RESEARCH ARTICLE | FEBRUARY 27 2024

Bio-inspired, adaptive acoustic sensor: Sensing properties in dependence of feedback parameters

Kalpan Ved; Claudia Lenk ✉; Tzvetan Ivanov; Philipp Hövel; Martin Ziegler

AIP Conf. Proc. 3062, 040011 (2024)
https://doi.org/10.1063/5.0189488

Bio-inspired, Adaptive Acoustic Sensor: Sensing Properties in Dependence of Feedback Parameters

Kalpan Ved,[1] Claudia Lenk,[1, a)] Tzvetan Ivanov,[1] Philipp Hövel,[2] and Martin Ziegler[1, 3]

[1)]*Department of Micro- and Nanoelectronic Systems, Technische Universität Ilmenau, 98693 Ilmenau, Germany*
[2)]*Department of Electrical and Information Engineering, Christian-Albrechts-Universität zu Kiel, 24143 Kiel, Germany*
[3)]*Institute of Micro- and Nanotechnologies (IMN) MacroNano, Technische Universität Ilmenau, 98693 Ilmenau, Germany*

[a)]*Corresponding author: claudia.lenk@tu-ilmenau.de*

Abstract. Pre-processing of the sound signal during sensing is an integral functionality of the cochlea, the part of human hearing responsible for sound sensing. This pre-pocessing, which is integrated into the sensing stage directly, enables the remarkable properties of human hearing. Similarly, integrating some of these pre-processing functionalities in technological speech processing systems strongly improves their recognition performance.
We developed a bio-inspired, adaptive acoustic sensor with pre-processing capabilities like nonlinear amplification and frequency filtering functionality. The sensor is composed of a single clamped silicon beam with integrated deflection sensing and thermo-mechanical actuation, subjected to a real-time feedback. While the resonance frequency and bandwidth are determined by the geometry of the sensor beam, its transfer characteristics can be switched dynamically from linear to nonlinear regime by changing the feedback parameters. In the linear regime, the feedback controls the sensitivity and bandwidth of the sensors. Here, we elaborate on the influence of the sign of feedback strength and offset on the sensor behaviour. Changing the sign of the feedback parameters switches between amplification and damping behaviour, enabling the change of sensitivity by 44 dB. Thereby, complex oscillation modes are observed for feedback parameters with similar polarity.

INTRODUCTION

The process of human hearing is full of intriguing properties like a dynamic range of 120 dB sound pressure level (SPL), resolving tones only 0.2% apart in frequency and understanding speech in noisy environments [1, 2]. These properties originate partly from the sound capturing and pre-processing capabilities of the cochlea [3]. The cochlea is the organ for sound sensing, and it incorporates a number of signal conditioning and feature extraction steps in the sensing process as well. Thereby, the most important steps are a frequency decomposition and amplification of the input signal, compressive (nonlinear) transfer characteristics and adaptation of amplification and transfer characteristics properties based on local and efferent feedback. These steps take place before and during the transduction process. The nonlinear amplification characteristics are discussed to stem from the sensor system acting as a Hopf oscillator [4, 5]. This hypothesis is further underlined by the fact, that the ear itself can produce tones without sound input, the so-called oto-acoustic emissions [6].

Introducing some of these steps into a technological speech processing system strongly improves the performance of the system, as was recently shown by Araujo et al. [7]. Thereby in particular, the nonlinear transfer characteristics improve the clustering of data, as was tested for spoken digits, which simplifies the classification task for the subsequent neural network. Most speech processing systems nowadays include, after the actual sensing step, a signal conditioning and feature extraction step, typically calculating the envelope and the frequency of the signal and transforming it into a frequency-time representation, often referred to as spectrogram [8]. Thereby, software-based realisations are quite common, but hardware-based implementations, often referred to silicon or FPGA cochlea, exist as well (cf. Refs. [9, 10, 11, 12]). Despite the remarkable increase in performance of speech processing systems, these still struggle in particular with noisy conditions, i.e., low signal-to-noise ratios, or multi-source situations. To address this issue, a bio-inspired approach targets the integration of the signal conditioning and feature extraction step into the sensing process, similar to the operation of the cochlea, rather than after the sensing process, as it is done in current technological realisations.

Several bio-inspired acoustic sensors have been developed, which differ mainly in the integrated functionalities and the number of sensors necessary to cover the frequency range [13]. The most common approach is the integration of frequency decomposition into the sensor system. However, these are the level used of bio-inspired sensors, which has been tested and applied already in speech processing systems [14]. Other approaches of bio-inspired sensors include nonlinear transformation and/or adaptation.

Nonlinearity and Hearing: Advances in Theory and Experiment
AIP Conf. Proc. 3062, 040011-1–040011-10; https://doi.org/10.1063/5.0189488
Published by AIP Publishing. 978-0-7354-4844-5/$30.00

We developed an acoustic sensor which integrates nonlinear transfer characteristics, amplification and frequency filtering functionality into the sensor and incorporates dynamic adaptation [15, 16]. These functionalities are achieved by combining a passive transducer with a high-speed (real-time) feedback. Here, we present our study on the sensing properties in dependence of the sign of the feedback parameters after a short introduction of the sensor system.

