Implementing the Fourier transform in a sensor: a benchmark application for neuromorphic acoustic sensing

IOP Publishing Neuromorph. Comput. Eng. 5 (2025) 034007 https://doi.org/10.1088/2634-4386/ade7ac

NEUROMORPHIC
Computing and Engineering

OPEN ACCESS

RECEIVED
23 October 2024

REVISED
3 June 2025

ACCEPTED FOR PUBLICATION
24 June 2025

PUBLISHED
24 July 2025

PAPER

Implementing the Fourier transform in a sensor: a benchmark application for neuromorphic acoustic sensing

Hermann Folke Johann Rolf[1,*], **Petro Feketa**[2], **Alexander Schaum**[3] and **Thomas Meurer**[1]

[1] Digital Process Engineering Group, Institute of Mechanical Process Engineering and Mechanics (MVM), Karlsruhe Institute of Technology, Karlsruhe, Germany
[2] School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand
[3] Department of Process Analytics, Computational Science Hub (CSH), University Hohenheim, Stuttgart, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: folke.rolf@kit.edu

Keywords: neuromorphic acoustic sensing, Andronov–Hopf bifurcation, frequency tunability, envelope model, sampling

Abstract

In changing environments, the mammalian hearing perception so far outperforms technical speech processing. This is enabled by the nonlinear dynamics of the cochlea. Inside of it, the processing is based on a frequency decomposition and a sophisticated feedback loop, so that the amplification of spectrum of an external signal can vary from linear to compressive. This behavior can be mimicked by implementing a controllable Andronov–Hopf bifurcation into neuromorphic oscillators, which enables a compressive and frequency-selective response. However, the frequency decomposition with these neuromorphic oscillators has not been investigated yet. Here, we show that any oscillator, which exhibits an Andronov–Hopf bifurcation, has a unique response to external stimuli, if its bifurcation parameter is in a neighborhood of the critical point. In addition, we propose three different algorithms to enable the frequency decomposition by implementing the Fourier transform in an acoustic sensor. We found that this Fourier transform can be done by applying amplitude demodulation to the output of any oscillator exhibiting at least one Andronov–Hopf bifurcation and investigated the convergence time of the different algorithms. Our results demonstrate that the Fourier transform can be utilized for either a single oscillator, which is simple to implement, or an array of oscillators, which has a fast convergence time. Thereby, it is shown that the neuromorphic acoustic sensor consisting of these oscillators can both mimic the processing of the cochlea and be implemented into the technical unit. Finally, a potential experimental implementation using micro-electromechanical system sensors is proposed.

1. Introduction

Traditionally, speech processing is structured in a feedforward architecture, which involves three stages [1]. First, the acoustic stimuli is measured by a microphone, which has a linear transfer function [2]. Second, in the pre-processing step the measured signal is transformed into the frequency domain and undergoes nonlinear filtering. Third, the pre-processed signal is classified (and thus recognized) by artificial neural networks [1, 3]. In particular, the nonlinearities for pre-processing and recognition enabled the rapid progression in speech processing in recent years. For instance, it has been demonstrated in [3] that the word success rate, i.e. the number of successfully recognized words divided by the total number of words, can be substantially increased by choosing a compressive nonlinearity in the pre-processing stage. This nonlinearity is characterized by its response to the signal amplitude. If the amplitude is small, it is amplified. Contrary, the amplification is attenuated, if the amplitude is large. Technological implementations of this nonlinearity are motivated by the mammalian hearing perception and are, e.g. the cochleagram [1, 4, 5], the Gammatone filter [6–8], the Mel-frequency cepstral coefficient [1, 3, 9], and in-materia computing units such as the combination of dopant network processing units and an in-memory computing chip consisting of memristive devices [10].

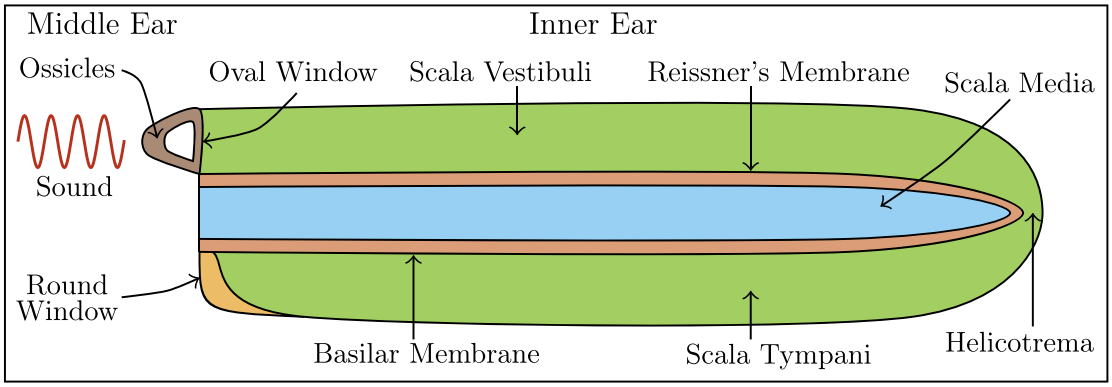

Figure 1. Sketch of the uncoiled cochlea. It consists of three tubes—the scala vestibuli, scala media and the scala tympani. These tubes are separated by Reissner's membrane and the basilar membrane. The basilar membrane is in-homogeneous over the space, so that the frequency components of an acoustic stimuli are separated over space.

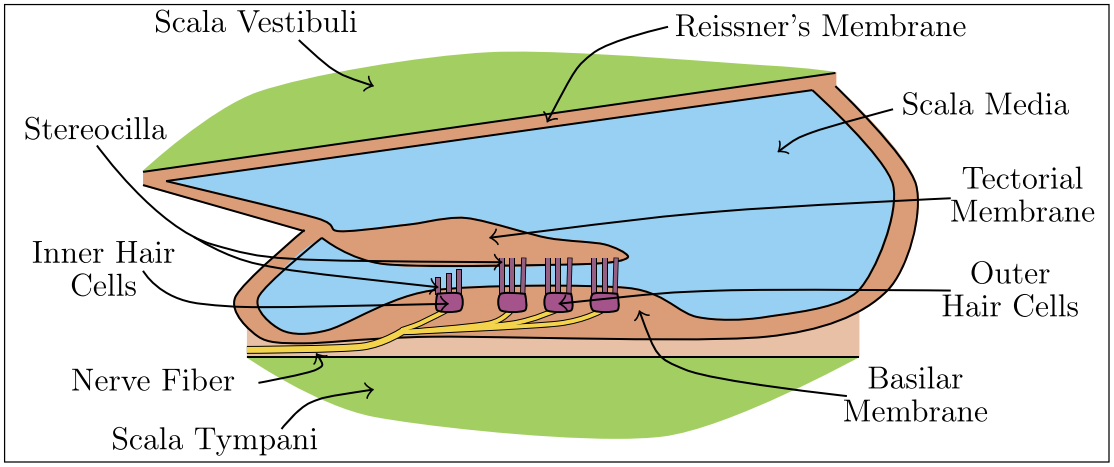

Figure 2. Sketch of the cross-section of the cochlea. The acoustic stimuli is transduced into an electrical signal by the inner hair cells. Their response can be controlled by an feedback loop from the nervous system to the outer hair cells and the tectorial membrane.

Similar to technological solutions, the mammalian hearing perception shows a compressive nonlinearity. However, the overall structure is fundamentally different since pre-processing and sensing is done simultaneously. This is enabled by a feedback loop in the cochlea, which can be actively adapted to varying stimuli [11, 12]. Hence, the dynamical range of the mammalian hearing perception is significantly larger than (technological) speech processing, i.e. faint sound can be detected in weakly distorted environments, while heavily distorted signal can also be reconstructed. The latter is called the cocktail party effect [13, 14]. Additionally, by exploiting the feedback loop its efficiency in terms of time and energy consumption is tremendously increased in comparison to technological speech processing. Subsequently, the anatomy and physiology of the cochlea are summarized. For further details, the reader is referred to [15, 16].

The sensing and pre-processing is done inside the cochlea, which is a snail shell structure within the inner ear. Once uncoiled and laid out straight, the cochlea consists of three tubes: the scala vestibuli, the scala media, and the scala tympani (see figure 1). At the base of the cochlea, there are the oval and round window and the scala vestibuli and the scala tympani are connected over the helicotrema at its apex. In contrast to this, the scala media is separated from the other tubes by Reissner's membrane and the basilar membrane. On these membranes, there are ion pumps, which transport potassium ion K^+ into the scala media, so that there is potential difference between the scala media and the other tubes.

Moreover, hair cells are scattered on the basilar membrane inside the scala media (see figure 2). These cells are negatively charged, they have hair-like outgrowths, which are called stereocilla. These cells are also connected to the nervous system and they can be distinguished into the outer hair cells and the inner hair cells. On the one hand, the outer hair cells are grouped into a V-shape, the tip of their stereocilla is connected to an additional membrane, the so-called tectorial membrane, and there are motor neurons within these hair cells. On the other hand, the stereocilla of the inner hair cells are arranged in a staircase and they have a free

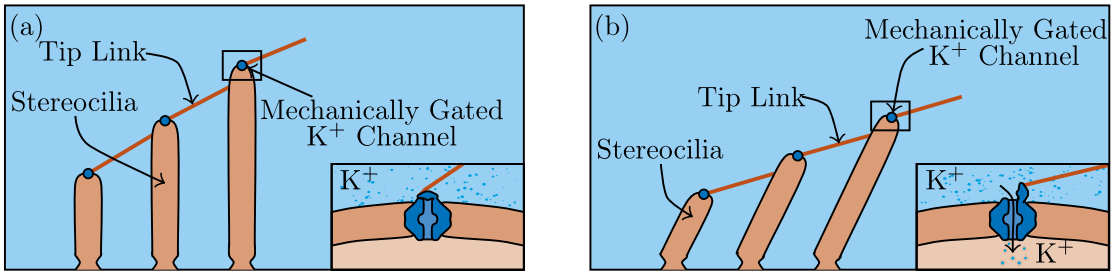

Figure 3. Close-up of the stereocilla of an inner hair cell. (a) The stereocilla are not bent towards their larger neighbor. Hence, the mechanically gated K^+ channel is not opened by the tip link and the charge of the inner hair cell remains constant (and negative). (b) The stereocilla are bent towards its larger neighbor. Thus, the tip link opens the mechanically gated K^+ channel and the inner hair cell is depolarized, so that a neuron in the cochlea is excited.

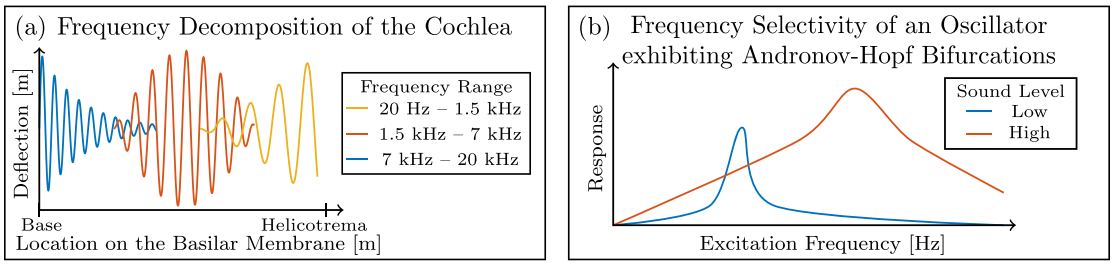

Figure 4. Visualization of (a) the spatial frequency decomposition in the cochlea and (b) the frequency decomposition over time done by an (frequency tunable) oscillator.

tip. Additionally, there is a mechanically gated K^+ channel on the top of these cells, which is connected by the tip link to its larger neighbor (see figure 3).

An acoustic stimuli is processed in the cochlea as follows: after the stimuli has been transferred from the middle ear into the inner ear by the ossicles, it travels through the scala vestibuli. As the basilar membrane has a varying width and stiffness, low frequencies cross the basilar membrane close to the helicotrema, while high frequencies pass the basilar membrane close its base (see figure 4(a)). This can be interpreted as a Fourier transform over space. The frequency components excite the hair cells, so that the stereocilla are oscillating. The mechanically gated K^+ channel are then closed, if the stereocilla are bent towards their smaller neighbors (see figure 3(a)). In contrast, if the stereocilla are bent towards their larger neighbors, the tip link opens the mechanically gated K^+ channel and the hair cell is depolarized resulting in an electrical signal (see figure 3(b)). To control the inner hair cells, the motor neurons of the outer hair cells are exploited.

To model the interplay the physiology, two different methods can be applied. First, a black box model can be followed, where the transfer function of the cochlea is based on the inverse correlation of the response of the neurons to an external stimuli [6]. This approach leads to the Gammatone filter and the cochleagramm. These models can be easily utilized in the pre-processing stage of speech processing, since they can be implemented by a digital filter [1, 3, 17]. Secondly, the complete dynamics of this system can be modeled by oscillators exhibiting Andronov–Hopf bifurcations [18–21]. To determine a bifurcation, a bifurcation parameter has to be identified. With this parameter, the dynamics of the system can be changed, when the parameter surpasses a certain threshold, the so-called critical point. For an Andronov–Hopf bifurcation, a system can exhibit two different regimes: First, the equilibrium of the system is (locally) asymptotically stable, if the bifurcation parameter has not surpassed the critical point. In this case the system is in the sub-critical regime. Second, if the bifurcation surpasses the critical point, a (potentially unstable) limit cycle emerges in the system. This regime is called super-critical [22–24]. To show that this bifurcation emerges, the system is usually compared with the so-called normal form of the Andronov–Hopf bifurcation by employing the Normal Form Theorem (or theorems based on it) [22–24]. It is worth noting that the properties of a system below its critical point of an Andronov–Hopf bifurcation are similar to the properties of the response of the cochlea. Therefore, a neuromorphic acoustic sensor mimicking the functionality of the cochlea can be realized by exploiting oscillators, which exhibit an Andronov–Hopf bifurcation. For instance, potential candidates showing such behavior are and thermally actuated micro-electromechanical systems (MEMSs), see, e.g. [25–28].

The scope of this work is to propose an application for a neuromorphic acoustic sensor, which mimics the functionality of the cochlea. For this, the frequency decomposition done by the basilar membrane is imitated by oscillators, which exhibit at least one Andronov–Hopf bifurcation, so that an in-sensor Fourier transform is enabled. In contrast to the frequency decomposition of the basilar membrane, this is a frequency decomposition over time and it is motivated by the frequency selectivity, i.e. the oscillator only reacts to a small frequency band in the frequency domain (see figure 4(b)). Hence, the frequency response satisfies approximately the sampling property[4] of the Dirac impulse and this concept is thus called sampling of the frequency domain.

