_n(t)).$$

We extract the functions A_n and ϕ_n from $a_n[s]$ using the Hilbert transform [24, 34, 35]. In particular, we have that

$$a_n[s](t) + \mathrm{i}H(a_n[s])(t) = a_n[s](t) + \frac{\mathrm{i}}{\pi} \int_{-\infty}^{\infty} \frac{a_n[s](u)}{t - u} \, \mathrm{d}u = A_n(t) e^{\mathrm{i}(\mathrm{Re}(\omega_n)t + \phi_n(t))},$$

from which we can extract A_n and ϕ_n by taking the complex modulus and argument, respectively.

Remark 5.1. *It is not obvious that the Hilbert transform $H(a_n[s])$ is well-defined. Indeed, we must formally take the principal value of the integral. For a signal that is integrable and has finite support, $H(a_n[s])(t)$ exists for almost all $t \in \mathbb{R}$.*

Given the functions A_n and ϕ_n, the power spectra $S_{A_n}(f)$ and $S_{\phi_n}(f)$ can be computed by applying the Fourier transform and squaring. We estimate the relationships of the form (5.1) by first averaging the N power spectra, to give $\overline{S}_A(f) := \frac{1}{N} \sum_n S_{A_n}(f)$ and $\overline{S}_\phi(f) := \frac{1}{N} \sum_n S_{\phi_n}(f)$ before fitting curves $f^{-\gamma_A}$

(a) A histogram of the normalised time-averaged log-amplitude, fitted to a distribution of the form (5.2).

(b) A histogram of the normalised time-averaged instantaneous frequency, fitted to a distribution of the form (5.3).

(c) The averaged (over different frequency bands) power spectra of the instantaneous amplitude, fitted to a distribution of the form (5.1).

(d) The averaged (over different frequency bands) power spectra of the instantaneous phase, fitted to a distribution of the form (5.1).
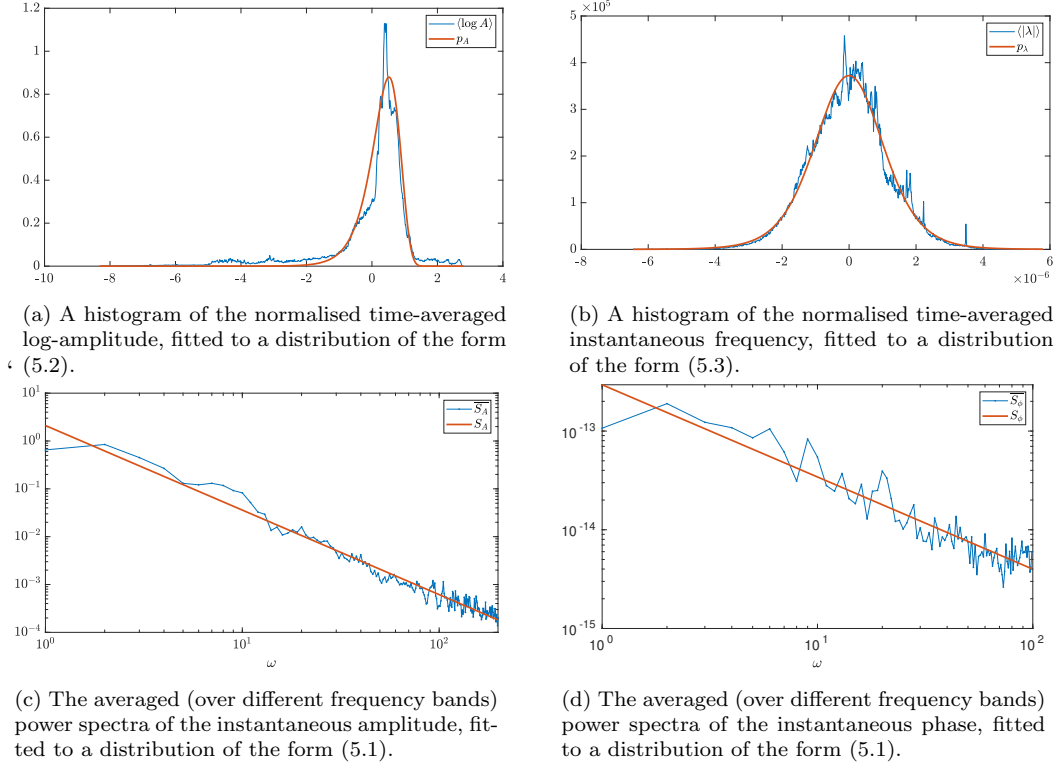
Figure 6: The fitted distributions for a trumpet playing a single note.