BIO-INSPIRED, ACOUSTIC SENSOR SYSTEM

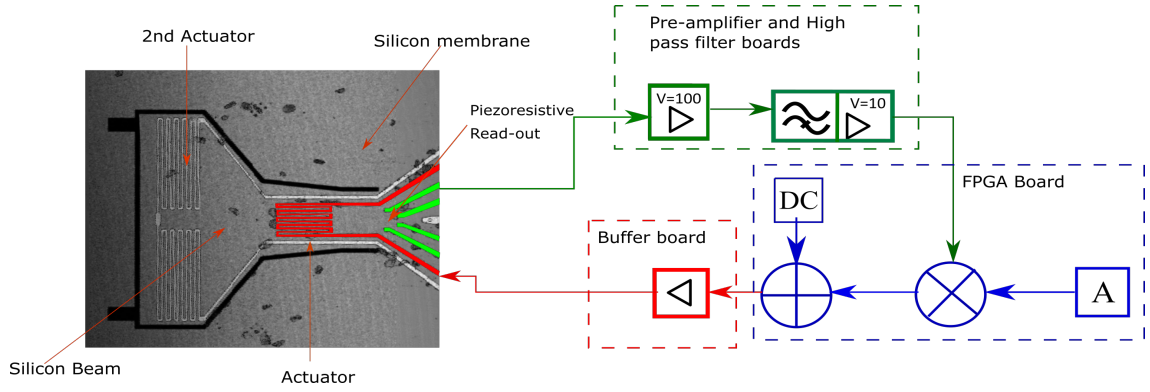


FIGURE 1: Schematic representation of the sensor system with microscope image (left) of silicon-based transducer with integrated thermo-mechanical actuator (red) and piezo-resistive deflection sensor (green) and the feedback algorithm (right) used to tune sensing properties of transducer.

To obtain the above described sensing functionalities, the developed bio-inspired sensor system consists of a transducer in combination with a high-speed feedback, as schematically shown in fig 1. Thereby, a single-side mounted silicon beam with integrated piezo-resistive deflection sensing is used as transducer. The dimensions of the silicon beam are 775 µm−1354 µm (length), 775 µm (width) and $2-10$ µm thickness. The transducer is surrounded by a silicon membrane to decrease the acoustic short circuit effect and thus increase the sensitivity. The beam is operated in resonance mode with a quality factor of $100-200$ and a gain factor of ≈ 3 V/Pa upon excitation with sound [17]. Through the resonant operation of the sensor, it effectively acts as band-pass filter with a bandwidth of $20-100$ Hz, depending on the design. The frequency f of each sensor is determined by its geometrical dimensions:

$$f = \frac{\omega_0}{2\pi} = \delta_n^2 \frac{d_{\mathrm{Si}}}{2\pi l_{\mathrm{Si}}^2} \sqrt{\frac{E_{\mathrm{Si}}}{12\rho_{\mathrm{Si}}}}$$

with l_{Si} the length of the beam, d_{Si} its thickness, E_{Si} the elastic module of silicon, ρ_{Si} the density of Silicon, and δ_n a positive pre-factor, depending on the oscillation mode. Thus, the resonance frequency can be tuned by the length and the thickness. However, the majority is tuned by changing the length of the beam [17]. To cover the audible range and implement frequency decomposition, an array of sensors with different resonance frequencies is necessary. Despite the small bandwidth, a spectrogram-like graph can be obtained from a relatively small number of sensors as shown in figure 2. Therefore, the envelope of the sensor signals, obtained from sensors with different resonance frequencies, are plotted as time series. Each signal is drawn on the y-axis according to the resonance frequency of its beam. Thereby, the sensor signals were simulated using a previously described model of the sensor system [18], which calculates the beam deflection in dependence of the introduced heating of the beam by the applied feedback voltage (see below). The input for the sensor array was a recording of a spoken digit (here: 'zero') [19]. A spectrogram of the same input, obtained from short-time fast Fourier filtering of overlapping time intervals of 5 ms length, is plotted for comparison in figure 2. In both cases, the frequency scale was taken according to the mel-scale with 16 channels for the frequency range of 100 Hz to 4 kHz. As can be seen, the beam-based spectrogram resembles the Fourier-based spectrogram quite well, despite the fixed bandwidth of 50 Hz.

Furthermore, implementation of high-speed feedback, as schematically shown in figure 1, is enabled by the thermo-mechanical actuator integrated into the silicon beam as well as the piezo-resistive read-out for deflection sensing [20,

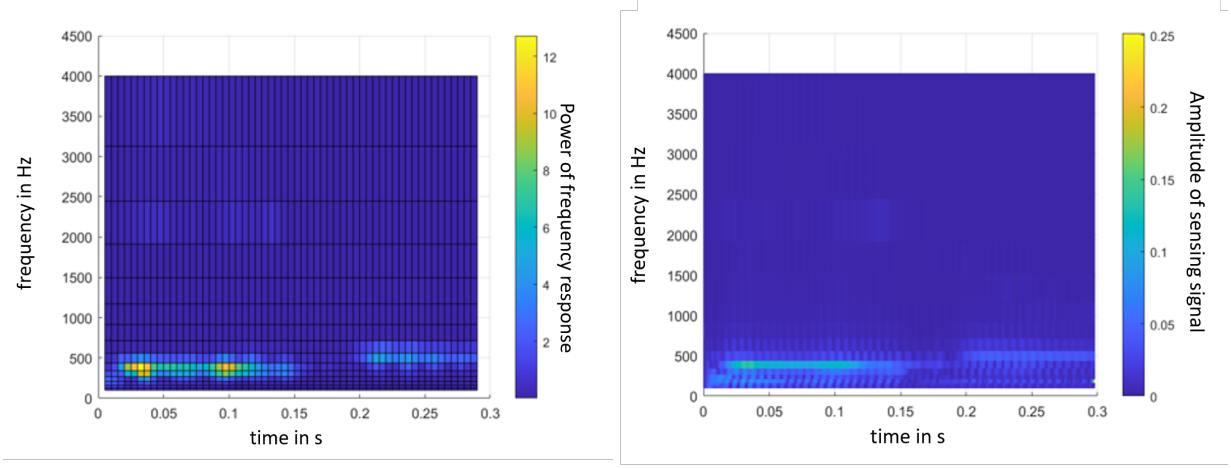


FIGURE 2: Spectrogram of spoken digit 'zero' obtained with short-time fast Fourier filtering (left graph) and from envelope of sensor signal from simulated beam array (right graph). Thereby, the color code refers to the power of the signals for the Fourier-based spectrogram (left) or the amplitude of the simulated sensing signals in the case of the beam array (right). In both cases, 16 channels distributed according the mel-scale were used for the frequency range of 100 Hz to 4 kHz.