To implement the in-sensor Fourier transform, two major problems have to be resolved. First, it has to be shown that the steady state of the excited oscillators is unique as the nonlinearities in systems can induce multiple (attractive) steady states, which leads to chaotic motion [22]. Hence, uniqueness of the steady state implies that the frequency response of the external stimuli can be uniquely reconstructed by the response of the oscillators, which are excited by this external stimuli. Second, the convergence time[5] of a neuromorphic acoustic sensor has to be investigated. If the convergence time is small enough, the external stimuli can be treated as a stationary signal, so that the oscillators lock onto the frequency components of the signal. However, the convergence time is increased by moving the bifurcation parameter closer to the critical point of the system, as this will move a pair of complex-conjugated eigenvalues of the linearized system closer to the imaginary axis [22, 23]. Hence, if the bifurcation parameter is in a small neighborhood of the critical point, the convergences time can be deduced by this pair of eigenvalues and it can be too slow for some acoustic sensing applications such as speech recognition. In particular, the convergence time for pre-processing are constrained by the vowel time, which can be as fast as 10 ms. Thus, the time frames of automatic speech recognition is usually around 20–30 ms [30].

In this paper, these two problems are resolved. This is done by analyzing the steady state of the Andronov–Hopf oscillator under an external stimuli. This oscillator is described by the (super-critical) normal form of the Andronov–Hopf bifurcation. Hence, any system exhibiting at least a super-critical Andronov–Hopf bifurcation converges to a sub-manifold[6], which can be transformed by an isomorphism into the normal form of the Andronov–Hopf bifurcation, if its bifurcation parameter is in a neighborhood of the critical point [22–24]. Therefore, the Andronov–Hopf oscillator can be viewed as a benchmark model for neuromorphic acoustic sensors. In addition, the convergence time of two different designs of neuromorphic acoustic sensors is analyzed. The in-sensor Fourier transform can be implemented by either an array of oscillators with constant frequency or a frequency tunable oscillator, i.e. the characteristic frequency of this oscillator can be adapted by a controllable input [32, 33]. In the latter approach, the frequency tunable oscillator covers a frequency range by changing the characteristic frequency. By assuming that the convergence time of the oscillator is faster than the variation of the external stimuli, different sampling algorithms can be designed, since the system can then treat the external stimuli as stationary signal. Hence, the oscillator can sample multiple frequencies and the sequence of sampled frequencies can be assigned arbitrary. With this, a benchmark for sampling the frequency domain can be defined and different algorithms for the in-sensor Fourier transform can be investigated. In particular, the premise of reinforcement learning is enables an efficient implementation.

The paper is structured as follows: the notations are briefly introduced in section 2. In section 3, the principles of a neuromorphic acoustic sensor are summarized by introducing the benchmark oscillator model and an approach to design of a neuromorphic acoustic sensor. In particular, the benchmark oscillator model is the Andronov–Hopf oscillator. This is followed by formally defining the concept of sampling in the frequency domain in section 4 and it is shown that by sampling the Fourier coefficients with the Andronov–Hopf oscillator, the frequency response can be reconstructed uniquely. Then algorithms are proposed to sample the frequency domain in section 5. The first algorithm is based on an array of oscillators. While, the other algorithms are based sequential sampling or sampling based on reinforcement learning with one frequency tunable oscillator. A numerical study of these algorithms is done in section 6 and the consequences of these algorithms and some remarks conclude the paper.

[4] The sampling property of the Dirac impulse refers to measuring one specific value of a function. In the frequency domain, it is given by $f(\omega)\delta_0(\omega - a) = f(a)\delta_0(\omega)$ with a function $f(\omega) \in \mathbb{C}$ and the Dirac impulse $\delta_0(\omega)$ [29].

[5] Convergence time is the time until the system is converged to a (small) neighborhood of an equilibrium.

[6] For the purpose of the paper, a sub-manifold $\mathcal{N}$ is a subset of an $n \in \mathbb{N}$ dimensional Euclidean vector space, which can be described locally by a (different) Euclidean vector space [31]. A simple example for a sub-manifold is the surface of a two-dimensional sphere with radius $r = 1$ embedded in a three-dimensional Euclidean space, i.e. $S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 \,|\, \|\boldsymbol{x}\|_2 = 1\} \subset \mathbb{R}^3$.

2. Notation

The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ stand for the sets of natural numbers, integers, real numbers, and complex numbers, respectively. Other sets are denoted by capital, calligraphic letters $\mathcal{A}$. We use small letters $a \in \mathbb{R}$ for variables. Vectors and matrices are denoted by bold, small letters $\boldsymbol{b} \in \mathbb{R}^n$ and capital letters $C \in \mathbb{R}^{n\times n}$, respectively. Given functions $f : \mathbb{R} \to \mathbb{R}$ we use $f(t) \in \mathbb{R}$ to denote the function value at given t. For functions $g : \mathbb{Z} \to \mathbb{R}$ defined over the integers we $g[k] \in \mathbb{R}$ to refer to their value. Finally, the mixed notation $h[k](t)$ is used for scalar fields given by $h : \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$.

3. Brief summary on neuromorphic acoustic sensing

To mimic the cochlea, a neuromorphic acoustic sensor is subsequently assumed to consists of an array of oscillators. This comes form the fact that the dynamics of the cochlea can be modeled by such a system [21]. These oscillators are assumed to exhibit at least one Andronov–Hopf bifurcation, so that this acoustic sensor mimics the functionality of the cochlea and it has the same properties as to the mathematical model of its physiology. Subsequently, a benchmark oscillator model for this setup, the functionality of the neuromorphic acoustic sensor, and two different approaches to implement it are discussed.

3.1. A benchmark oscillator model for neuromorphic acoustic sensing

To describe a neuromorphic acoustic sensor, it is assumed that the oscillators inside a neuromorphic acoustic sensor can be modeled by the normal form of the Andronov–Hopf bifurcation. This can be done since the cochlea can be modeled by oscillators, which exhibit this bifurcation [18–21], and any system exhibiting an Andronov–Hopf bifurcation converges to a sub-manifold that can be transformed by an isomorphism into the normal form of the Andronov–Hopf bifurcation, if the bifurcation parameter is in a neighborhood of the critical point [22–24]. Classically, this normal form is described in real-valued coordinates

$$\dot{\boldsymbol{x}} = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix} \boldsymbol{x} - \eta \|\boldsymbol{x}\|_2^2 \boldsymbol{x} + \beta \boldsymbol{v}, \qquad t > 0, \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{1}$$

with the state vector $\boldsymbol{x}(t) \in \mathbb{R}^2$, the 2-norm $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2}$, and an (unknown) input $\boldsymbol{v}(t) \in \mathbb{R}^2$. The parameters are given by the damping $\mu \in \mathbb{R}$, the characteristic (angular) frequency $\omega \geqslant 0$, the cubic damping $\eta \in \mathbb{R}$, the input constant $\beta \in \mathbb{R}$, and the initial conditions $\boldsymbol{x}_0 \in \mathbb{R}^2$. To simplify the derivations, (1) is transformed into complex coordinates. This transformation is done by denoting the complex-valued state and the complex-valued input by $z = x_1 + \mathrm{i}x_2 \in \mathbb{C}$ and $u = v_1 + \mathrm{i}v_2 \in \mathbb{C}$. Taking the derivative of the state z and sorting terms, yields the normal form of the Andronov–Hopf bifurcation in complex coordinates, which is given by

$$\dot{z} = (\mu + \mathrm{i}\omega) z - \eta |z|^2 z + \beta u, \qquad t > 0, \qquad z(0) = z_0. \tag{2}$$

Note that the dynamics of the unexcited system (2), i.e. $u \equiv 0$, fundamentally changes by varying the damping μ. Subsequently, the emerging dynamics from an Andronov–Hopf bifurcation is recalled. If $\mu \leqslant 0$, (2) is in the sub-critical regime and it has an asymptotically stable equilibrium $z_{\mathrm{eq}} = 0$, i.e. the system (2) converges to z_{eq}. If $\mu > 0$, the system is in the super-critical regime and a (potentially unstable) limit cycle emerges, i.e. the system (2) starts to oscillate, if the cubic damping satisfies $\eta > 0$. This change in dynamics is called bifurcation and the damping μ, which induces the change in the qualitative behavior of the solutions of (2), is denoted as bifurcation parameter. In addition, the point, where the dynamics change from asymptotically stable to oscillatory, is called critical point. For (2) the critical point is given by $\mu_{\mathrm{c}} = 0\,\frac{1}{\mathrm{s}}$ and the sub-critical and super-critical regimes are illustrated in figure 5, respectively. It should be noted that any system exhibiting an Andronov–Hopf bifurcation will converge to a sub-manifold, which can be transformed by some isomorphism, so that this system can be described by (2), if the bifurcation parameter is in a neighborhood of the critical point [22–24]. As a consequence the subsequent results can be generalized to any system exhibiting an Andronov–Hopf bifurcation, if the system can be transformed into the normal form. Hence, (2) can be seen as a benchmark oscillator model for neuromorphic acoustic sensing.

For sensor applications, an emerging limit cycle is a distortion and it will jam the external stimuli u. Therefore, system (2) should stay in the sub-critical regime for sensor applications. In this regime, (2) reacts in two different ways: On the one hand, (2) reacts linearly with respect to amplitude and phase, if the bifurcation parameter μ has a large distance to the critical point μ_{c}, i.e. $\mu \ll \mu_{\mathrm{c}}$. On the other hand, (2) reacts frequency-selective and compressive, if the bifurcation parameter μ is sufficiently close to the bifurcation point μ_{c}, i.e. μ fulfills $\mu \lessapprox \mu_{\mathrm{c}}$. We recall a compressive system amplifies small amplitudes and attenuates strong amplitudes, while frequency selectivity implies that the oscillator reacts to a (narrow) frequency band.

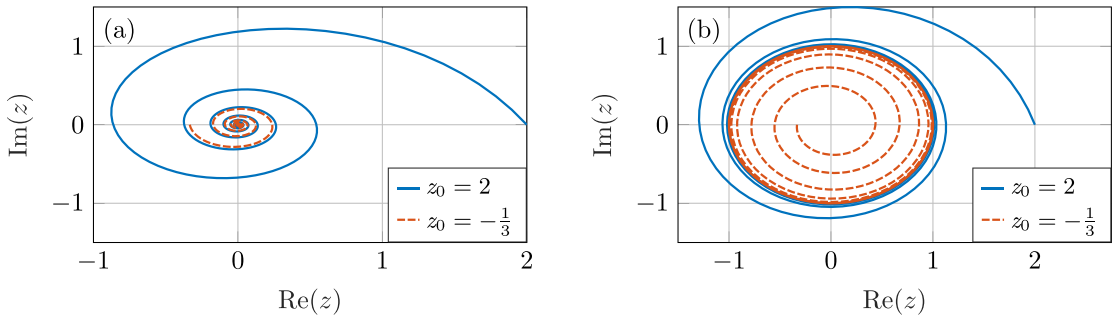

Figure 5. Phase plane $(\mathrm{Re}(z), \mathrm{Im}(z))$ of the normal form of the Andronov–Hopf bifurcation given by (2) in the sub-critical and super-critical regime for different initial conditions. The characteristic frequency and the cubic damping are given by $\omega = 10\,\frac{1}{\mathrm{s}}$ and $\eta = 1\,\frac{1}{\mathrm{s}}$. (a) Trajectories $(\mathrm{Re}(z), \mathrm{Im}(z))$ of oscillator (2) in the sub-critical regime for $\mu = -1\,\frac{1}{\mathrm{s}}$. (b) Trajectories $(\mathrm{Re}(z), \mathrm{Im}(z))$ of oscillator (2) in super-critical regime for $\mu = 1\,\frac{1}{\mathrm{s}}$.

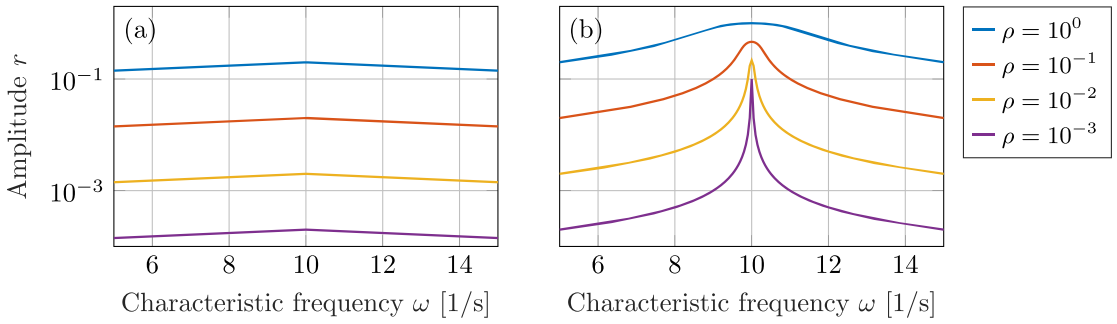

Figure 6. Linear and nonlinear response of system (2) as a function of the characteristic frequency ω and the external amplitude ρ with the cubic damping $\eta = 1\,\frac{1}{\mathrm{s}}$ and the input constant $\beta = 1\,\frac{1}{\mathrm{s}}$. (a) Linear response of system (2). Here, the damping is given by $\mu = -5\,\frac{1}{\mathrm{s}}$. (b) Nonlinear compressive, frequency-selective response of system (2). For this, the damping reads $\mu = 0\,\frac{1}{\mathrm{s}}$.

These regions in the sub-critical regime can be illustrated by analyzing the influence of a harmonic stimuli $u = \rho \mathrm{e}^{\mathrm{i}(\nu t + \phi)}$ that excites the system (2). The parameters of the harmonic stimuli are given by the amplitude $\rho > 0$, the phase $\phi \in [0, 2\pi)$, and the external frequency $\nu > 0$. By assuming that (2) is 1:1 frequency-locked[7] on this stimuli, its solution is given by $z = r\mathrm{e}^{\mathrm{i}(\nu t + \varphi)}$ with the amplitude $r > 0$ and the phase $\varphi \in [0, 2\pi)$. Inserting this assumption into (2) and dividing by the harmonic $\mathrm{e}^{\mathrm{i}\nu t}$, yields

$$0 = [\mu + \mathrm{i}(\omega - \nu)]\, r\mathrm{e}^{\mathrm{i}\varphi} - \eta r^3 \mathrm{e}^{\mathrm{i}\varphi} + \beta \rho \mathrm{e}^{\mathrm{i}\phi}. \tag{3}$$

Equation (3) can be further simplified by subtracting $\rho \mathrm{e}^{\mathrm{i}\phi}$ and taking the absolute value squared of the result. This results in

$$0 = \eta^2 r^6 - 2\mu\eta r^4 + \left[\mu^2 + (\omega - \nu)^2\right] r^2 - \beta\rho^2, \tag{4}$$

which can be easily solved for r numerically. This is subsequently done by assuming that the amplitude r is a function of the characteristic frequency $\omega \in [5, 15]\,\frac{1}{\mathrm{s}}$[8]. To visualize the properties of the linear and nonlinear regimes, the numerical parameters are given by the damping $\mu \in \{-5, 0\}\,\frac{1}{\mathrm{s}}$, the amplitude $\rho \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}\,\frac{1}{\mathrm{s}}$, the phase $\varphi = 0$, and the frequency $\nu = 10\,\frac{1}{\mathrm{s}}$ of the external stimuli. The linear response and frequency-selective, compressive response of (2) are illustrated in figure 6, respectively. System (2) is frequency-selective, if the bifurcation parameter μ is in a neighborhood of the critical point μ_c[9], so that it will only react to excitation with its characteristic frequency. By exploiting this property, (2) fulfills

[7] A detailed discussing on this assumption in given in appendix A.