and $f^{-\gamma_\phi}$ using least-squares regression. We estimate the parameters of the probability distributions (5.2) and (5.3) by normalising both $\log_{10} A_n(t)$ and $\lambda_n(t)$ so that

$$\langle \log_{10} A_n \rangle = 0, \qquad \langle (\log_{10} A_n)^2 \rangle = 1,$$

and similarly for $\lambda_n(t)$, before repeatedly averaging the normalised functions over intervals $[t_1, t_2] \subset \mathbb{R}$. Curves of the form (5.2) and (5.3) are then fitted to the resulting histograms (which combine the temporal averages from different filters $n = 1, \ldots, N$ and different time intervals $[t_1, t_2]$) using non-linear least-squares optimisation.

5.3 Discussion

The observations of Section 5.1 give us six coefficients $(\gamma_A, \alpha, \beta, \gamma_\phi, \zeta, \eta) \in \mathbb{R}^6$ that portray global properties of a natural sound. Table 1 shows some examples of these parameters, estimated using the approach described in Section 5.2. Our hypothesis is that these parameters capture, in some sense, the quality of

	trumpet	violin	cello	thunder	baby speech	adult speech	running water	crow call
γ_A	1.767	1.563	1.528	1.415	1.763	1.808	1.466	1.571
α	1.244	0.375	0.284	0.474	0.517	0.528	0.336	0.649
β	2.390	0.783	0.841	0.596	0.747	0.894	0.484	0.896
γ_ϕ	0.763	0.871	0.6977	0.446	1.192	1.125	1.088	0.908
ζ $(\times 10^{-6})$	2.878	3.433	6.1149	6.322	4.773	5.176	5.200	4.212
η	8.579	11.824	8.679	8.315	9.660	9.358	9.290	10.475

Table 1: Values of the estimated distribution parameters for different samples of natural sounds.

a signal. Thus, incorporating these parameters into the representation of a signal, alongside *e.g.* temporal averages (4.2), will improve the 'perceptual' abilities of any classification algorithm based on this representation. The space of natural sounds that is characterised by the six parameters is likely to have a highly non-trivial (and non-Euclidean) structure which will need to be learnt from data, *cf.* [36, 37].

6   Concluding remarks

We have studied an array of subwavelength resonators that has similar dimensions to the cochlea and mimics its biomechanical properties. We proved, from first principles, that the pressure field scattered by this structure satisfies a modal expansion with spatial eigenmodes and gammatone time dependence. We, then, explored how these basis functions could be used as kernels in a convolutional signal processing routine. In particular, we proposed an algorithm for extracting meaningful global properties from the band-pass coefficients tailored to the class of natural sounds.

An advantage of two-step approach (physical scattering followed by neural processing) is that the subtleties of the auditory system can be readily incorporated, *e.g.* the non-linear amplification that takes place in the cochlea [38]. This work studied linear representations of signals followed by a non-linear algorithm for extracting the natural sound parameters. While analyses of non-linear networks (*i.e.* with the activation function different from the identity) have been conducted in other settings [32, 33], an amplification mechanism based on a compressive non-linearity can be incorporated directly into the resonator-array model, as studied in [1, 2].

Acknowledgements

References

[1] H. Ammari and B. Davies. Mimicking the active cochlea with a fluid-coupled array of subwavelength Hopf resonators. *Proc. R. Soc. A*, 476(2234):20190870, 2020.

[2] M. Rupin, G. Lerosey, J. de Rosny, and F. Lemoult. Mimicking the cochlea with an active acoustic metamaterial. *New J. Phys.*, 21:093012, 2019.

[3] H. Ammari and B. Davies. A fully-coupled subwavelength resonance approach to filtering auditory signals. *Proc. R. Soc. A*, 475(2228):20190049, 2019.

[4] T. Duke and F. Jülicher. Critical oscillators as active elements in hearing. In G. A. Manley, R. R. Fay, and A. N. Popper, editors, *Active Processes and Otoacoustic Emissions in Hearing*, pages 63–92. Springer, New York, 2008.

[5] B. S. Joyce and P. A. Tarazaga. Developing an active artificial hair cell using nonlinear feedback control. *Smart Mater. Struct.*, 24(9):094004, 2015.

[6] B. S. Joyce and P. A. Tarazaga. Mimicking the cochlear amplifier in a cantilever beam using nonlinear velocity feedback control. *Smart Mater. Struct.*, 23(7):075019, 2014.

[7] C. F. Babbs. Quantitative reappraisal of the Helmholtz-Guyton resonance theory of frequency tuning in the cochlea. *J. Biophys.*, 2011:1–16, 2011.