21]. The feedback algorithm is given by

$$u_{\mathrm{act}} = a_{\mathrm{f}} u_{\mathrm{sens}} + u_{\mathrm{DC}} \tag{1}$$

and was realized in FPGA and analog circuits [16, 22]. Here, u_{act} refers to the actuation voltage applied to the actuator, u_{sens} is the sensing voltage obtained from the piezo-resistive read-out after amplification and high-pass filtering, a_{f} is the feedback strength and u_{DC} a bias offset. The feedback is used to tune the sensing properties of the system. In particular, one obtains three operation modes of the sensor depending on the feedback strength a_{f} or the bias voltage u_{DC}:

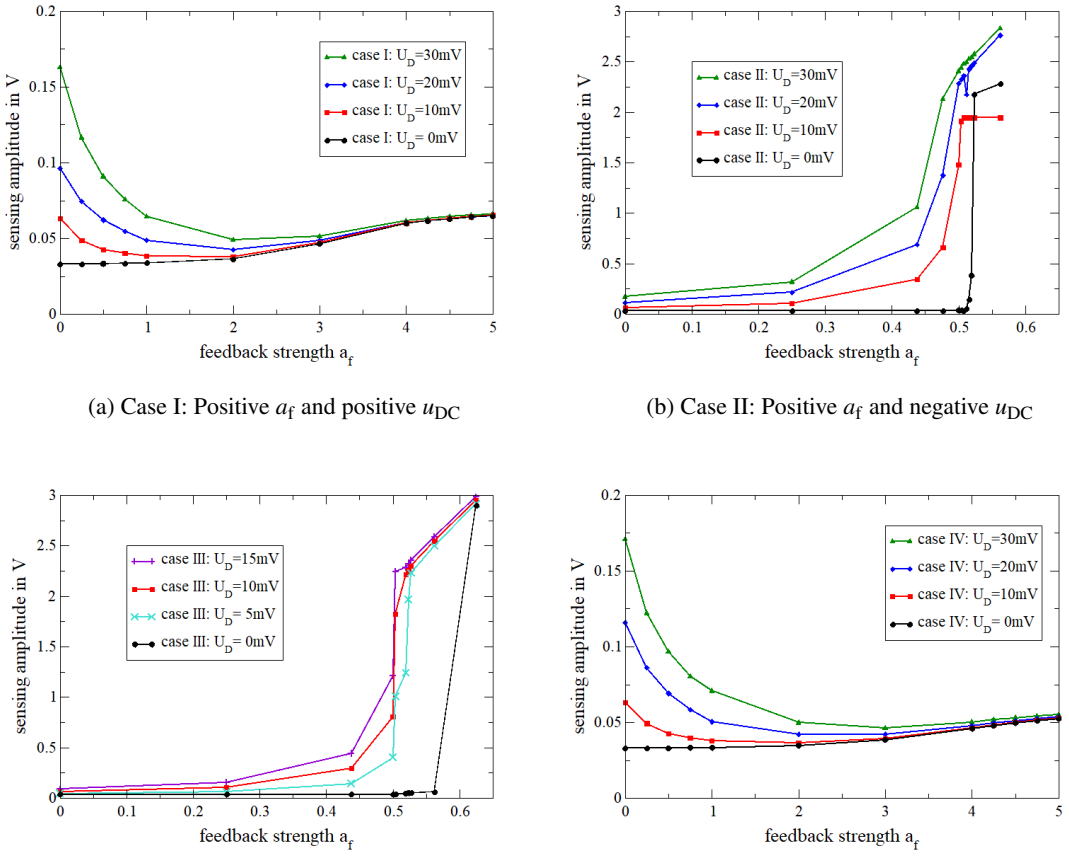
- (i) a passive mode with linear transfer characteristics, if both feedback parameters are zero,
- (ii) an active, linear mode for intermediate feedback strengths or DC values, which is characterised by an increased gain and sensitivity compared to the passive mode and linear transfer characteristics and
- (iii) an active, nonlinear mode for high feedback strengths or small DC values, which exhibits a gain depending on the sound pressure level (nonlinear transfer characteristics).

Further increase of the feedback strength or further decrease of the u_{DC} value will result in self-excited oscillations due to the system undergoing a Hopf bifurcation [18]. Interestingly, measurements with a reference microphone (MT Gefell MV203), which was placed 5.5 cm beside our developed bio-inspired microphone in an anechoic chamber, revealed that these self-excited oscillations produce measurable sounds. This is quite similar to the oto-acoustic emissions from the inner ear, which are thought to occur due to parts of the cochlea acting as Hopf oscillators as well [4]. An additional mechanism implemented in the sensor system is adaptation to the input or the acoustic environment [22]. Thereby, an amplitude-based switching of the feedback strength a_{f} was implemented to model sensor adaptation to constant sound inputs, to increase the dynamic range (for sensing with high gain values) and to enhance the sound onset for subsequent feature extraction.