[8] The response of (2) is invariant to the frequencies ω and ν since $(\omega - \nu)^2 = (\nu - \omega)^2$ holds.

[9] Here, equation (2) with $\mu = 0\,\frac{1}{\mathrm{s}}$ is only close to the critical point μ_c, if the amplitude ρ is small enough. This comes from the fact that the critical point μ_c is moved by the amplitude of the external excitation. Indeed, this effect is one of the underlying mechanisms for the compressive behavior of an Andronov–Hopf bifurcation.

the sampling property in the frequency domain, i.e. the amplitude is approximately a Dirac impulse, which reads

$$\delta_0(\omega-\nu)=\begin{cases}\infty, & \text{if } \omega=\nu\\ 0, & \text{else}\end{cases}. \tag{5}$$

In addition, it should be noted that amplification of harmonic inputs with small amplitude is large at the characteristic frequency of (2), e.g. the amplitude of (2) is given by $r=10^{-1}$ for an amplitude $\rho=10^{-3}\,\frac{1}{\mathrm{s}}$. While, amplification of harmonic inputs with large amplitude is small at the characteristic frequency of (2), e.g. the amplitude of (2) is given by $r=1$ for an amplitude $\rho=10^{0}\,\frac{1}{\mathrm{s}}$. This is called compression and it has to be stressed that combining both properties can improve the performance and efficiency for speech recognition [3, 34].

3.2. Functionality and design of neuromorphic acoustic sensors

By exploiting the frequency selectivity, an array of oscillators can be used to sample the frequency domain. With this, an in-sensor Fourier transform is enabled and the functionality of the cochlea is mimicked by using the compressive nonlinearity. In addition to the inherent properties of an Andronov–Hopf bifurcation, it is useful to introduce the notion of frequency tunability. Here, an oscillator is called frequency tunable, if its characteristic frequency can be adjusted by a controllable input [33, 35]. For instance, tunability is induced by controlling a delay feedback [35, 36], exploiting the asymmetry between coupled oscillators [33, 37], or using a nonlinearity to change the local dynamics [32]. Based on these preliminaries, there are two different implementations for a neuromorphic acoustic sensor.

First, the neuromorphic acoustic sensor is an array consisting of a large number of untunable oscillators, which have different, constant characteristic frequencies and exhibit at least one Andronov–Hopf bifurcation. This approach will have a fast convergence time. However, the number of these systems might be limited physically by space. Second, by asserting that the oscillators are frequency tunable, the number of oscillators decreases, as each of these oscillators can be used to sample a frequency interval. However, the convergence time increases, since these oscillators have to react to multiple frequencies. Thus, quickly varying inputs might not be measurable.

To implement a neuromorphic acoustic sensor in hardware, an array consisting of thermally actuated, cantilevered MEMSs sensors can be used. This approach is sketch in figure 7. In principle, these sensors are a mechanical structure composed of a silicon layer as a fundament, a siliconoxid layer for isolation, and 1 μm thick aluminum wires for electric actuation. In addition, there is a piezoelectric strain gauge at the base of the MEMS sensor, so that the deflection can be deduced from the measured strain [26, 38, 39]. The deflection is then filtered by a high pass and further processed by a field-programmable gate array (FPGA). Controllable Andronov–Hopf bifurcations can then be induced in this setup, e.g. by using proportional feedback [26, 28, 32, 40], by coupling the MEMS sensors [33, 41], or exploiting a controllable delayed feedback [36]. As these MEMS sensors are cantilevers, their characteristic frequency is constant for the proportional feedback [28, 33, 42]. Hence, to enable frequency tunability either the geometry of the MEMS sensor has to be changed [32, 43] or different feedback mechanisms can be used [33, 36, 41].

4. Sampling using an Andronov–Hopf bifurcation

In the following, sampling of the frequency domain is formulated mathematically. We aim to show that the frequency response of (2) to an unknown external stimuli u can be uniquely reconstructed by evaluating the steady state of (2). For this, system (2) is rewritten as a (time dependent) Fourier series and it is assumed that the unknown stimuli u can be treated as a stationary signal.

To analyze the temporal evolution of the Fourier series of an ODE, the so-called envelope model has to be employed [44–46]. This is done by rewriting the state z and the external stimuli u by a truncated Fourier series

$$z=\sum_{k=0}^{n} q_k \mathrm{e}^{\mathrm{i}k\nu_\Delta t} \tag{6a}$$

$$u=\sum_{k=0}^{n} F_k \mathrm{e}^{\mathrm{i}k\nu_\Delta t} \tag{6b}$$

with the Fourier coefficients $q_k(t)\in\mathbb{C}$, $F_k(t)\in\mathbb{C}$, the fundamental frequency ν_Δ and the number of Fourier coefficients $n\in\mathbb{N}$ for all $k=0,1,\ldots,n$. The time evolution of the Fourier coefficients q_k follows by

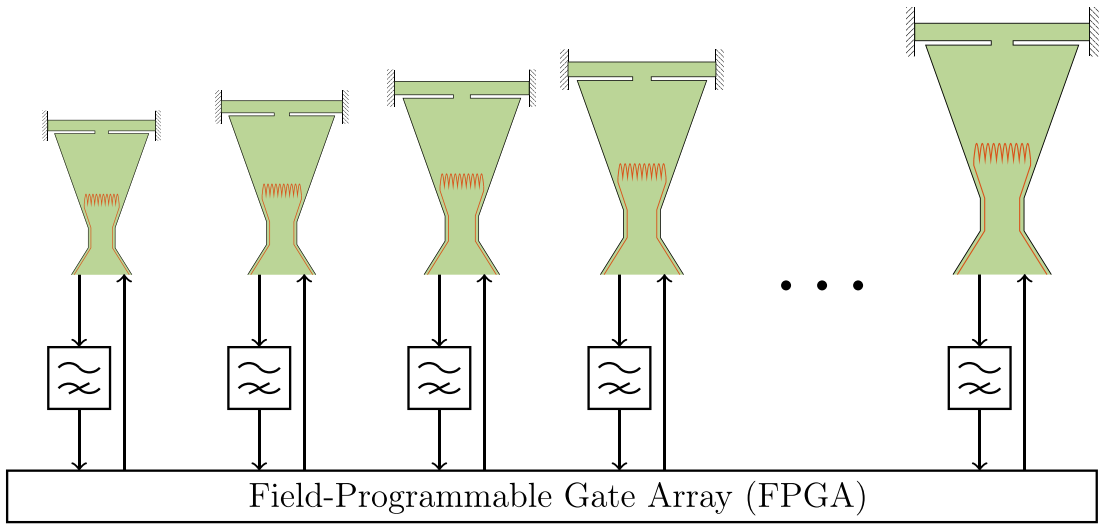

Figure 7. Sketch of an oscillator-based neuromorphic acoustic sensor. It consists of MEMS sensors, which exhibit Andronov–Hopf bifurcations. Indicated by the increasing size of the MEMS sensor, each MEMS sensor has a different characteristic frequency. To control these bifurcations, feedback mechanisms are implemented in a field-programmable gate array (FPGA). The MEMS sensor has been designed by partners from Ilmenau and Kiel University. For further details, the reader is referred to [26, 39, 43].

substitution of (6) into (2) and sorting terms in $\mathrm{e}^{\mathrm{i}k\nu_\Delta t}$ for all $k = 0, 1, \ldots, n$. This yields

$$\dot{\boldsymbol{q}} = A\boldsymbol{q} - \eta \boldsymbol{f}(\boldsymbol{q}) + \beta \boldsymbol{F} \tag{7}$$

with the vector of the Fourier coefficients $\boldsymbol{q} = [q_0, q_1, \cdots, q_n]^{\mathrm{T}} \in \mathbb{C}^{n+1}$, $\boldsymbol{F} = [F_0, F_1, \ldots, F_n]^{\mathrm{T}} \in \mathbb{C}^{n+1}$, and the vector-valued nonlinearity $\boldsymbol{f}(\boldsymbol{q}) \in \mathbb{C}^{n+1}$ for all $k \in \{0, 1 \ldots, n\}$. Herein, the system matrix reads

$$A = \begin{bmatrix} \mu + \mathrm{i}\omega & 0 & \cdots & 0 \\ 0 & \mu + \mathrm{i}(\omega - \nu_\Delta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu + \mathrm{i}(\omega - n\nu_\Delta) \end{bmatrix} \in \mathbb{R}^{(n+1)\cdot(n+1)}. \tag{8}$$

Equation (7) is the so-called envelope model of (2) and its Fourier coefficients $\boldsymbol{q}$ are coupled by the nonlinearity $\boldsymbol{f}$[10]. Therefore, the nonlinearity $\boldsymbol{f}$ has to be determined to analyze the response of the system (2) to an external input u. This is done by comparing the frequency components $\mathrm{e}^{\mathrm{i}k\nu_\Delta t}$ of

$$|z|^2 z = \left(\sum_{k=0}^{n} q_k \mathrm{e}^{\mathrm{i}k\nu_\Delta t}\right)\left(\sum_{k=0}^{n} q_k \mathrm{e}^{\mathrm{i}k\nu_\Delta t}\right)\left(\sum_{k=0}^{n} q_k^* \mathrm{e}^{-\mathrm{i}k\nu_\Delta t}\right), \tag{9}$$

where q_k^* denotes the complex conjugate of the Fourier coefficients q_k for all $k = 0, 1 \ldots, n$. Equation (9) is simplified, if (2) is frequency-selective, since frequency selectivity implies that the coupling between the Fourier coefficients vanishes, i.e. $q_i q_j q_k^* \approx 0$ for all $i, j, k \in \{0, 1, \ldots n\}$ and $(i \neq j) \vee (i \neq k) \vee (j \neq k)$. Hence, by asserting $\mu \lessapprox \mu_\mathrm{c}$, the nonlinearity (9) can be approximated by

$$|z|^2 z \approx \sum_{k=0}^{n} \left(|q_k|^2 q_k \mathrm{e}^{\mathrm{i}k\nu_\Delta}\right). \tag{10}$$

Therefore, if the bifurcation parameter μ is in a neighborhood of the critical point μ_c, the Fourier coefficients q_k can be analyzed individually and (7) reads

$$\dot{q}_k = [\mu + \mathrm{i}(\omega - \nu_\Delta)]\, q_k - \eta |q_k|^2 q_k + \beta F_k \tag{11}$$

[10] It follows from this that equilibrium of Fourier coefficients of a linear system will unique, since linearity implies that the nonlinearity vanishes, i.e. $\boldsymbol{f} \equiv 0$. Hence, a linear, frequency-selective system can be used to sample the frequency domain.

for all $k = 0, 1, \ldots, n$. With this, the Fourier coefficients $\boldsymbol{F}$ of the external signal u can be determined the steady state of (11), i.e. $\dot{q}_k = 0$. This results in

$$0 = [\mu + \mathrm{i}(\omega - \nu_\Delta)]\, q_{\mathrm{eq},k} - \eta |q_{\mathrm{eq},k}|^2 q_{\mathrm{eq},k} + \beta F_{\mathrm{eq},k} \tag{12}$$

for all $k = 0, 1, \ldots, n$. Equation (12) might have multiple solutions in terms of the Fourier coefficients $q_{\mathrm{eq},k}$. Hence, the number of solutions of (12) is subsequently investigated. Transforming (12) into polar coordinates $q_{\mathrm{eq},k} = r_{\mathrm{eq},k} \mathrm{e}^{\mathrm{i}\varphi_{\mathrm{eq},k}}$ and $F_k = \rho_{\mathrm{ex},k} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{ex},k}}$ with amplitudes $r_{\mathrm{eq},k}, \rho_{\mathrm{ex},k} \geq 0$ and phases $\varphi_{\mathrm{eq},k}, \phi_{\mathrm{ex},k} \in [0, 2\pi)$ for all $k = 0, 1, \ldots, n$, yields

$$0 = \left([\mu + \mathrm{i}(\omega - \nu_\Delta)] - \eta r_{\mathrm{eq},k}^3 \right) \mathrm{e}^{\mathrm{i}\varphi_{\mathrm{eq},k}} + \beta \rho_{\mathrm{ex},k} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{ex},k}}. \tag{13}$$

Equation (13) is analyzed by separating the amplitude $r_{\mathrm{eq},k}$ and the phase $\varphi_{\mathrm{eq},k}$ of the Fourier coefficient q_k. On the one hand, the phase $\varphi_{\mathrm{eq},k}$ can be directly determined by solving (13) for it. On the other hand, the equilibrium conditions for the amplitude $r_{\mathrm{eq},k}$ are derived by subtracting (13) with $\rho_{\mathrm{ex},k} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{ex},k}}$ and taking the absolute value squared from the result. This results in

$$\varphi_{\mathrm{eq}} = \arctan\left[\frac{\sqrt{\mu^2 + \omega_\Delta^2} \sin(\varphi)}{r_{\mathrm{eq}}^2 - \sqrt{\mu^2 + \omega_\Delta^2} \cos(\varphi)} \right] + \phi_{\mathrm{ex}}, \tag{14a}$$

$$0 = \eta^2 r_{\mathrm{eq},k}^6 - 2\mu\eta r_{\mathrm{eq},k}^4 + \left[\mu^2 + \omega_\Delta^2\right] r_{\mathrm{eq},k}^2 - \beta \rho_{\mathrm{ex},k}^2 \tag{14b}$$

with the phase $\varphi = \arctan(\omega_\Delta / \mu)$ and the difference frequency $\omega_\Delta = \omega - k\nu_\Delta$. Thus, the number of different solutions of (12) is given by number of positive, real-valued solutions of (14*b*). It can be shown that there is only one real-valued solution, e.g. by substituting $x = r_{\mathrm{eq},k}^2$ and applying Descartes' rule of signs, see, e.g. [47]. If $\mu\eta < 0$, $\omega_\Delta \in \mathbb{R}$, and $\rho_{\mathrm{ex},k} \in \mathbb{R}$ hold, it turns out that there is only one change of signs. Hence, (12) has two unique, positive, real-valued zero and the claim is shown. This result implies that the equilibrium of the Fourier coefficients $q_{\mathrm{eq},k}$ are unique, if (2) is frequency-selective.