[8] V. Leroy, A. Bretagne, M. Fink, H. Willaime, P. Tabeling, and A. Tourin. Design and characterization of bubble phononic crystals. *Appl. Phys. Lett.*, 95(17):171904, 2009.

[9] V. Leroy, A. Strybulevych, M. Scanlon, and J. Page. Transmission of ultrasound through a single layer of bubbles. *Eur. Phys. J. E*, 29(1):123–130, 2009.

---

[10] R. F. Lyon. *Human and machine hearing*. Cambridge University Press, 2017.

[11] J. F. Alm and J. S. Walker. Time-frequency analysis of musical instruments. *SIAM Rev.*, 44(3):457–476, 2002.

[12] L. Cohen. *Time-frequency analysis*. Prentice Hall, Englewood Cliffs, NJ, 1995.

[13] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, 1998.

[14] I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, 1992.

[15] X. Yang, K. Wang, and S. A. Shamma. Auditory representations of acoustic signals. *IEEE T. Inform. Theory*, 38(2):824–839, 1992.

[16] J. J. Benedetto and A. Teolis. A wavelet auditory model and data compression. *Appl. Comput. Harmon. A.*, 1(1):3–28, 1993.

[17] J. Andén and S. Mallat. Deep scattering spectrum. *IEEE T. Signal Proces.*, 62(16):4114–4128, 2014.

[18] E. C. Smith and M. S. Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.

[19] M. J. Hewitt and R. Meddis. A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *J. Acoust. Soc. Am.*, 95(4):2145–2159, 1994.

[20] R. D. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. APU report 2341: An efficient auditory filterbank based on the gammatone function. Applied Psychology Unit, Cambridge, 1988.

[21] A. Bell and H. P. Wit. Cochlear impulse responses resolved into sets of gammatones: the case for beating of closely spaced local resonances. *PeerJ*, 6:e6016, 2018.

[22] F. E. Theunissen and J. E. Elie. Neural processing of natural sounds. *Nat. Rev. Neurosci.*, 15(6):355–366, 2014.

[23] R. F. Voss and J. Clarke. '1/f noise' in music and speech. *Nature*, 258(5533):317–318, 1975.

[24] H. Attias and C. E. Schreiner. Temporal low-order statistics of natural sounds. In M. C. Mozer, M. Jordan, and T. Petsche, editors, *Advances in neural information processing systems 9*, pages 27–33, Cambridge, MA, 1997. MIT Press.

[25] H. Attias and C. E. Schreiner. Coding of naturalistic stimuli by auditory midbrain neurons. In M. C. Mozer, M. Jordan, M. Kearns, and S. S, editors, *Advances in neural information processing systems 10*, pages 103–109, Cambridge, MA, 1998. MIT Press.

[26] H. Ammari, B. Fitzpatrick, H. Kang, M. Ruiz, S. Yu, and H. Zhang. *Mathematical and computational methods in photonics and phononics*, volume 235 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, 2018.

[27] H. Ammari, B. Fitzpatrick, H. Lee, S. Yu, and H. Zhang. Double-negative acoustic metamaterials. *Q. Appl. Math.*, 77(4):767–791, 2019.

[28] I. Gohberg and J. Leiterer. *Holomorphic operator functions of one variable and applications: methods from complex analysis in several variables*, volume 192 of *Operator Theory Advances and Applications*. Birkhäuser, Basel, 2009.

[29] H. Ammari, B. Davies, and S. Yu. Close-to-touching acoustic subwavelength resonators: eigenfrequency separation and gradient blow-up. *arXiv preprint arXiv:2001.04888*, 2020.

[30] H. Ammari, B. Davies, H. Lee, E. O. Hiltunen, and S. Yu. Exceptional points in parity–time-symmetric subwavelength metamaterials. *arXiv preprint arXiv:2003.07796*, 2020.

[31] S. T. Neely and D. Kim. A model for active elements in cochlear biomechanics. *J. Acoust. Soc. Am.*, 79(5):1472–1480, 1986.

[32] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1872–1886, 2013.

[33] S. Mallat. Group invariant scattering. *Comm. Pure Appl. Math.*, 65(10):1331–1398, 2012.

[34] J. Flanagan. Parametric coding of speech spectra. *J. Acoust. Soc. Am.*, 68(2):412–419, 1980.

[35] B. Boashash. Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals. *Proc. IEEE*, 80(4):520–538, 1992.

[36] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proc. computer vision and pattern recognition, IEEE*, pages 652–660, 2017.

[37] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017.

[38] A. Hudspeth. Making an effort to listen: mechanical amplification in the ear. *Neuron*, 59(4):530–545, 2008.