Summarising, the developed, bio-inspired sensor system demonstrates functionalities observed from the inner ear in human hearing [22, 23]. These are (i) frequency decomposition of the input, (ii) amplification, (iii) nonlinear transfer characteristics, (iv) dynamic adaptation of sensing properties and (v) emission of sounds (without input). Thereby, amplification and nonlinear properties are obtained for positive feedback strength a_{f} and negative offset u_{DC}. However, for the inner ear a frequency-selective reduction of the gain is discussed as well besides the above discussed amplification [24, 25]. It is assumed to arise from damping of the activity of outer hair cells controlled by efferent

feedback. Since outer hair cells are believed to provide the amplification, dampening their activity leads to a reduction of the effective gain of the inner ear. Since outer hair cells are activated by the basilar membrane, which performs a frequency decomposition of the sound signal by creating a frequency-place coding, the amplification and damping is specific for each frequency bands. This effect is thought to enhance speech processing in noisy situations by reducing the response for the noisy signals [26]. Here, we study if similar effects, like frequency-selective amplification and dampening of the sensor response are possible by investigating the dependence of the dynamic and sensing properties in dependence of the sign of the feedback parameters a_f and u_DC.

DEPENDENCE OF SENSOR SYSTEM PROPERTIES ON SIGN OF FEEDBACK PARAMETERS



(a) Case I: Positive a_f and positive u_DC

(b) Case II: Positive a_f and negative u_DC

(c) Case III: Negative a_f and positive u_DC

(d) Case IV: negative a_f and negative u_DC

FIGURE 3: Amplitude of sensing signal in dependence of feedback strength a_f for different driving voltages of the loudspeaker (see legends), corresponding to different SPL.The microbeam has a resonance frequency of 3.63 kHz. The panels represent the four different feedback cases, as summarized in table I.

Since the feedback signal u_act depends on a_f and u_DC (see eq. 1), , four possible combinations of the signs of the feedback parameters are possible, i.e. both parameters positive, both parameters negative and two cases with one negative and one positive. However, since we use a thermo-mechanical actuation principle, we expect only two different cases, as explained in the following. For thermo-mechanical actuation, the feedback signal is used to heat the beam due the resistance of the aluminium wire on top of the beam, at which the feedback voltage is applied.

This heating results in a beam deflection based on the bimorph effect [21]. Since the electrically induced temperature increase of the beam is linearly depending on the introduced electrical power, the effective driving signal for the beam, i.e., the power of the electric signal, is proportional to the square of the feedback signal

$$P_h \approx (a_\mathrm{f} u_\mathrm{sens} + u_\mathrm{DC})^2 = a_\mathrm{f}^2 u_\mathrm{sens}^2 + 2 a_\mathrm{f} u_\mathrm{sens} u_\mathrm{DC} + u_\mathrm{DC}^2.$$

Thus, combinations of similar signs of a_f and u_DC will result in only positive signs of the introduced power, whereas opposite signs of a_f and u_DC will yield a negative term.

To study for the four different feedback cases (summarized in table I) the dependence of sensor properties on the sign of feedback parameters, the response of a sensor with a resonance frequency of 3.63 kHz to acoustic excitation was measured. The results are shown in figure 3 (a)-(d). Thereby, a single tone signal of 3.63 kHz, generated by a signal generator, was applied to drive the loudspeaker. Different driving voltages are used from $U_D = 0$ mV (no sound) to 30 mV, corresponding to different sound pressure levels (SPL). Note that the sound pressure is linearly depending on the amplitude of the driving signal. For each combination of feedback strength a_f and amplitude of the driving signal U_D, one time series was recorded and the maximal amplitude of the sensing signal was determined, occurring at the resonance frequency. In all cases $|u_\mathrm{DC}| = 200$ mV was used.

As discussed above, it is clearly visible from Fig. 3 that always two feedback cases yield similar responses. The case $sign(a_\mathrm{f}) = sign(u_\mathrm{DC})$ yields a decrease of amplitude with increasing feedback strength, whereas the case of opposite signs leads to an increase of sensing amplitude with increasing feedback strength a_f. The initial sensitivity, given by the slope of $u_\mathrm{sens}/U_\mathrm{D}$, is ≈ 5 for the passive mode in all four cases. It increases up to ≈ 80 for $a_\mathrm{f} = 0.5$ for the cases II and III (opposite signs) and decreases to ≈ 0.5 for $a = 2$ for the cases I and IV (similar signs). This corresponds to a total change in sensitivity by 44 dB. Thus by changing the sign of the feedback parameters, we can control the sensitivity of the sensor in a wide range. This enables not only to increase the sensor gain, as was previously shown (see e.g. [27]), but also to dampen the sensors response. The latter effect can be used to implement frequency-selective and sound processing based damping of unwanted signals, as discussed for the inner ear in the frame of efferent feedback [24, 25].

Besides this, for the cases II and III with opposite signs, a strong increase for a certain feedback strength ($a_\mathrm{f} \approx 0.55$) is observed. This is the indication of the Hopf bifurcation [16, 18]. As a result of this, self-excited oscillations with sine wave shape are observed for larger feedback strengths with a frequency of 3.63 kHz corresponding to the resonance frequency of the beam. For the cases with similar signs, this strong increase in amplitude is observed as well, if very large feedback strengths are considered. This is visible from figure 4, which shows results from simulations of the sensor system for the cases III and IV.

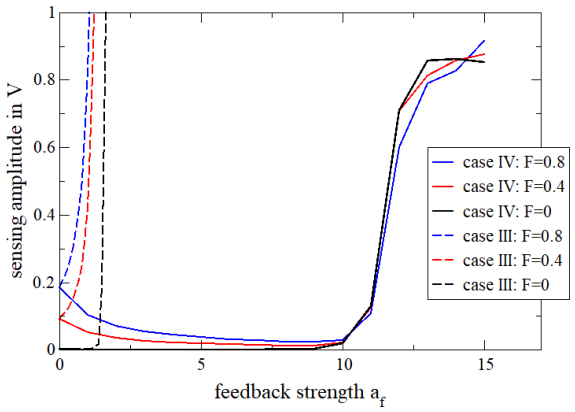
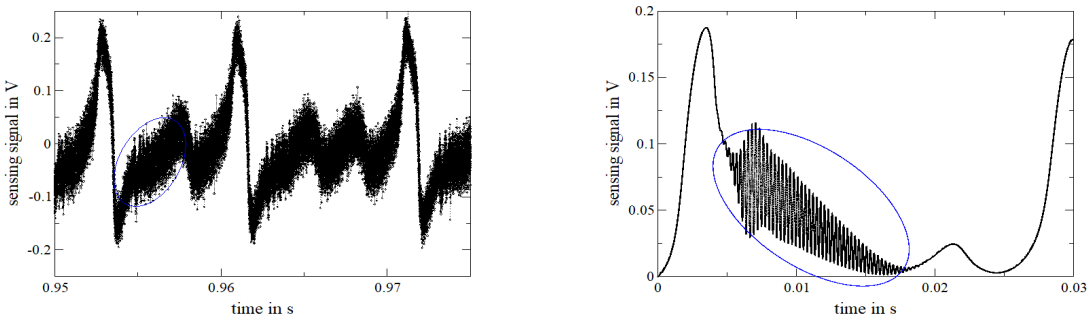