The results are verified numerically by comparing the equilibria of the original envelope model (7) and the simplified envelope model (11). This is done by exciting these models with a two-tone signal and varying one amplitude of this two-tone signal, so that (2) is either satisfying or violating frequency selectivity. In particular, the change of frequency selectivity comes from the fact that the critical point is also dependent on the external excitation[11]. Here, the two-tone signal is given by

$$u = \rho_1 \mathrm{e}^{\mathrm{i}\nu_\Delta t} + \rho_k \mathrm{e}^{\mathrm{i}2\nu_\Delta t} \tag{15}$$

with the amplitudes $\rho_1, \rho_k > 0$ and the fundamental frequency $\nu_\Delta > 0$. With this, the (original) envelope model (7) reads

$$0 = (\mu + \mathrm{i}\omega)\, q_1 - \eta |q_1|^2 q_1 - 2\eta |q_2|^2 q_1 + \beta F_1, \tag{16a}$$

$$0 = [\mu + \mathrm{i}(\omega + \nu_\Delta)]\, q_2 - 2\eta |q_0|^2 q_2 - \eta |q_2|^2 q_2. \tag{16b}$$

For the simulations, the numerical parameters are given by the fundamental frequency $\nu_\Delta = 10\,\frac{1}{\mathrm{s}}$ and the amplitudes $\rho_1 \in \{10^0, 10^{-1}, 10^{-2}\}\,\frac{1}{\mathrm{s}}$ and $\rho_2 \in \{10^0, 10^{-1}\}\,\frac{1}{\mathrm{s}}$. The results are visualized in figure 8. Frequency selectivity is satisfied in figures 8(a) and (b), since the amplitude ρ_2 is small enough. Here, it is showcased that the measured frequency response of the original envelope model (7) and the simplified envelope model (11) is approximately the same. In contrast to this, frequency selectivity is violated in figures 8(c) and (d), since the amplitude ρ_2 is to large. In this case, it turns out that the measured frequency response is not sufficiently approximated by the simplified envelope model (11), if the amplitude ρ_1 is too small.

In summary, it is shown that the frequency domain can be reconstructed by a neuromorphic acoustic sensor based on oscillators, which exhibit at least a controllable Andronov–Hopf bifurcation, if the bifurcation parameter is in a neighborhood of the critical point. In this case, the response of these oscillators is frequency-selective and their trajectories will converge to the sub-manifold, which is isomorphic to an invariant manifold of (2) [22–24]. Therefore, following the results in this section, the steady state of the excited oscillators is unique, which implies the claim.

[11] This comes from the fact that the critical point of an Andronov–Hopf bifurcation is determined by analyzing the linearization of a system. As (2) is nonlinear and the equilibrium of (2) is shifted by the external excitation, the linearization changes. Hence, the critical point can be adjusted by the external excitation.

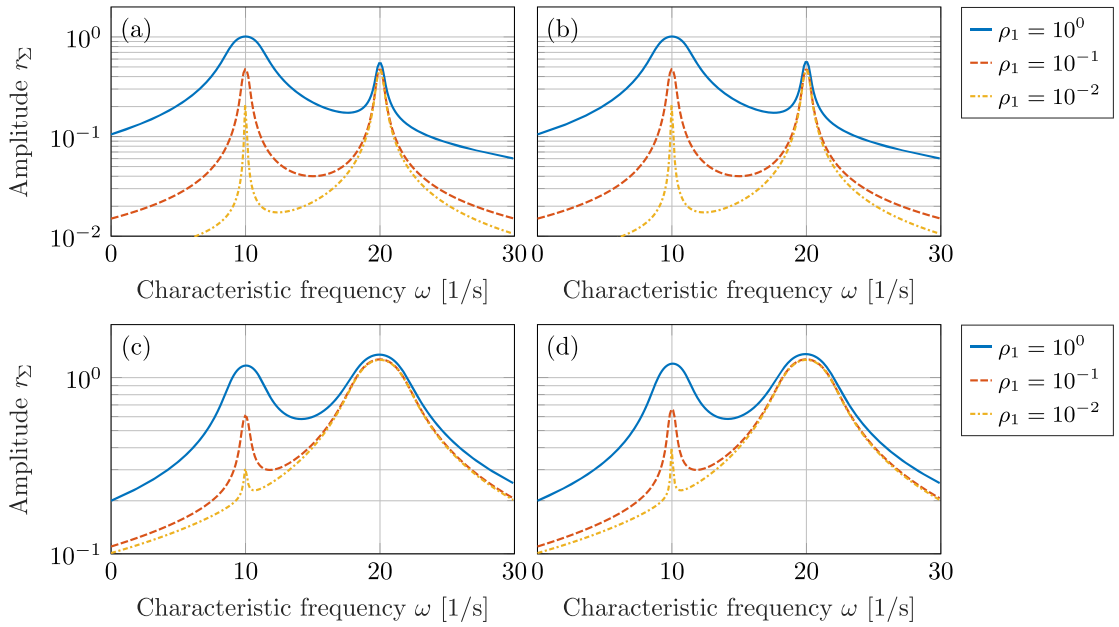

Figure 8. Comparison between the amplitude responses $r_\Sigma = r_1 + r_2$ of (7) and (11) as a function of the characteristic frequency ω and different harmonic signals with various external amplitudes ρ_1 and ρ_2 at the critical point, i.e. $\mu = 0\,\frac{1}{s}$. Additional parameters are given by the cubic damping $\eta = 1\,\frac{1}{s}$ and and the input constant $\beta = 1\,\frac{1}{s}$. (a),(b) Small excitation amplitude $\rho_2 = 10^{-1}$. (c),(d) Large excitation amplitude $\rho_2 = 10^0$. (a),(c) Amplitude response of the original envelope model (7). (b),(d) Amplitude response of the simplified envelope model (11).

5. Algorithms to sample the frequency domain

Subsequently, three different algorithms are introduced to sample the frequency domain. To access the Fourier coefficients $\boldsymbol{q}$ for an external input in (6), demodulation has to be employed. Here, amplitude demodulation is used, since the most important information of the external input is encoded on the envelope of the oscillator. For (2), amplitude demodulation can be done by taking the absolute value squared of its output, so that information of the phase is lost. With this demodulation scheme, the relevant features of the acoustic signal, e.g. the energy and the periodicity, can be determined [30]. Thus, the sampled spectrum can be in principle used for automatic speech recognition. In contrast, other demodulation schemes can also extract the phase of the (modulated) signal. By exploiting the frequency selectivity of (2), the amplitude demodulation is approximated by

$$|z|^2 = \left(\sum_{k=0}^{n} q_k \mathrm{e}^{\mathrm{i}k\nu_\Delta t}\right)\left(\sum_{k=0}^{n} q_k^* \mathrm{e}^{-\mathrm{i}k\nu_\Delta t}\right) \approx \sum_{k=0}^{n}\left(|q_k|^2\right) \approx \begin{cases} |q_k|^2, & \text{if } \omega \approx k\nu_\Delta \\ 0, & \text{else} \end{cases}. \tag{17}$$

Thus, the state z is a function in time t and the characteristic frequency ω and the Fourier coefficients $\boldsymbol{q}$ can be sampled by varying the characteristic frequency ω. From practical viewpoint, z is assumed to be discretized in time t and frequency ω. On the one hand, the frequency domain has to be discretized, since the continuous frequency domain has infinitely many frequencies to sample. Thus, a constrained frequency domain $\mathcal{F} = [\omega_{\min}, \omega_{\max}]$ with the minimal frequency and the maximum frequency $0 < \omega_{\min} < \omega_{\max}$ is discretized into $n_\omega + 1$ subintervals defined by $[\omega_k, \omega_{k+1}]$ for all $k = 0, 1, \ldots, n_\omega$. Herein, the limits are assumed to be equidistant and they are given by $\omega_k = \omega_{\min} + k\omega_s$ with the frequency spacing $\omega_s = \frac{\omega_{\max} - \omega_{\min}}{n_\omega + 1}$. On the other hand, the equilibrium $\boldsymbol{q}_{\mathrm{eq}}$ of the envelope model is approximated by holding the characteristic frequency ω constant for a sampling time $T_s > 0$, so that the oscillator locks onto the frequency of the external stimuli and $z(t, \omega)$ converges to the equilibrium $z_{\mathrm{eq}}(\omega)$ of (2). This implies that the change of the external stimuli F has to be much slower than the convergence time of system (2) and that the state $z(t, \omega)$ has to be discretized in time, i.e. $z(lT_s, \omega)$ for all $l \in \mathbb{N}$.

With these preliminaries, different algorithms to sample the frequency domain can be discussed. This is done by analyzing the convergence time of the design with frequency untunable and frequency tunable oscillators[12]. In particular, sampling with a frequency tunable oscillator is not necessarily sequential as the

[12] An oscillator is called frequency tunable, if its characteristic can be changed by a controllable input [33, 35].

external stimuli is assumed to be stationary for system (2). Hence, different algorithms can be applied. To this end, sequential sampling and a sampling approach based on reinforcement learning are introduced for a frequency tunable oscillator. As these algorithms are applied iteratively, the lth iteration of the sampled frequency response for sequential sampling and sampling based on reinforcement learning is denoted by

$$\mathrm{h}_{\mathrm{seq}}[l,k] = [\delta_{-1}(\omega - \omega_{\min} - k\omega_{\Delta}) - \delta_{-1}(\omega - \omega_{\min} - (k+1)\omega_{\Delta})]\, q_{\mathrm{seq},k}[l] \tag{18a}$$

$$\mathrm{h}_{\mathrm{RF}}[l,k] = [\delta_{-1}(\omega - \omega_{\min} - k\omega_{\Delta}) - \delta_{-1}(\omega - \omega_{\min} - (k+1)\omega_{\Delta})]\, q_{\mathrm{RF},k}[l] \tag{18b}$$

with the Heavyside function

$$\delta_{-1}(\omega) = \begin{cases} 1, & \text{if } \omega > 0 \\ 0, & \text{else} \end{cases}, \tag{19}$$

the sequentially sampled Fourier coefficients $q_{\mathrm{seq},k}[l]$ and the sampled Fourier coefficients $q_{\mathrm{seq},k}[l]$ based on reinforcement learning for all $k = 0, 1, \ldots, n_{\omega}$ and $l \in \mathbb{N}$. Finally, the continuous frequency response of the oscillators, i.e. $k \to \infty$, is denoted by $\mathrm{h}_{\mathrm{eq}}[l](\omega)$.

5.1. Algorithm 1: array of untunable oscillators

The design of a neuromorphic acoustic sensor can consist of an array of oscillators with constant and different characteristic frequencies. This is, e.g. sketched in figure 7. Because the frequency domain is determined by sampling the steady state of the envelope model, the convergence of the oscillators to their steady state, i.e. the locking of the oscillators onto a frequency, has to be characterized. Following [48], this can be approximated by linearizing the simplified envelope model (13) and assuming that their initial conditions are zero[13], which results in

$$p_z = 1 - \mathrm{e}^{\mu_{\mathrm{C}} T_{\mathrm{s}}} = 1 - \mathrm{e}^{-m_{\mathrm{conv}}} \tag{20}$$

with the real part $\mu_{\mathrm{C}} < 0$ of dominant eigenvalue of the system the sampling time $T_{\mathrm{s}} > 0$, and the convergence number $m_{\mathrm{conv}} = |\mu_{\mathrm{C}}| T_{\mathrm{s}}$. Here, the real part of the dominant eigenvalue for (13) is given by $\mu_{\mathrm{C}} = \mu$ and the convergence number m_{conv} can be used to characterize p_z independent from the oscillator. The convergence number m_{conv} is a design parameter of algorithm 1 and relevant values of p_z in terms of m_{conv} read [48]

$$p_z = \begin{cases} 95{,}0\%, & \text{if } |\mu_{\mathrm{C}}| T_{\mathrm{s}} = 3 \\ 98{,}1\%, & \text{if } |\mu_{\mathrm{C}}| T_{\mathrm{s}} = 4 \\ 99{,}3\%, & \text{if } |\mu_{\mathrm{C}}| T_{\mathrm{s}} = 5 \end{cases}, \tag{21}$$

so that the convergence time can be determined by

$$T_{\mathrm{s}} = \frac{m_{\mathrm{conv}}}{|\mu_{\mathrm{C}}|}. \tag{22}$$

The term $\mathrm{e}^{\mu_{\mathrm{C}} T_{\mathrm{s}}}$ can be viewed as an additional noise source, such that m_{conv} has to be chosen in terms of the signal-to-noise ratio. If the noise is too dominant, larger sampling times do not improve the overall result of the sampling algorithm.

In addition to the tuning of the convergence time T, the real part μ_{C} of the dominant eigenvalue adjusts the 3 dB bandwidth of the oscillator. The relationship between μ_{C} and the 3 dB bandwidth can be derived by comparing the experimentally determined Q-factor [49]

$$Q_{\mathrm{exp}} = \frac{f_{\mathrm{C}}}{f_{3\mathrm{dB}}}. \tag{23}$$

with the 2π-normalized characteristic frequency $f_{\mathrm{C}} = \omega_{\mathrm{C}}/(2\pi)$ and the 3 dB bandwidth $f_{3\mathrm{dB}} > 0$ and the Q-factor of a damped oscillator [50]

$$Q_{\mathrm{th}} = \left| \frac{\omega_{\mathrm{C}}}{\mu_{\mathrm{C}}} \right| \tag{24}$$

[13] This assumption results in an overestimation of the convergence time as the amplitude demodulated signal is always positive.

with the imaginary part $\omega_C \in \mathbb{R}$ of the dominant eigenvalue. As the dominant eigenvalue is considered, it follows that $Q_{th} = Q_{exp}$ and hence

$$f_{3dB} = \frac{|\mu_C|}{2\pi}. \tag{25}$$

Finally, (25) is inserted into (22), which results in

$$T_s = \frac{m_{conv}}{2\pi f_{3dB}}. \tag{26}$$

Therefore, the convergence time T and 3 dB bandwidth f_{3dB} are anti-proportional, which implies the sampling of the frequency domain is either fast or accurate.

5.2. Algorithm 2: sequential sampling

Regarding algorithm 2, the discretized frequency domain $\mathcal{F}$ is sampled by decreasing or increasing the characteristic frequency ω of a single frequency tunable oscillator and the frequency response is recovered by demodulating the output with (17). The procedure can be summarized as follows. First, the lower (or upper, respectively) limit of the frequency domain $\mathcal{F}$ is chosen as the characteristic frequency of the oscillator and this Fourier coefficient is sampled by the oscillator for a sampling time $T_s > 0$. Second, the characteristic frequency is increased (or decreased, respectively) by the frequency spacing ω_s and the next Fourier coefficient is sampled by the oscillator. Third, this is repeated until the whole frequency range $\mathcal{F}$ is sampled. The characteristic frequency is then set to its initial value. To compute the convergence time T_{seq} of this algorithm, note that the sampling time T_s satisfies (20), so that the convergence time T_{seq} of one iteration reads

$$T_{seq} = (n_\omega + 1)\, T_s. \tag{27}$$

Here, the sampling time T_s can be approximated with (21) and (22). Hence, the convergence time T_{seq} can be computed for any system without any statistical analysis, so that algorithm 2 can be considered to be the benchmark for an in-sensor Fourier transform with frequency tunable oscillators.

5.3. Algorithm 3: sampling based on reinforcement learning

Because the sampling time T_s of a system close to its critical point is rather slow, the sampler might miss important information of the acoustic signal. Hence, the important information has to be measured in a very restricted time interval. This issue can be resolved by assuming that information about the acoustic signals is encoded by a fundamental frequency and its harmonics. For instance, this assumption holds for vowels in speech [51]. In this case, only the most dominant Fourier coefficients have to be determined. Following the demodulation (17), noting that these Fourier coefficients are the maximum of the frequency spectrum, and discretizing z, an optimization problem arises, which reads

$$\max_{\omega \in \mathcal{F}} |z(t,\omega)| = \max_{k \in \{0,1,\ldots,n_\omega\}} |z[l,k]|. \tag{28}$$

It should be noted that gradient-based methods are not suitable to solve (28). This is caused by the nonlinear, frequency-selective transfer function of (2). Thus, the gradient of the cost function $|z_{eq}|$ is large, if the frequency ω is close to one maximizer, and it is small, if the frequency ω is far away from the maximizer (see figure 6(b)).

To avoid gradient-based methods, the optimization problem is reformulated based on the discretization of the frequency domain $\mathcal{F}$. With this, (28) is similar to the multi-armed bandit, see, e.g. [52–54], since the equilibrium value of the Fourier coefficient is only approximated. The Fourier coefficients can be interpreted as the probabilities for winning and the optimization (28) can be, e.g. solved by reinforcement learning [53].

For this, a variation of the epsilon-greedy strategy is applied to determine $n_{max} > 0$ maxima of (28). With this approach, either the frequency domain $\mathcal{F}$ is explored or (28) is optimized based on a value function $Q[l,k]$ with the iteration $l \in \mathbb{N}$ and the number $k = 0, 1, \ldots, n_\omega$ of the kth subinterval of the frequency domain $\mathcal{F}$. The value function $Q[l,k]$ represents the knowledge of the explored frequency domain and the correct amplitude is determined and processed by the compressive nonlinearity in the optimization phase. To explore the frequency domain, the policy π has to be defined and the set of possible actions is assumed to be given by $\mathcal{A} = \{0, 1, \ldots, n_\omega\}$. For frequency sampling, an action can be viewed as the choice of an frequency and the policy chooses a frequency accordingly. In particular, if all actions are made in kth

iteration, the value function $Q[l,k]$ is equivalent to the sampled frequency response $\mathrm{h}_{\mathrm{RF}}[l,k]$, i.e. $Q[l,k] = \mathrm{h}_{\mathrm{RF}}[l,k]$. For the purpose of this paper, the value function is assumed to be given by

$$Q[l+1,k] = aQ[l,k] + bp[l], \qquad l>0, \qquad Q[0,k] = Q_0[k] \tag{29}$$

with the initial condition $Q_0[k] \in \mathbb{R}$. The parameters are given by the forgetting rate $a \in [0,1)$ and the input parameter $b > 0$ for all $l \in \mathbb{N}$ and $k \in \mathcal{A}$. The forgetting rate a represents the speed in which the learned information is forgotten. While, the input parameter can be used to change the equilibrium of (29). Moreover, the reward $p[l]$ follows (17) and (28), so that the reward reads $p[l] = |z(lT_\mathrm{s})|$, and the policy of the epsilon-greedy strategy is given by

$$\pi[Q[l,k]] = \begin{cases} \text{random action from } \mathcal{A}, & \text{if } (i-1)\, n_\mathrm{s}^{\mathrm{RF}} \leq l < in_\mathrm{ex}^{\mathrm{RF}} \\ \text{exploiting } n_{\max}\text{maxima of } Q[l,k], & \text{if } in_\mathrm{ex}^{\mathrm{RF}} \leq l < in_\mathrm{s}^{\mathrm{RF}} \end{cases} \tag{30}$$

with the number of explorations $n_\mathrm{ex}^{\mathrm{RF}} \in \mathbb{N}$, the number of samples $n_\mathrm{s}^{\mathrm{RF}} \in \mathbb{N}$ for all $i \in \mathbb{N}$. Particularly, the policy chooses the Fourier coefficient, which is measured in the next time step. Thus, the optimizer is purely greedy, if $n_\mathrm{ex}^{\mathrm{RF}} = 0$, and purely exploratory, if $n_\mathrm{ex}^{\mathrm{RF}} = n_\mathrm{s}^{\mathrm{RF}}$. To increase the accuracy of this approach, the policy is separated into a exploration phase and a exploitation phase, since a larger sampling time improves the accuracy of the discovered maxima in the exploitation phase. In addition, the random action in the exploration phase is constructed to decrease the number of explorations $n_\mathrm{ex}^{\mathrm{RF}}$ as follows: a permutation $\mathcal{P}(\mathcal{A})$ of the set of action $\mathcal{A}$ is constructed. Then the actions are draw sequentially from $\mathcal{P}(\mathcal{A})$. After the set of permutation $\mathcal{P}(\mathcal{A})$ is empty, a new permutation is generated based on the set $\mathcal{A}$. With this, it is possible to draw every action from random sequence. As the maximum of the frequency response will be estimated by this approach, the sampling time T_s is decreased drastically, while increasing the accuracy of the sampled maxima.

Moreover, the parameters a and b of (29) have to be chosen properly to ensure proper convergence. This can be, e.g. done by determining analytic solution of (29), which reads

$$Q[l,k] = a^l Q[0,k] + \frac{b}{1-a} p[l]. \tag{31}$$

From this, it follows that (29) will converge to the quantitative values of the frequency domain, if $a \in [0,1)$ and $b = 1-a$. Moreover, the distance between (29) and its steady state is given by

$$p_Q = 1 - a^l. \tag{32}$$

Thus, it follows that the forgetting rate a has to be small, i.e. $a \ll 1$, since (29) has to converge after sampling all frequencies.

To compare the performance between algorithms 2 and 3, the time until one iteration of the reinforcement learning algorithm terminates has to be determined. Subsequently, the number of samples n_{RF}, which algorithm 3 takes, is fixed, so that the convergence time T_{RF} can be approximated by

$$T_{\mathrm{RF}} = n_{\mathrm{RF}} T_{\mathrm{RF}} \tag{33}$$

with the convergence time $T_{\mathrm{RF}} = m_{\mathrm{conv}}^{\mathrm{RF}}/|\mu|$ and the convergence number $m_{\mathrm{conv}}^{\mathrm{RF}}$. As the algorithm is motivated by the multi-armed bandit problem, the sampling time can be decreased substantially in comparison to the convergence time of the system, i.e. $T_{\mathrm{RF}} \ll T_\mathrm{s}$ and $m_{\mathrm{conv}}^{\mathrm{RF}} \ll m_{\mathrm{conv}}$. However, more samples have to be taken, which implies $n_{\mathrm{RF}} > n_\omega + 1$.

6. Numerical evaluation

Subsequently, the frequency responses $\mathrm{h}_{\mathrm{seq}}[l,k]$ and $\mathrm{h}_{\mathrm{RF}}[l,k]$ of algorithms 1 and 2 and the (continuous) frequency response $\mathrm{h}_{\mathrm{eq}}[l](\omega) = \lim_{n_\omega \to \infty} \mathrm{h}_{\mathrm{seq}}[l,k]$ are compared for different (harmonic) signals. For this, the parameters of the oscillator (2) are given in table B1. In particular, the interval length ω_Δ and the limits of the frequency domain $\mathcal{F}$ imply that 13 frequencies are measured, i.e. $n_\omega = 12$. As the convergence number is given by $m_{\mathrm{conv}} = 4$, algorithms 1 and 2 require each iteration $T_\mathrm{s} = 5$ ms (for algorithm 1) and $T_{\mathrm{seq}}(= 13 \cdot 5) = 65$ ms for algorithm 2 to sample the frequency response. Hence, algorithm 2 is not fast enough to sample the frequency domain $\mathcal{F}$ for automatic speech processing as around 20–30 ms are allowed [30]. In addition, the convergence number m_{conv} implies that each frequency is converged to 98.1%. To evaluate the statistics of algorithm 3, it is simulated 1000 times and its average and its confidence interval, which is assumed to be by three times its standard deviation, is analyzed. By choosing three times the

Table 1. Comparison between the convergence time per iteration of the algorithms for the four different scenarios.

Convergence time	Scenario 1	Scenario 2	Scenario 3	Scenario 4
T_s	5 ms	5 ms	5 ms	5 ms
T_{seq}	65 ms	65 ms	65 ms	65 ms
$T_{RF,1}$	65 ms	65 ms	65 ms	65 ms
	$100\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$100\% \cdot T_{seq}$
$T_{RF,2}$	56.875 ms	65 ms	65 ms	56.875 ms
	$87.5\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$87.5\% \cdot T_{seq}$
$T_{RF,3}$	48.75 ms	65 ms	65 ms	48.75 ms
	$75\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$100\% \cdot T_{seq}$	$75\% \cdot T_{seq}$
$T_{RF,4}$	32.5 ms	48.75 ms	48.75 ms	32.5 ms
	$50\% \cdot T_{seq}$	$75\% \cdot T_{seq}$	$75\% \cdot T_{seq}$	$50\% \cdot T_{seq}$

standard deviation, 99.8% of all solutions lie in the confidence interval. Moreover, it is assumed for algorithm 3 that the complete frequency domain $\mathcal{F}$ is sampled in the exploration phase. The discovered maxima are then maximized in the exploitation phase. Hence, if the convergence numbers m_{conv}^{seq} and m_{conv}^{RF} of the algorithms 1 and 2 are equivalent, the convergence time T_{RF} of algorithm 3 is bounded from above by the convergence time T_{seq} of algorithm 2 and from below by the convergence time T_s of algorithm 1, i.e. $T_{RF} \in [T_s, T_{seq}]$.

To compare the performance of the algorithms, the accuracy of the sampled frequency domain $\mathcal{F}$ is analyzed with respect to the convergence number m_{conv}^{RF}, the convergence time T_{RF}, and the number of maxima n_{max} of algorithm 3. For this, (2) is excited by four different external stimuli, a time-varying signal u_{tv}, a two-tone signal u_{tt}, a harmonic input with a harmonic distortion u_{hd}, and a harmonic input with additive white Gaussian noise. These excitation are given by

$$u_{tv} = \rho(t)\, e^{i[\nu(t)t+\phi(t)]}, \tag{34a}$$

$$u_{tt} = \rho_{1,tt} e^{i(\omega_{1,tt}t+\varphi_{1,tt})} + \rho_{2,tt} e^{i(\omega_{2,tt}t+\varphi_{2,tt})}, \tag{34b}$$

$$u_{hd} = \rho_{1,hd} e^{i(\omega_{1,hd}t+\varphi_{1,hd})} + \rho_{2,hd} e^{i(\omega_{2,hd}t+\varphi_{2,hd})}, \tag{34c}$$

$$u_{awgn} = \rho_{awgn} e^{i\left(\omega_{awgn}t+\varphi_{awgn}\right)} + N \tag{34d}$$

with

$$\rho(t) = \begin{cases} \rho_{1,tv}, & \text{if } t < t_{th} \\ \rho_{2,tv}, & \text{if } t \geq t_{th} \end{cases}, \; \nu(t) = \begin{cases} \omega_{1,tv}, & \text{if } t < t_{th} \\ \omega_{2,tv}, & \text{if } t \geq t_{th} \end{cases}, \; \phi(t) = \begin{cases} \varphi_{1,tv}, & \text{if } t < t_{th} \\ \varphi_{2,tv}, & \text{if } t \geq t_{th} \end{cases}, \tag{35}$$

and the additive white Gaussian noise $N(t) \in \mathcal{N}(0, \sigma^2)$. The parameters are given by the amplitudes $\rho_{i,tv}, \rho_{i,tt}, \rho_{i,hd}, \rho_{awgn} > 0$, the frequencies $\omega_{i,tv}, \omega_{i,tt}, \omega_{i,hd}, \omega_{awgn} > 0$, the phases $\varphi_{i,tv}, \varphi_{i,tt}, \varphi_{i,hd}, \varphi_{awgn} \in [0, 2\pi)$, the threshold $t_{th} > 0$, and the standard deviation $\sigma > 0$ for all $i = 1, 2$. In addition, these scenarios are evaluated for four different cases. The numerical values for these cases are given in table B2, and the threshold t_{th} is assumed to be equivalent to the convergence time of algorithms 2 and 3, respectively. In particular, the convergence of algorithm 3 is verified by the first parameter set as these imply a pure exploration.

6.1. Scenario 1: time-varying signal

The results of scenario 1 are illustrated in figure 9 and the second column of table 1. It turns out that by choosing $m_{conv}^{RF} = m_{conv}$, the results of algorithms 2 and 3 are identical. This is depicted in figure 9(a). In comparison to $m_{conv,1}^{RF}$, the convergence numbers $m_{conv,2}^{RF}$, $m_{conv,3}^{RF}$, and $m_{conv,4}^{RF}$ are decrease. With this, it is possible to increase the number of samples in one iteration, so that the frequency response can be first estimated and the equilibria of its maxima are then precisely measured. Thereby, the frequency domain $\mathcal{F}$ can be sampled in a shorter time as shown in the second column of table 1, while preserving the same accuracy of the maximum. This is visualized in figures 9(b)–(d) in particular, the convergence time can be decreased by at least 50% in this scenario, while algorithm 3 can still extract the most important information of the external stimuli. Thus, it is possible to enable speech processing by decreasing the convergence time with different algorithms.