FIGURE 4: Amplitude of sensing signal in dependence of feedback strength a_f for case III and case IV obtained from numerical simulations of a beam with resonance frequency of 3.63 kHz.

Thereby, the feedback strength at bifurcation, i.e., the critical feedback strength a_c, is much larger for case IV (similar signs) than for case III (opposite signs). The strong increase in amplitude observed for $|a_\mathrm{f}| > a_c$ in fig. 4 corresponds to the onset of self-excited oscillations, similar to case II and III. Thereby, particularly interesting is the



(a) Case IV: Measured time series for $a_f = -5$ and $u_{DC} = -200\,\text{mV}$.

(b) Case II: Simulated time series $a_f = 11$ and $u_{DC} = 200\,\text{mV}$.

FIGURE 5: Time series of sensing signal for case IV obtained from measurements and for case II obtained from numerical simulations. In both cases, two oscillations are visible: a slow one with a frequency of ≈ 100 Hz and a fast one (marked by blue ellipse), on the slope of the slow one, with a frequency 3.63 kHz corresponding to the first mode response of the cantilever.

shape of the self-excited oscillations in case IV and II (see figure 5). In these cases, a complex oscillatory pattern is observed, composed of a slow oscillation ($\approx 100\,\text{Hz}$) with large amplitude and a faster oscillation (3.63 kHz) with small amplitude, visible in the rise/fall of the slow oscillation. However, if this behaviour is due to a Hopf bifurcation as well (as in the cases II and III), and if these oscillations correspond to a limit cycle and their nature requires further research.

SUMMARY

We presented a bio-inspired acoustic sensor that provides several functionalities of the cochlea. These are frequency decomposition and amplification of the sound signal, nonlinear transfer characteristics and dynamic adaptation of sensing properties to the input and acoustic environment. The presented sensor design can help solve problems that speech processors currently have, for instance, in changing acoustic environment with low signal-to-noise ratios (*cocktail party problem*). There has been a rapid development in terms of technological transfer to sound processing systems that make use of neural networks. However, pre-processing steps are usually applied to the microphone signal before feeding it into the neural networks, and feature representation is critically affected by perturbations to the signal. Integrating this bio-physical pre-processing can strongly enhance the performance, e.g., in multiple-source problems [7, 28]. The presented, bio-inspired, adaptive acoustic sensor integrates the signal processing functionality already at the sensing stage. Those functionalities are obtained by resonant operation of the transducer in combination with a high-speed feedback. The transducer is a silicon beam with integrated actuation and deflection sensing. The feedback loop gives precise control for the sensitivity and the linearity of the sensor. The polarity of the feedback parameters a_f and u_{DC} determines if the sensitivity and gain are increased (opposite sign of feedback parameters) or decreased (similar sign of feedback parameters) compared to the passive case without feedback, as summarized in table I. Combining both polarities, the sensitivity can be tuned in a large range by up to 44 dB. A benchmark-

TABLE I: Feedback cases in dependence of the sign of feedback strength a_f and offset u_{DC} and their effect on the gain of the sensor system for excitation by sound.

sign of a_f \ sign of u_{DC}	positive	negative
positive	case I: decrease of gain with feedback (damping)	case II: increase of gain with feedback (amplification)
negative	case III: increase of gain with feedback	case IV: decrease of gain with feedback

ing of the sensitivity and other characteristics can be found in Ref. [29]. If the feedback parameters cross a certain threshold, self-excited oscillations occur, which can be addressed to the system undergoing a Hopf bifurcation. Interestingly, in the case of opposite polarity, the self-excited oscillations exhibit a sine-wave shape and the sensor can emit sound. This resembles the oto-acoustic emissions of the inner ear. However, in the case of similar polarity of the feedback parameters, the self-excited oscillations exhibit a complex dynamic behaviour, which contains slow and fast oscillations.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) - Project-ID 434434223 – SFB 1461 and the Carl-Zeiss-Stiftung in the project 'Memristive Materials for Neuromorphic Engineering (MemWerk)'.

COMMENTS AND QUESTIONS

[Online forum]

Brian Frost: Really great work! I appreciate the idea of implementing compressive nonlinearity on the sensor side rather than the processor side. Two comments/questions: 1) I feel that this work could greatly benefit from a figure clearly comparing the output of this sensor to the output of another sensor and, for example, the compound axon potential or basilar membrane response. This would really drive home the fact that this system better represents the features of the cochlea than other sensors; 2) I feel that I need a bit more convincing that this design will help solve the problems that speech processors currently have – specifically, the cocktail party problem or multiple source problem mentioned in the introduction comes to mind. Even just a few sentences in the conclusion would help to make the value of this device clear.