6.2. Scenario 2: two-tone signal

The results of scenario 2 are shown in figure 10 and in the third column of table 1. In particular, it can be observed that the accuracy of both algorithms is inhibited by the two-tone signal. Thus, algorithm 2 and the

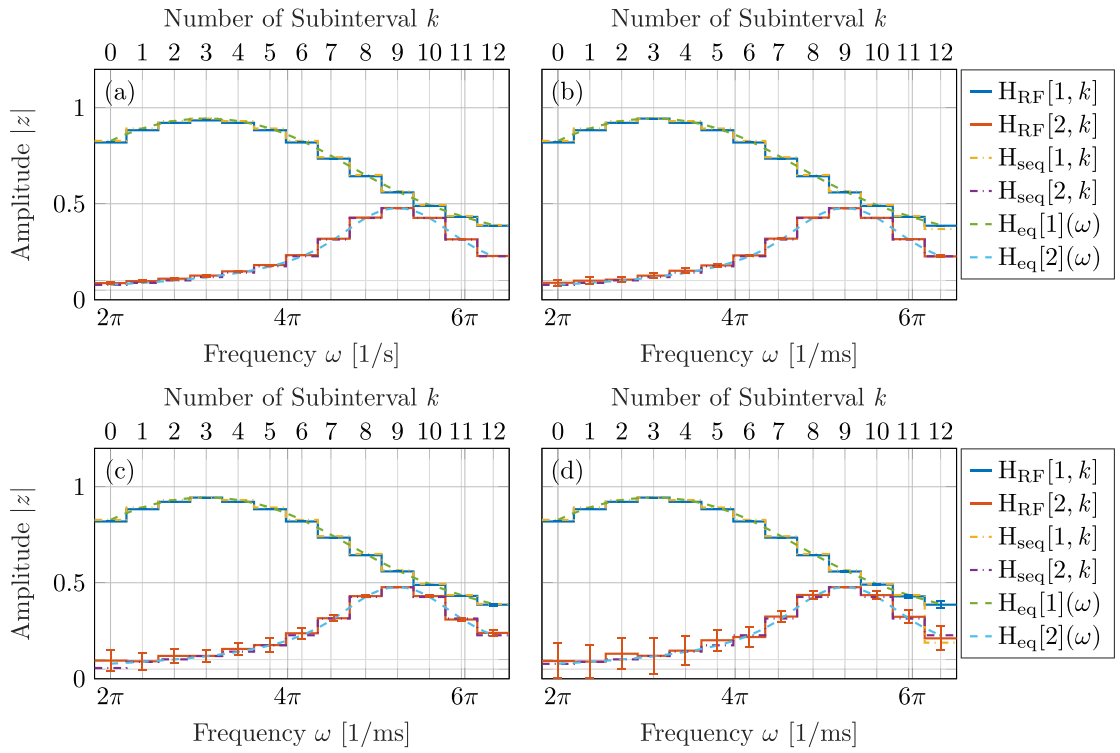

Figure 9. Comparison between the sampled frequency responses $\mathrm{h_{seq}}[l,k]$ and $\mathrm{h_{RF}}[l,k]$ of algorithms 2 and 3 and the (continuous) frequency response $\mathrm{h_{eq}}[1]$ of oscillator (2) after two iterations, i.e. $l=1,2$. The frequency domain is sampled from $2\pi\cdot 1000\,\frac{1}{\mathrm{s}}$ to $2\pi\cdot 3160\,\frac{1}{\mathrm{s}}$ and separated into 13 subintervals. A time varying signal is sampled for different convergence number $m^{\mathrm{RF}}_{\mathrm{conv},i}$, number of maxima $n_{\mathrm{max},i}$, and sampling time $T_{\mathrm{RF},i}$ for all $i=1,2,3,4$. (a) Validation for the equivalence between algorithms 2 and 3 with the parameters given by convergence number $m^{\mathrm{RF}}_{\mathrm{conv},1}=4$, number of maxima $n_{\mathrm{max},1}=1$, and sampling time $T_{\mathrm{RF},1}=T_{\mathrm{seq}}$. (b)–(d) Comparison between algorithms 2 and 3 with the convergence number $m^{\mathrm{RF}}_{\mathrm{conv},2}=3$, $m^{\mathrm{RF}}_{\mathrm{conv},3}=2$, and $m^{\mathrm{RF}}_{\mathrm{conv},4}=1$ and the number of maxima $n_{\mathrm{max},2}=n_{\mathrm{max},3}=n_{\mathrm{max},4}=1$. With these parameters, the sampling time of algorithm 3 is decreased to $T_{\mathrm{RF},2}=87.5\%\cdot T_{\mathrm{seq}}$ (for (b)), $T_{\mathrm{RF},3}=75\%\cdot T_{\mathrm{seq}}$ (for (c)), and $T_{\mathrm{RF},4}=50\%\cdot T_{\mathrm{seq}}$ (for (d)).

purely exploratory algorithm 3 cannot converge to the equilibrium of the perfectly sampled frequency response $\mathrm{h_{eq}}[k]$. However, by decreasing the convergence number $m^{\mathrm{RF}}_{\mathrm{conv}}$ and thus increasing the time of exploitation of the maxima, either the accuracy can be improved or the convergence time can be decreased. For instance, the maxima are robustly identified by decreasing the convergence number to 2 as depicted in figures 10(b) and (c). By additionally decreasing the convergence time $T^{\mathrm{tt}}_{\mathrm{RF}}$, the accuracy of the result becomes similar to the sequential sampling as is illustrated in figure 10(d).

6.3. Scenario 3: harmonic disturbance

The results of scenario 3 are shown in figure 11 and in the forth column of table 1. In contrast to scenario 2, the accuracy is not drastically inhibited by the harmonic disturbance, which is far away from a harmonic input. In addition, it is showcased that by choosing the parameters accordingly the accuracy and the convergence time can be improved by using algorithm 3. This is depicted in figures 11(b) and (d). Moreover, it is illustrated in figure 11(c) that by search for an additional maximum the accuracy of the maximum is similar to algorithm 2.

6.4. Scenario 4: additive white Gaussian noise

Finally, a harmonic excitation distorted by additive white Gaussian noise is considered. The results are visualized in figure 12 and in the fifth column of Table 1, which demonstrate that the most important features of the input signal can be reconstructed even though it is distorted by noise and that the convergence time can be substantially improved by algorithm 3. However, this evaluation is done in an environment with small noise as the signal-to-noise ratio reads

$$\mathrm{SNR}=20\log_{10}\left(\frac{\rho_{\mathrm{awgn}}}{\sigma}\right)=0\,\mathrm{dB}. \tag{36}$$

Thus, by sampling the frequency domain with an oscillator, which exhibits an Andronov–Hopf bifurcation, noise on external stimuli can be suppressed. These observations can be generalized also for other noise

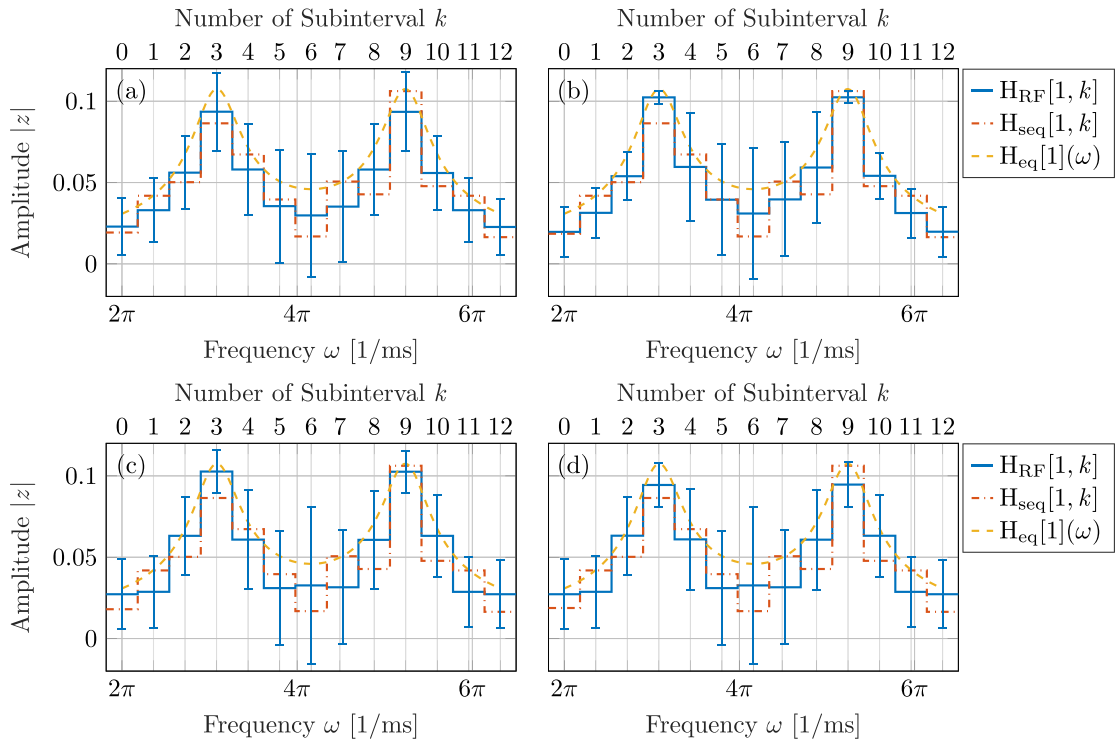

Figure 10. Comparison between the sampled frequency responses $\mathrm{h_{seq}}[1,k]$ and $\mathrm{h_{RF}}[1,k]$ of algorithms 2 and 3 and the (continuous) frequency response $\mathrm{h_{eq}}[1]$ of oscillator (2) after one iteration. The frequency domain is sampled from $2\pi \cdot 1000\,\frac{1}{\mathrm{s}}$ to $2\pi \cdot 3160\,\frac{1}{\mathrm{s}}$ and separated into 13 subintervals. A two-tone signal is sampled for different convergence number $m^{\mathrm{RF}}_{\mathrm{conv},i}$, number of maxima $n_{\mathrm{max},i}$, and sampling time $T_{\mathrm{RF},i}$ for all $i = 1,2,3,4$. (a) Validation for the equivalence between algorithms 2 and 3 with the parameters given by convergence number $m^{\mathrm{RF}}_{\mathrm{conv},1} = 4$, number of maxima $n_{\mathrm{max},1} = 1$, and sampling time $T_{\mathrm{RF},1} = T_{\mathrm{seq}}$. (b)–(d) Comparison between algorithms 2 and 3 with the convergence number $m^{\mathrm{RF}}_{\mathrm{conv},2} = 3$ and $m^{\mathrm{RF}}_{\mathrm{conv},3} = m^{\mathrm{RF}}_{\mathrm{conv},4} = 2$ and the number of maxima $n_{\mathrm{max},2} = n_{\mathrm{max},3} = n_{\mathrm{max},4} = 2$. With these parameters, the sampling time of algorithm 3 is given by $T_{\mathrm{RF},2} = T_{\mathrm{RF},3} = 100\% \cdot T_{\mathrm{seq}}$ (for (b) and (c)) and $T_{\mathrm{RF},4} = 75\% \cdot T_{\mathrm{seq}}$ (for (d)).

models added to the external input. Then the power spectrum of the noise becomes a function of the frequency ω. This, however, will not change the results tremendously as the dependency of the frequency ω only implies that the distortion is changing for each sampled frequency.

7. Consequences for neuromorphic acoustic sensing

Subsequently, the consequences of sections 5 and 6 are briefly discussed for an neuromorphic acoustic sensor consisting of thermally actuated MEMS sensor. For this, the MEMS sensor with a geometric nonlinearity is considered [32, 43] and the setup to enable an in-sensor Fourier transform is illustrated in figure 13. This setup consists of a MEMS sensor (with a geometric nonlinearity), a high pass filter, an envelope detection, an algorithm to choose the characteristic frequency, and a controller to assign the desired frequency to the MEMS sensor. This setup works as follows: First, the deflection of the MEMS sensor is high pass filtered and the envelope of the signal is determined. With this, the amplitude of the Fourier coefficient for the characteristic frequency is measured. This amplitude is used to determine the next sampled frequency with, e.g. an algorithm based on reinforcement learning. Then the desired frequency is converted in to a DC-voltage to tune the characteristic frequency of the MEMS sensor.

To determine the maximum number of sampled frequency, the convergence time of this setup has to be analyzed. This is done by solely analyzing the dominant eigenvalue of the MEMS sensor. For this, it is assumed that the MEMS sensor is described by the dominant mode model [32]. In contrast to this, the