Author: To demonstrate the benefits of the sensor compared to a standard microphone, the sensor was exposed to the natural sound database (consisting of ferret calls, speech or water flowing etc.). Here, the blue curve (fig. 6) shows the exiter signal for the loudspeaker, which closely correlates to the response of a conventional microphone. The red curve shows the response of the bio-inspired sensor without feedback, which responds only to its resonance frequency at 5.19 kHz. This is comparable to the signal of a microphone after bandpass-filtering. If the feedback strength a is increased to slightly smaller than the critical value, the signal-to-noise ratio is strongly improved and quiet sounds (e.g. at $t \approx 9.4 - 9.5$ s) can now be resolved.

Indeed, there has been a rapid development in terms of technological transfer to sound processing systems that make use of, e.g., convolutional, recurrent, and spiking neural networks. This has led to an increased performance for tasks like keyword spotting, speaker identification, or speech analysis. Then, pre-processing steps are usually applied to the microphone signal before feeding it into the neural networks. However, in tasks like the mentioned cocktail party or multiple source problem, feature representation is critically affected by perturbations to the signal (room reverberation, interfering noise etc.), in particular, at low signal-to-noise ratios. The presented, bio-inspired, adaptive acoustic sensor integrates the signal processing functionality already at the sensing stage. We argue that the dynamical switching between linear and nonlinear characteristics improves detection of signals in noisy conditions, increases the dynamic range of the sensor, and enables adaptation to changing acoustic environments.

As suggested, we expanded the Conclusion section to highlight the benefits of the proposed dynamic MEMS device, including additional references.

Jonathan Siegel: I had a reaction similar to Brian's. It wasn't apparent how this device would perform better at a meaningful task than alternatives. What are you benchmarking performance against? Spontaneous oscillation of a device intended to detect external acoustic signals is not generally a good thing. The inner ear can generate spontaneous OAEs, but these are generally at a very low level and don't impair encoding of external sounds. Nonlinearity may make detecting signals in noise even harder than linear amplification, so this potential application would need to be tested explicitly.

Author: For comparison with human (mammalian) and artificial cochleae as well as other bio-inspired sensors, we performed a benchmarking that considers a number of quantities. These include frequency range, self-noise, maximal sound-pressure level, size of sensor, frequency decomposition, gain change to mention but a few. The results can be found in a paper that has been accepted and is currently at the production stage [29]. In brief, we found that (i) we can obtain a change of gain of up to 44 dB, (ii) feedback and adaptation algorithms have a

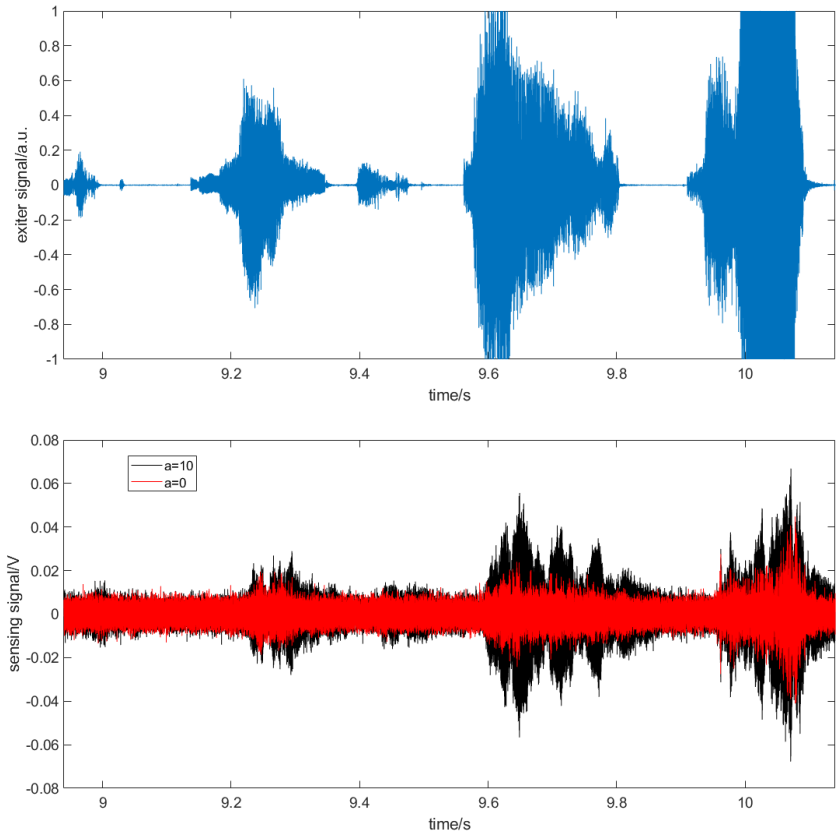


FIGURE 6: Comparison of exciter signal for loudspeaker (top, blue), which closely correlates to conventional microphone recording of this sound, and MEMS device without (bottom, red) and with (bottom, black) feedback when exposed to a natural sound (consisting of ferret calls, speech or water flowing etc.) [30].

small circuit overhead per channel (that is, beam), and (iii) a high resilience against device tolerances is due to the simplicity of the feedback loop and its multiple control parameters. For the readers' convenience, we added a sentence in the Conclusion section.

According to hypothesis, Oto-Acoustic Emissions (OAEs) are generated because the inner ear is a critically tuned oscillator near a Hopf bifurcation point [31]. Spontaneous oscillations are shown to describe the analogy between our sensor and inner ear. Nevertheless, encoding of the sound is done before the bifurcation point, i.e. before the onset of the spontaneous oscillations regime.

Concerning the question of nonlinearity and noise, we would like to stress that frequency decomposition reduces noise due to bandpass filtering properties. Furthermore, the critically tuned Hopf oscillator induces compressive nonlinearity, thus making it a small signal amplifier [32]. Thereby, the signal-to-noise ratio is improved for small signals in noise, as can be seen from fig. 7. Hence, this cumulative effect has the potential to improve the signal-to-noise ratio. This is in agreement with previous studies that showed how nonlinearity can help in noisy condition, see e.g. [33, 34, 35].