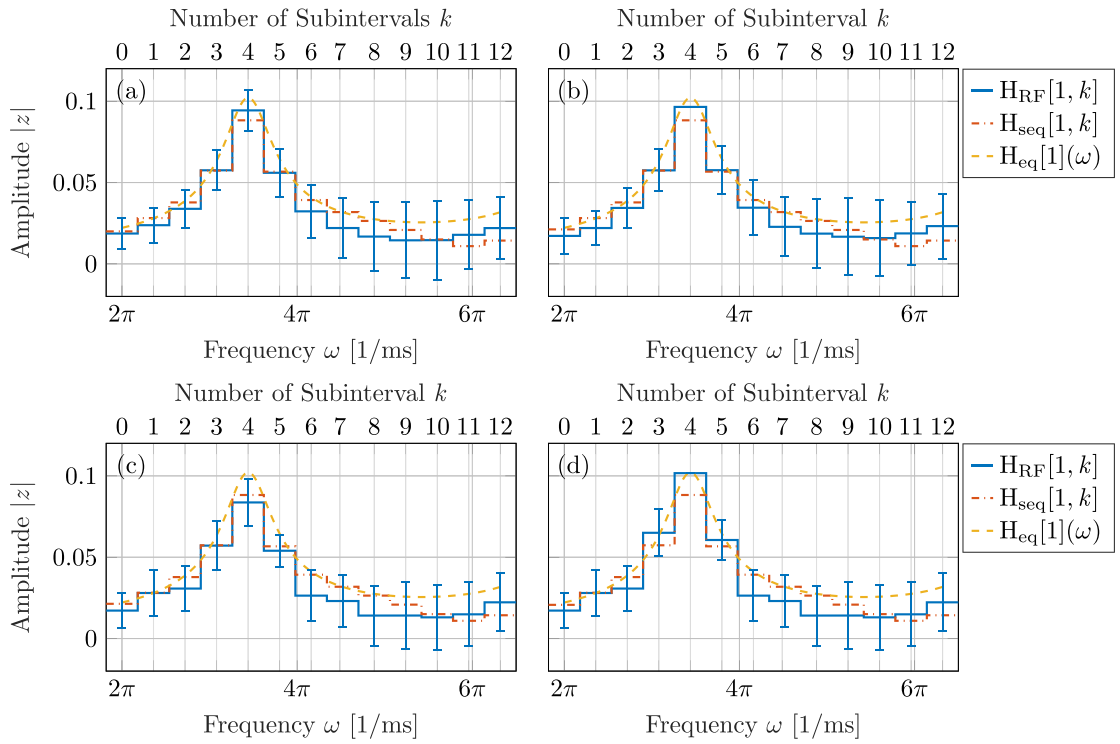

Figure 11. Comparison between the sampled frequency responses $\mathrm{h_{seq}}[1,k]$ and $\mathrm{h_{RF}}[1,k]$ of algorithms 2 and 3 and the (continuous) frequency response $\mathrm{h_{eq}}[1]$ of oscillator (2) after one iteration. The frequency domain is sampled from $2\pi \cdot 1000\,\frac{1}{\mathrm{s}}$ to $2\pi \cdot 3160\,\frac{1}{\mathrm{s}}$ and separated into 13 subintervals. A harmonically disturbed input is sampled for different convergence number $m^{\mathrm{RF}}_{\mathrm{conv},i}$, number of maxima $n_{\mathrm{max},i}$, and sampling time $T_{\mathrm{RF},i}$ for all $i = 1,2,3,4$. (a) Validation for the equivalence between algorithms 2 and 3 with the parameters are given by convergence number $m^{\mathrm{RF}}_{\mathrm{conv},1} = 4$, number of maxima $n_{\mathrm{max},1} = 1$, and sampling time $T_{\mathrm{RF},1} = T_{\mathrm{seq}}$. (b)–(d) Comparison between algorithms 2 and 3 with the convergence number $m^{\mathrm{RF}}_{\mathrm{conv},2} = 3$ and $m^{\mathrm{RF}}_{\mathrm{conv},3} = m^{\mathrm{RF}}_{\mathrm{conv},4} = 2$ and the number of maxima $n_{\mathrm{max},2} = n_{\mathrm{max},4} = 1$ and $n_{\mathrm{max},3} = 2$. With these parameters, the sampling time of algorithm 3 is given by $T_{\mathrm{RF},2} = T_{\mathrm{RF},3} = 100\% \cdot T_{\mathrm{seq}}$ (for (b) and (c)) and $T_{\mathrm{RF},4} = 75\% \cdot T_{\mathrm{seq}}$ (for (d)).

filters, i.e. the high pass filter and the envelope detector, are asserted to be given by 1st order filters. Thus, the setup is governed by

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \overbrace{\begin{bmatrix} -x_2 \\ -c_1 x_1 - c_3 x_1^3 - \mu x_2 + \alpha x_3 + \frac{1}{m} F_{\mathrm{ex}} \\ -\beta x_3 + \zeta\,(k x_4 + \nu)^2 \\ -\frac{1}{T_1} x_4 + \kappa_1 x_2 \\ -\frac{1}{T_2} x_5 + \kappa_2 |x_4|^2 \end{bmatrix}}^{=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})}, \qquad t > 0, \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{37}$$

with the input vector $\boldsymbol{u}(t) = [F_{\mathrm{ex}}(t), \nu(t)]^{\mathrm{T}} \in \mathbb{R}^2$, the state vector $\boldsymbol{x}(t) = [x_1(t), x_2(t), x_3(t), x_4(t), x_5(t)]^{\mathrm{T}} \in \mathbb{R}^5$ and the initial conditions $\boldsymbol{x}_0 \in \mathbb{R}^5$. The state vector $\boldsymbol{x}$ is composed of the deflection $x_1(t) \in \mathbb{R}$, the velocity $x_2(t) \in \mathbb{R}$, the ambient temperature $x_3(t) \in \mathbb{R}$, the voltage $x_4(t) \in \mathbb{R}$ of the high pass filter, and the voltage $x_5(t) \in \mathbb{R}$ of the envelope detector, while the input vector consists of the thermal actuator $\nu(t) \in \mathbb{R}$ and the external input $F_{\mathrm{ex}}(t) \in \mathbb{R}$. Additional parameters are given by the spring constants $c_1 > 0$, $c_3 > 0$, damping $\mu > 0$, the transfer factors $\zeta = \gamma / R^2 > 0$, $\alpha, \gamma > 0$, the time constants $\beta, T_1, T_2 > 0$, the mass $m > 0$, the resistance $R > 0$, the feedback strength $k \in \mathbb{R}$, and the calibration factors $\kappa_1, \kappa_2 > 0$. It is showcased in [32] that (37) has two Andronov–Hopf bifurcations in terms of the feedback strength and the frequency can be tuned by exploiting the geometric nonlinearity. Hence, system (37) can be used to implement the in-sensor Fourier transform. The effects of the geometric nonlinearity and the convergence rate are described by the local dynamics of (37). This is given by the system matrix, which is determined by computing the Jacobian of

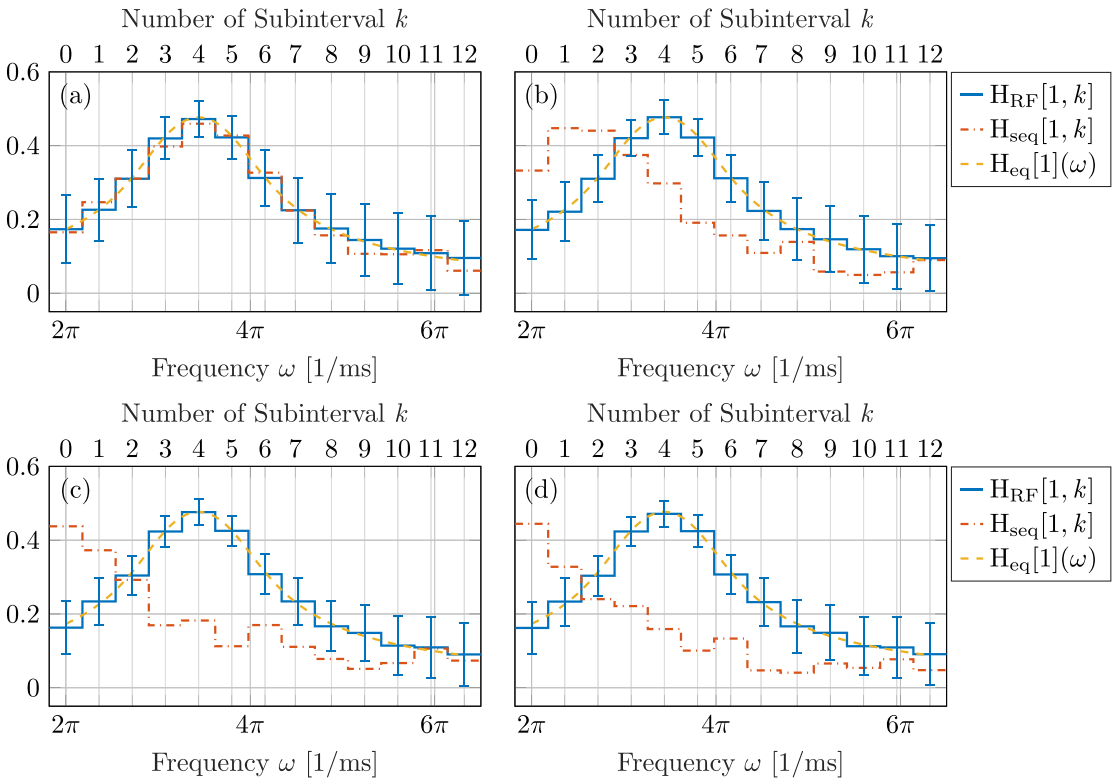

Figure 12. Comparison between the sampled frequency responses $\mathrm{h_{seq}}[1,k]$ and $\mathrm{h_{RF}}[1,k]$ of algorithms 2 and 3 and the (continuous) frequency response $\mathrm{h_{eq}}[1]$ of oscillator (2) after one iteration. The frequency domain is sampled from $2\pi \cdot 1000\,\frac{1}{\mathrm{s}}$ to $2\pi \cdot 3160\,\frac{1}{\mathrm{s}}$ and separated into 13 subintervals. A harmonic input with additive Gaussian noise is sampled for different convergence number $m^{\mathrm{RF}}_{\mathrm{conv},i}$, number of maxima $n_{\mathrm{max},i}$, and sampling time $T_{\mathrm{RF},i}$ for all $i = 1,2,3,4$. (a) Validation for the equivalence between algorithms 2 and 3 with the parameters are given by convergence number $m^{\mathrm{RF}}_{\mathrm{conv},1} = 4$, number of maxima $n_{\mathrm{max},1} = 1$, and sampling time $T_{\mathrm{RF},1} = T_{\mathrm{seq}}$. (b)–(d) Comparison between algorithms 2 and 3 with the convergence number $m^{\mathrm{RF}}_{\mathrm{conv},2} = 3$, and $m^{\mathrm{RF}}_{\mathrm{conv},3} = m^{\mathrm{RF}}_{\mathrm{conv},4} = 2$ and the number of maxima $n_{\mathrm{max},2} = n_{\mathrm{max},3} = n_{\mathrm{max},4} = 1$. With these parameters, the sampling time of algorithm 3 is decreased to $T_{\mathrm{RF},2} = 87.5\% \cdot T_{\mathrm{seq}}$ (for (b)), $T_{\mathrm{RF},3} = 75\% \cdot T_{\mathrm{seq}}$ (for (c)), and $T_{\mathrm{RF},4} = 50\% \cdot T_{\mathrm{seq}}$ (for (d)).

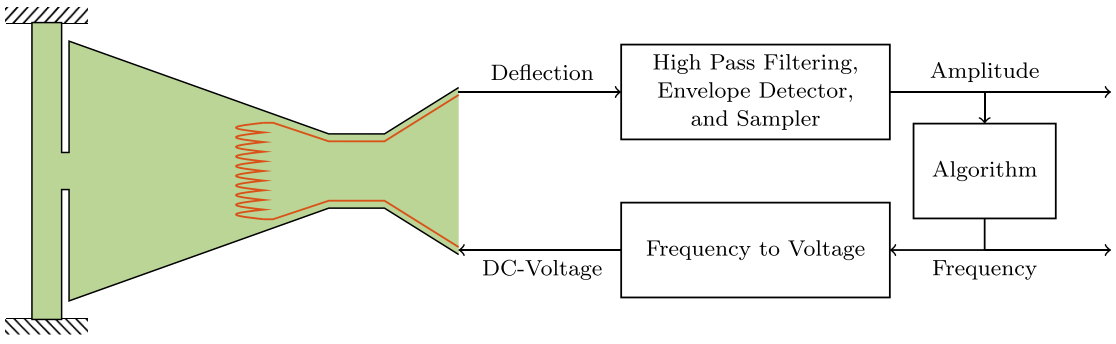

Figure 13. Bio-inspired system design to enable sampling in the frequency domain $\mathcal{F}$. Frequency tunability is implemented by inducing geometric nonlinearities in a thermally actuated MEMS sensor, so that the strain-stress relation becomes nonlinear. The deflection of the MEMS sensor is deduced by measuring the strain with a piezoelectric strain gauge and its characteristic frequency can be tuned by changing the pre-deflection with a DC-voltage. To enable frequency sampling in this device, the deflection is high pass filtered and its envelope is determined. Based on the magnitude of this envelope, the next sampled frequency is chosen and the desired DC-voltage is computed from this frequency to properly tune the MEMS sensor. To implement the algorithm and the relationship between frequency and voltage, a FPGA can be used [26].

$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})$ and inserting the equilibrium values $\boldsymbol{x}_{\mathrm{eq}} \in \mathbb{R}^5$ of (37). This yields

$$A = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\left(\boldsymbol{x}_{\mathrm{eq}}, \boldsymbol{u}_{\mathrm{eq}}\right) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -\Omega^2 & -\mu & \alpha & 0 & 0 \\ 0 & 0 & -\beta & 2\zeta u_{\mathrm{DC}} k & 0 \\ 0 & \kappa_1 & 0 & -\frac{1}{T_1} & 0 \\ 0 & 0 & 0 & \kappa_2 & -\frac{1}{T_2} \end{bmatrix} \tag{38}$$

with the frequency $\Omega^2 = c_1 + 3c_3 x_{1,\mathrm{eq}}^2$. Note that (37) will also be frequency-selective with $k = 0$. This will simplify the analysis of the convergence time, since (38) can be divided into 4 independent block matrices. Then the eigenvalues of (38) read

$$\lambda_1 = -\beta, \qquad \lambda_2 = -\frac{1}{T_1}, \qquad \lambda_3 = -\frac{1}{T_2}, \qquad \lambda_{4,5} = -\frac{\mu}{2} \pm \mathrm{i}\Omega\left(u_{\mathrm{DC}}\right). \tag{39}$$

Inserting the relevant, numerical parameters given in table B3, yields

$$\lambda_1 = -256\,\frac{1}{\mathrm{s}}, \qquad \lambda_2 = -1000\,\frac{1}{\mathrm{s}}, \qquad \lambda_3 = -1000\,\frac{1}{\mathrm{s}}, \qquad \lambda_{4,5} \approx -125 \pm \mathrm{i}\Omega\left(u_{\mathrm{DC}}\right)\,\frac{1}{\mathrm{s}}, \tag{40}$$

so that the convergence time is given by

$$T_{\mathrm{s}} = \begin{cases} 7.9\mathrm{ms}, & \text{if } m_{\mathrm{conv}} = 1 (\text{or } p = 63,1\%) \\ 15.7\mathrm{ms}, & \text{if } m_{\mathrm{conv}} = 2 (\text{or } p = 86,5\%) \\ 23.6\mathrm{ms}, & \text{if } m_{\mathrm{conv}} = 3 (\text{or } p = 95,0\%) \\ 31.5\mathrm{ms}, & \text{if } m_{\mathrm{conv}} = 4 (\text{or } p = 98,1\%) \\ 39.4\mathrm{ms}, & \text{if } m_{\mathrm{conv}} = 5 (\text{or } p = 99,3\%) \end{cases}. \tag{41}$$

Asserting that the proposed setup should have an accuracy of $p = 86,5\%$, it is able to sample at least two frequencies in 31.4 ms. This is (approximately) in the time frame from 20 ms to 30 ms of automatic speech recognition [30]. Hence, the number of MEMS sensors in a neuromorphic acoustic sensor can be at least halved with this approach.