[Post-talk Q&A]

Audience member: The system you presented is in my understanding basically a feature extraction stage for automatic speech recognition systems rather than a sensor. Have you tried to connect this with convolutional neural networks to do ASR tasks and benchmark e.g. word error rate? Are you planning to do such benchmarking?
Author: We are currently working on this, i.e., using the systems output as input for neural network-based speech recognition systems, but cannot yet present definite results on word error rate or similar. Feature extraction

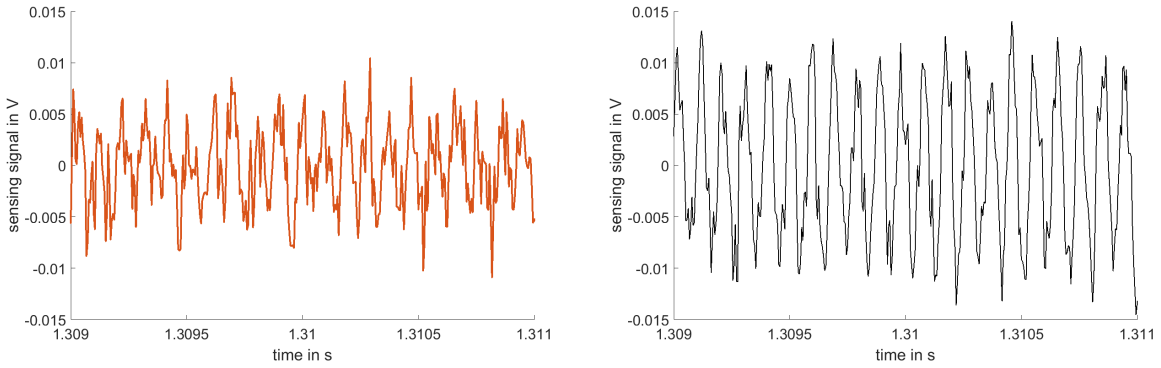


FIGURE 7: Sensing signal either without feedback (left) or with feedback (right) to demonstrate improvement of signal-to-noise ratio by feedback operation.

is typically used to improve the clustering of data so that recognition of classes by the network is improved. First simulations of our system demonstrate an increase in clustering, but the data were not yet applied to neural networks. That bio-inspired pre-processing is indeed increasing accuracy and reducing error rate was already shown for other systems [7, 14].

Audience member: How fast is the thermomechanical actuation?
Author: Thermomechanical actuation can be used for AC signals up to 1 MHz [20, 36], although the static thermal time constant is roughly 30 μs. This high frequency can be achieved, since the dimensions of the beams, in particular the thickness, are small and only small fluctuations of the peak-to-peak value of the temperature, in the range of 10%, are needed to drive the beam bending.

Audience member: I can use my Iphone for dictating messages with high word rates and approximately less than 1%error. On a conference, I was trying to talk to a person using a cochlea-implant during dinner, i.e. in a very noisy environment. Direct talking and understanding was not possible. Thus, we talked into the Iphone and it converted the speech into text. This worked fine. Is this not in contradiction to the statement that speech recognition systems struggle with noisy environments? Why not just use a microphone or mobile phone everywhere, since everyone has a mobile phone?
Author: That the Iphone works well in these situations is due to multi-microphone recording (2-3 microphones per mobile phone), which allows to do background noise cancellation and other noise suppression algorithm (e.g. using directionality) to pre-process the signals before feeding them into speech recognition stage. Thus, the signal-to-noise ratio of the pre-processed signal is much better than that of unprocessed one. However, this requires multi-mic recordings and (partially) extensive algorithms for pre-processing. Depending on the application, exploiting these algorithms is not always feasible/possible, due to restrictions regarding power consumption and computation power (e.g. hearing aids), direction of important sound source not clear (or changing) (e.g. machine supervision in production sites), and strongly changing hearing conditions (echos, reverberations etc). For these purposes, a low-power, efficient and adaptive feature extraction integrated into the sensor might be helpful.

REFERENCES

1. E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of Neural Science* (McGraw-Hill, 2012).
2. M. A. Ruggero, S. S. Narayan, A. N. Temchin, and A. Recio, "Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla," Proceedings of the National Academy of Sciences **97**, 11744–11750 (2000).
3. A. Hudspeth, "Integrating the active process of hair cells with cochlear function," Nature Review Neuroscience **15**, 600 (2014).
4. A. J. Hudspeth, F. Jülicher, and P. Martin, "A critique of the critical cochlea: Hopf—a bifurcation—is better than none," Journal of Neurophysiology **104**, 1219–1229 (2010).
5. T. A. J. Duke and F. Jülicher, "Critical oscillators as active elements in hearing," in *Active Processes and Otoacoustic Emissions in Hearing* (2008) p. 63.

6. D. T. Kemp, "Stimulated acoustic emissions from within the human auditory system," The Journal of the Acoustical Society of America **64**, 1386–1391 (1978).
7. F. Abreu Araujo, M. Riou, J. Torrejon, S. Tsunegi, D. Querlioz, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, "Role of non-linear data processing on speech recognition task in the framework of reservoir computing," Scientific Reports **10**, 1 (2020).
8. J. L. Flanagan, "Speech synthesis," in *Speech Analysis Synthesis and Perception* (Springer, 1972) pp. 204–276.
9. J. Lazzaro and C. Mead, "Circuit models of sensory transduction in the cochlea," in *Analog VLSI Implementation of Neural Systems* (1989) p. 85.
10. T. J. Hamilton, C. Jin, A. van Schaik, and J. Tapson, "An active 2-D silicon cochlea," IEEE Transactions on Biomedical Circuits and Systems **2**, 30–43 (2008).
11. S.-C. L. (Editor), T. D. (Co-Editor), G. I. (Co-Editor), A. Whatley, and R. Douglas, *Event-based neuromorphic systems* (John Wiley & Sons, 2015).
12. A. Jiménez-Fernandez, E. Cerezuela-Escudero, L. Miro-Amarante, M. J. Domínguez-Morales, F. de Asís Gómez-Rodríguez, A. Linares-Barranco, and G. Jiménez-Moreno, "A binaural neuromorphic auditory sensor for FPGA: A spike signal processing approach," IEEE Transactions on Neural Networks and Learning Systems **28**, 804–818 (2017).
13. C. Lenk, K. Ved, S. Durstewitz, T. Ivanov, M. Ziegler, and P. Hövel, "Bio-inspired (neuromorphic) acoustic sensing," in *Springer Series in Bio- and Neurosciences* (Springer, 2022).
14. H. S. Wang, S. K. Hong, J. H. Han, Y. H. Jung, H. K. Jeong, T. H. Im, C. K. Jeong, B.-Y. Lee, G. Kim, C. D. Yoo, and K. J. Lee, "Biomimetic and flexible piezoelectric mobile acoustic sensors with multiresonant ultrathin structures for machine learning biometrics," Science Advances **7**, eabe5683 (2021).
15. C. Lenk, S. Gutschmidt, and I. W. Rangelow, "Vorrichtung und Verfahren zur Detektion von Schall in Gasen und Flüssigkeiten," Patent **DE 102018117481B8** (2019).
16. C. Lenk, L. Seeber, M. Ziegler, P. Hövel, and S. Gutschmidt, "Enabling adaptive and enhanced acoustic sensing using nonlinear dynamics," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, 2020) pp. 1–4.
17. C. Lenk, T. Ivanov, V. Gubbi, K. Ved, M. Ziegler, T. Fritsch, J. Küller, and D. Beer, "Bio-inspired, nonlinear and adaptive acoustic sensing - study of sensor design," in *Fortschritte der Akustik - DAGA2022* (2022) pp. 1–4.
18. N. Lam, S. Hayashi, and S. Gutschmidt, "A novel mems sensor concept to improve signal-to-noise ratios," International Journal of Non-Linear Mechanics **139**, 103863 (2022).
19. Z. Jackson, C. Souza, J. Flaks, Y. Pan, H. Nicolas, and A. Thite, "Jakobovski/free-spoken-digit-dataset: v1.0.8," (2018).
20. I. W. Rangelow, T. Ivanov, A. Ahmad, M. Kaestner, C. Lenk, I. S. Bozchalooi, F. Xia, K. Youcef-Toumi, M. Holz, and A. Reum, "Review article: Active scanning probes: A versatile toolkit for fast imaging and emerging nanofabrication," Journal of Vacuum Science & Technology B **35**, 06G101 (2017).
21. T. Ivanov, "Piezoresistive cantilevers with an integrated bimorph actuator," Thesis (2003), urn:nbn:de:hebis:34-1153.
22. S. Durstewitz, C. Lenk, and M. Ziegler, "Bio-inspired acoustic sensor with gain adaptation enhancing dynamic range and onset detection," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)* (2022).
23. C. Lenk, K. Ved, S. Gutschmidt, T. Ivanov, P. Hövel, T. Meurer, and M. Ziegler, "Dynamically adaptable acoustic sensor with nonlinear filtering functionality," in *47th international conference on Micro and Nano Engineering* (2021).
24. A. B. Elgoyhen, "Cochlear efferent innervation: function, development and plasticity," Current Opinion in Physiology **18**, 42–48 (2020).
25. P. Dallos, A. Popper, and R. Fay, *The cochlea* (Springer, 1996).
26. J. J. Guinan, "Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses," Hearing Research **362**, 38–47 (2018), annual Reviews 2018.
27. C. Lenk, L. Seeber, and M. Ziegler, "Tuning acoustic sensing properties of mems cantilever by nonlinear operation," in *Mikro-Nano-Integration; 8th GMM-Workshop* (2020) pp. 85–87.
28. W. Hemmert and H. Wang, "Offset adaptation in the inner hair cell synapses is important for speech coding." in *Proc. 31st ARO Midwinter Meeting* (2008).
29. C. Lenk, P. Hövel, V. Kalpan, S. Durstewitz, T. Meurer, T. Fritsch, A. Männchen, J. Küller, D. Beer, T. Ivanov, and M. Ziegler, "Neuromorphic acoustic sensing with the dynamic MEMS cochlea: a highly adaptive, nonlinear and frequency-filtering acoustic sensor," Nature Electronics (2023), in print.
30. J. Schnupp, "Network receptive fields-natural sounds database," (2016).
31. T. A. Duke and F. Jülicher, "Critical oscillators as active elements in hearing," in *Active processes and otoacoustic emissions in hearing* (Springer, 2008) pp. 63–92.
32. K. Wiesenfeld and B. McNamara, "Period-doubling systems as small-signal amplifiers," Physical Review Letters **55**, 13 (1985).
33. M. Akilli, N. Yilmaz, and K. Gediz Akdeniz, "Automated system for weak periodic signal detection based on Duffing oscillator," IET Signal Processing **14**, 710–716 (2020).
34. S. Tiwari and R. N. Candler, "Using flexural mems to study and exploit nonlinearities: A review," Journal of Micromechanics and Microengineering **29**, 083002 (2019).
35. A. Hajjaj, N. Jaber, S. Ilyas, F. Alfosail, and M. I. Younis, "Linear and nonlinear dynamics of micro and nano-resonators: Review of recent advances," International Journal of Non-Linear Mechanics **119**, 103328 (2020).
36. T. Michels and I. W. Rangelow, "Review of scanning probe micromachining and its applications within nanoscience," Microelectronic Engineering **126**, 191–203 (2014).