8. Conclusion

Motivated by the frequency decomposition inside the cochlea, an in-sensor Fourier transform is proposed for a neuromorphic acoustic sensor based on an array of oscillators, which exhibit Andronov–Hopf bifurcations. It is demonstrated by analyzing the uniqueness of the response of the Andronov–Hopf bifurcation that these oscillators have a unique response to external stimuli, if their bifurcation parameter is in the neighborhood of the critical point of their respective Andronov–Hopf bifurcation. With this, the Fourier coefficients of an external stimuli can be uniquely reconstructed, so that an in-sensor Fourier transform can be implemented by employing amplitude demodulation on the output of the oscillator. Finally, algorithms have been proposed to improve the convergence time of the in-sensor Fourier transform. Herein, two approaches have been investigated in detail: on the one hand, an array of oscillators with untunable frequency is analyzed with respect to its convergence time. It is demonstrated that the convergence time can be deduced from the eigenvalues of its linearization and that the bandwidth of the oscillators and their convergence time are anti-proportional. On the other hand, two algorithms for a single, frequency tunable oscillator are proposed. Here, sequential sampling, which has been defined as a benchmark for the in-sensor Fourier transform, and sampling based on reinforcement learning are compared. It is showcased that the convergence time of the sequential sampling can be tremendously reduced by designing a algorithm based on reinforcement learning. Hence, the latter approach paves the way to decrease the number of oscillators in an neuromorphic acoustic sensor. For further research, the in-sensor Fourier transform can be implemented using an FPGA following section 7 and different algorithms based on models of pitch perception, see, e.g. [55–57], can be developed.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 434434223 – SFB 1461.

Author contributions

Hermann Folke Johann Rolf: Conceptualization (lead); Formal analysis (lead); Software (lead); Visualization (lead); Writing - original draft (lead); Writing - review & edit- ing (equal); **Petro Feketa:** Formal analysis (supporting); Supervision (equal); Visualization (supporting); Writing - review & editing (equal); **Alexander Schaum:** Formal analysis (supporting); Visualization (supporting); Writing - review & editing (equal); **Thomas Meurer:** Formal analysis (supporting); Funding acquisition (lead); Supervision (equal); Visualization (supporting); Writing - review & editing (equal);

Competing Interests

The authors declare no competing interests.

Code availability

The custom-developed codes for the MATLAB simulation are available from the corresponding author upon request.

Appendix A. Convergence to 1:1 frequency-lock

Subsequently, the assumption on 1:1 frequency-lock of the harmonically excited system (2) is shown, i.e. the input is given by $u = \rho e^{\mathrm{i}(\nu t+\phi)}$. This is done by analyzing the equilibria of the phase and their stability. Thus, (2) has to be transformed into the polar coordinates, which can be done by denoting the amplitude and phase by

$$r = |z|, \tag{A.1a}$$

$$\varphi = \tan\left(\frac{\mathrm{Im}\,(z)}{\mathrm{Re}\,(z)}\right). \tag{A.1b}$$

In particular, the amplitude and phase satisfy $r > 0$ and $\varphi \in (-\pi, \pi]$. Taking the derivative with respect to t of (A.1) and inserting (2), yields the polar coordinates

$$\dot{r} = \mu r - \eta r^3 + \rho \cos(\nu t + \phi - \varphi), \tag{A.2}$$

$$\dot{\varphi} = \omega - \frac{\rho}{r} \sin(\varphi - \nu t - \phi). \tag{A.3}$$

For 1:1 frequency lock, one equilibrium of the phase difference $\Delta\varphi = \varphi - \nu t - \phi$ has to be asymptotically stable. Here, the phase difference is governed by

$$\Delta\dot{\varphi} = \Delta\varphi - \frac{\rho}{r} \sin(\Delta\varphi) \tag{A.4}$$

with $\Delta\omega = \omega - \nu$, so that its equilibria is given by

$$\sin\left(\Delta\varphi_{\mathrm{eq}}\right) = \frac{r_{\mathrm{eq}}}{\rho}(\omega - \nu). \tag{A.5}$$

It has to be stressed that (A.5) has two distinct solution on $(-\pi, \pi]$ and one equilibrium is asymptotically stable, which is shown graphically in figure A1. This can be done since the amplitude satisfies $r > 0$. In particular, the slope around the equilibrium $\Delta\varphi_{\mathrm{s}}$ is negative, while the slope around the equilibrium $\Delta\varphi_{\mathrm{u}}$ is positive. Combining this with the fact that the phase is defined only on $(-\pi, \pi]$, asymptotic stability of $\Delta\varphi_{\mathrm{s}}$ and thus 1:1 frequency lock directly follow.

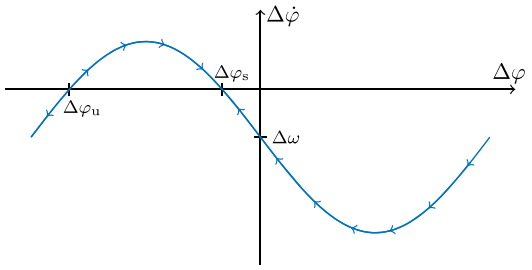

Figure A1. Sketch of (A.4) as a function of $\Delta\varphi$. Note that the slope around the equilibrium $\Delta\varphi_s$ is negative, while the slope around the equilibrium $\Delta\varphi_u$ is positive. Hence, it follows that $\Delta\varphi_s$ has to be stable.

Appendix B. Numerical parameters

Subsequently, the numerical parameters in section 6 are given in tables B1 and B2. In addition the numerical parameters of a MEMS sensor are given in table B3. It has to be stressed that these parameters are characterizing an actual MEMS sensor.

Table B1. Numerical parameters of the sampler of the frequency domain.

Parameters		Value
Damping	μ	$-800\,\frac{1}{s}$
Cubic damping	η	$3867\,\frac{1}{s}$
Input constant	β	$800000\,\frac{1}{s}$
Minimal frequency	$\omega_{\min}$	$2\pi\cdot 1000\,\frac{1}{s}$
Maximal frequency	$\omega_{\min}$	$2\pi\cdot 3160\,\frac{1}{s}$
Spacing	ω_{Δ}	$2\pi\cdot 180\,\frac{1}{s}$
Number of explorations	n_{ex}^{RF}	$12\,\frac{1}{s}$
Forgetting rate	a	0.01
Sequential sampling time	T_{seq}	5 ms

Table B2. Numerical parameters for the four different scenarios.

Parameters		Scenario 1	Scenario 2	Scenario 3	Scenario 4
Amplitude	ρ_1	$5\cdot 10^{-3}$	$1\cdot 10^{-4}$	$1\cdot 10^{-4}$	$1\cdot 10^{-3}$
	ρ_2	$1\cdot 10^{-3}$	$1\cdot 10^{-4}$	$1\cdot 10^{-4}$	—
Frequency	ω_1	$2\pi\cdot 1540\,\frac{1}{s}$	$2\pi\cdot 1540\,\frac{1}{s}$	$2\pi\cdot 1720\,\frac{1}{s}$	$2\pi\cdot 1720\,\frac{1}{s}$
	ω_2	$2\pi\cdot 2620\,\frac{1}{s}$	$2\pi\cdot 2620\,\frac{1}{s}$	$2\pi\cdot 3700\,\frac{1}{s}$	—
Phase	φ_1	0	0	0	0
	φ_2	$\frac{\pi}{2}$	$\frac{\pi}{2}$	$\frac{\pi}{2}$	—
Convergence number	$m_{conv,1}^{RF}$	4	4	4	4
	$m_{conv,2}^{RF}$	3	3	3	3
	$m_{conv,3}^{RF}$	2	2	2	2
	$m_{conv,4}^{RF}$	1.5	2	2	2
Number of maxima	$n_{max,1}$	1	2	1	1
	$n_{max,2}$	1	2	1	1
	$n_{max,3}$	1	2	2	1
	$n_{max,4}$	1	2	1	1
Standard deviation	σ	—	–	—	$1\cdot 10^{-3}\,\frac{1}{s}$

Table B3. Parameters of a MEMS sensor.

Parameter		Value
Damping	μ	$254.7\ \frac{1}{s}$
	β	$256.4\ \frac{1}{s}$
Time constant	T_1	$10^{-3}\ \frac{1}{s}$
	T_2	$10^{-3}\ \frac{1}{s}$

ORCID iDs

Hermann Folke Johann Rolf 0009-0004-3620-1311
Petro Feketa 0000-0002-7799-5996
Alexander Schaum 0000-0002-2710-1851
Thomas Meurer 0000-0002-9175-2157

References

[1] Abeßer J 2020 *Appl. Sci.* **10** 2020
[2] Zawawi S A, Hamzah A A, Majlis B Y and Mohd-Yasin F 2020 *Micromachines* **11** 484
[3] Abreu Araujo F *et al* 2020 *Sci. Rep.* **10** 328
[4] Lyon R F 1990 Automatic Gain Control in Cochlear Mechanics *The Mechanics and Biophysics of Hearing Lecture Notes in Biomathematics* vol 87, ed S Levin, P Dallos, C D Geisler, J W Matthews, M A Ruggero and C R Steele (Springer) pp 395–402
[5] Lyon R F 2011 Using a cascade of asymmetric resonators with fast-acting compression as a cochlear model for machine-hearing applications
[6] Johannesma P L M 1972 The pre-response stimulus ensemble of neurons in the cochlear nucleus *Symp. on Hearing Theory, 1972* (IPO)
[7] Lopez-Poveda E A and Meddis R 2001 *J. Acoust. Soc. Am.* **110** 3107–18
[8] Slaney M *et al* 1993 Apple Computer, Perception Group *Technical Report* 35
[9] Davis S and Mermelstein P 1980 *IEEE Trans. Acoust. Speech Signal Process.* **28** 357–66
[10] Zolfagharinejad M *et al* 2024 arXiv:2410.10434
[11] Gold T and Gray J 1948 *Proc. R. Soc. B* **135** 492–8
[12] Gold T, Pumphrey R J and Gray J 1948 *Proc. R. Soc. B* **135** 462–91
[13] Bronkhorst A W 2000 *Acta Acust. United Ac.* **86** 117–28
[14] Cherry E C 1953 *J. Acoust. Soc. Am.* **25** 975–9
[15] Hall J E and Hall M E 2020 *Guyton and Hall Textbook of Medical Physiology e-Book* (Elsevier Health Sciences)
[16] Saladin K S and Miller L 1998 *Anatomy & Physiology* (WCB/McGraw-Hill New York)
[17] Holdsworth J *et al* 1988 *Annex C of the SVOS Final Report: Part A: The Auditory Filterbank* 1 1–5
[18] Duke T A J and Jülicher F 2008 Critical oscillators as active elements in hearing *Active Processes and Otoacoustic Emissions in Hearing* (Springer) pp 63–92
[19] Eguíluz V M, Ospeck M, Choe Y, Hudspeth A J and Magnasco M O 2000 *Phys. Rev. Lett.* **84** 5232–5
[20] Kern A and Stoop R 2003 *Phys. Rev. Lett.* **91** 128101
[21] Lerud K D, Kim J C, Almonte F V, Carney L H and Large E W 2019 *Hear. Res.* **380** 100–7
[22] Guckenheimer J and Holmes P 2013 *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* vol 42 (Springer)
[23] Marsden J E and McCracken M 1976 *The Hopf Bifurcation and its Applications* (*Applied Mathematical Sciences*) vol 19 (Springer)
[24] Guo S and Wu J 2013 *Bifurcation Theory of Functional Differential Equations* vol 10 (Springer)
[25] Lenk C, Ekinci A, Rangelow I W and Gutschmidt S 2018 Active, artificial hair cells for biomimetic sound detection based on active cantilever technology *2018 40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 4488–91
[26] Lenk C *et al* 2023 *Nat. Electron.* **6** 370–80
[27] Lenk C *et al* 2020 Enabling adaptive and enhanced acoustic sensing using nonlinear dynamics *2020 IEEE Int. Symp. on Circuits and Systems (ISCAS)* pp 1–4
[28] Rolf H F J and Meurer T 2023 *IFAC-PapersOnLine* **56** 181–6
[29] Karris S T 2003 *Signals and Systems With Matlab Applications* (Orchard Publications)
[30] O'Shaughnessy D 2023 *Speech Commun.* **151** 64–75
[31] Frankel T 2011 *The Geometry of Physics: an Introduction* (Cambridge university press)
[32] Rolf H F J and Meurer T 2024 *IFAC-PapersOnLine* **58** 66–71
[33] Rolf H F J and Meurer T 2024 *Chaos* **34** 103135
[34] Kern A and Stoop R 2011 *Neural Comput.* **23** 2358–89
[35] Rolf H F J, Kumar R and Meurer T 2024 *IFAC-PapersOnLine* **58** 13–18
[36] Rolf H F J, Feketa P and Meurer T 2025 *IEEE Access* **13** 38940–51
[37] Rolf H F J and Meurer T 2025 *IFAC-PapersOnLine* **59** 217–22
[38] Roeser D, Gutschmidt S, Sattel T and Rangelow I W 2016 *J. Microelectromech. Syst.* **25** 78–90
[39] Ivanov T 2004 Piezoresistive cantilevers with an integrated bimorph actuator *PhD Thesis* Citeseer
[40] Ved K *et al* 2024 *AIP Conf. Proc.* **3062** 040011
[41] Ved K, Rolf H F J, Ivanov T, Meurer T, Ziegler M and Lenk C 2025 *Hear. Res.* **462** 109260
[42] Reddy J N 2017 *Energy Principles and Variational Methods in Applied Mechanics* (Wiley)
[43] Gubbi V, Ivanov T, Ved K, Lenk C and Ziegler M 2025 *Sensors Actuators A* **387** 116369
[44] Caliskan V A, Verghese O C and Stankovic A M 1999 *IEEE Trans. Power Electron.* **14** 124–33
[45] Egretzberger M and Kugi A 2010 *Microsyst. Technol.* **16** 777–86

[46] Egretzberger M, Mair F and Kugi A 2012 *Mechatronics* **22** 241–50
[47] Wang X 2004 *Am. Math. Mon.* **111** 525–6
[48] Albach M 2011 *Elektrotechnik* (Pearson Deutschland GmbH)
[49] Green E I 1955 *Am. Sci.* **43** 584–94
[50] Hering E, Martin R and Stohrer M 2007 *Physik für Ingenieure* (Springer)
[51] Rosen S and Fourcin A 1986 *Frequency selectivity in hearing* vol 373487
[52] Kuleshov V and Precup D 2014 arXiv:1402.6028
[53] Mahajan A and Teneketzis D 2008 Multi-armed bandit problems *Foundations and Applications of Sensor Management* (Springer) pp 121–51
[54] Robbins H 1952 Some aspects of the sequential design of experiments
[55] Meddis R and O'Mard L 1997 *J. Acoust. Soc. Am.* **102** 1811–20
[56] Stoop R and Gomez F 2012 *IFAC Proc. Volumes* **45** 70–74
[57] Stoop R 2022 *Front. Netw. Physiol.* **2** 868470