Design of a Silicon Cochlea System with Biologically Faithful Response

Shiwei Wang*, Thomas Jacob Koickal[†], Godwin Enemali[‡], Luiz. Gouveia[§], Lei Wang* and Alister Hamilton[‡]
*Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. (Email: wangsw@siat.ac.cn)

†Beach Theory, Trivandrum, India.

‡Inst. for Integrated Micro and Nano Systems, University of Edinburgh, Edinburgh, UK. (Email: alister.hamilton@ed.ac.uk) §School of Engineering, University of Glasgow, Glasgow, UK.

Abstract-This paper presents the design and simulation results of a silicon cochlea system that has closely similar behavior as the real cochlea. A cochlea filter-bank based on the improved three-stage filter cascade structure is used to model the frequency decomposition function of the basilar membrane; a filter tuning block is designed to model the adaptive response of the cochlea; besides, an asynchronous event-triggered spike codec is employed as the system interface with bank-end spiking neural networks. As shown in the simulation results, the system has biologically faithful frequency response, impulse response, and active adaptation behavior; also the system outputs multiple band-pass channels of spikes from which the original sound input can be recovered. The proposed silicon cochlea is feasible for analog VLSI implementation so that it not only emulates the way that sounds are preprocessed in human ears but also is able match the compact physical size of a real cochlea.

I. INTRODUCTION

The concept of silicon cochlea was introduced in late 1980s by Lyon and Mead [1]. In such systems, analog very large scale integrated (VLSI) circuits are used to replicate the hydrodynamics and neuromorphology of the biological cochlea [2]–[5]. By virtue of its bio-inspired nature, the neuromorphic silicon cochlea systems are able to perceive sounds in a similar way as human ear does, processing sound in real time and consuming only tens of milliwatts power. With these advantages, the silicon cochlea systems are recently becoming popular in intelligent machine audition [6], [7] and cochlea prosthesis [8] applications.

Nevertheless, there remain significant discrepancies between the response of existing silicon cochlea systems and the real biological cochlea. The physiological experiment in [9] shows that the biological cochlea has a sharp tuning of frequency selectivity, with the -10 dB quality factor (Q_{10}) reaching maximally over 6 when sound stimuli is weak, and the stop-band roll-off slope of the biological cochlea frequency response is as steep as over 300 dB/dec. The sharp frequency selectivity tuning and steep roll-off slope are key factors contributing to the cochlea's highly discriminative sound spectrum analysis and spectral decomposition capabilities. The silicon cochlea developed by Wen is thus far the most faithful replication of the cochlea hydrodynamics and anatomical structure, which not only models the 2-dimensional wave propagation in the cochlea fluid but also incorporates the active coupling effect between neighboring basilar membrane (BM) segments [5]; nevertheless, the maximal Q_{10} factor and steepest roll-off slope measured from the chip are respectively 2.7 and 179.4 dB/dec, which are still inferior compared with

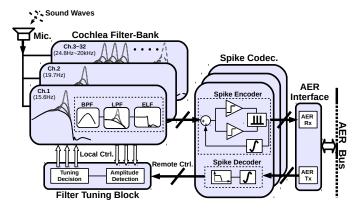


Fig. 1. Block diagram of the proposed silicon cochlea system

the specifications measured from the biological cochlea. Besides, the existing silicon cochleas only focus on modeling the cochlea hydrodynamics and neuromorphology, while limited efforts have been made to replicate the full cochlea response characteristics such as group delay, impulse response and critical bandwidth etc.

Under this background, a couple of research groups have attempted to employ high-order pole-zero CMOS filters to build better models of the BM which is the frequency decomposition organ of the cochlea. Katsiamis developed a CMOS cochlea channel based on an 8^{th} -order eight-pole one-zero analog filter which for the first time achieves the bio-realistic sharp frequency tuning and steep stop-band roll-off [10]. Recently, we developed and reported a 9^{th} order nine-pole five-zero CMOS analog filter which not only manifests the bio-realistic frequency tuning and roll-off slope but also resemble many more aspects of the cochlea response characteristics [11].

In this paper, we introduce the design of a silicon cochlea system on the basis of the CMOS cochlea filters reported in [11]. Section. II describes the system architecture. Section. III explains the design of the building blocks of the proposed system. Section. IV presents the key simulation results we have obtained. The paper is summarized and concluded in Section. V

II. SYSTEM ARCHITECTURE

Fig. 1 illustrates the block diagram of the proposed silicon cochlea system. The architecture is a close replication of the system-level structure of the biological cochlea. The sound

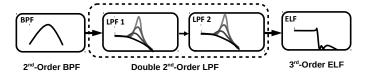


Fig. 2. Modified architecture of a single cochlea filter channel.

signal collected by the microphone is converted into multiple band-passed segments by a cochlea filter-bank that consists of 32 channels with center frequencies located in one-third octave distribution to cover the audio band. The cochlea filter-bank functions as a counterpart of the BM in biological cochlea and its response is controlled locally by a filter tuning block which, similarly as the out hair cell (OHC) does, adjust the filter gain and selectivity adaptively according to the sound input intensity. An array of event-driven spike time codecs encodes the filter outputs into a spike train which is transmitted onto the address-event-representation (AER) bus for inter-chip communication with back-end neuromorphic processors. Also, mimicking the brain-controlled loop that alters OHC response for selective attention, the spike codecs receive and decode feedback signals from the AER bus to perform remote control on the filter tuning block.

III. DESIGN OF BUILDING BLOCKS

A. Cochlea Filter Channel

As proposed in [11], each channel of the cochlea filterbank consists of three cascaded filter stages, namely the bandpass filter (BPF), the low-pass filter (LPF) and the elliptic filter (ELF). As proven in the VLSI hardware measurement results, this method contributes to a filter frequency response that resembles many of the real cochlea characteristics such as the sharp roll-off slope, the variable and selective midband and the constant low-frequency tail, etc. However, since the LPF which models the variable and selective mid-band response is only $2^{nd} - order$, its damping factor has to be reduced to an extremely low value in order to match the high Q factor in physiological results. Consequently, the measured maximum Q factor and gain data reported in [11] fall short of the simulation results reported in [12]. Besides, the measured impulse response shows a far longer settling time compared with physiological results due to the excessively low filter damping factor.

Therefore, in this design we modify slightly the filter architecture proposed in [11] and the new architecture is illustrated in Fig. 2. The 2^{nd} -order BPF is used to model the constant and gentle low-frequency tail in the same way as that in [11]; however, two identical 2^{nd} -order LPFs instead of one are used to model the variable and selective mid-band response; besides, a 3^{rd} -order instead of 5^{th} -order ELF is used to model the sharp roll-off slope. In this architecture, the total filter order remains the same as that in [11] while the filter sharp roll-off slope is traded off for a higher midband selectivity. As will be shown in the simulation results, this trade-off strategy optimizes the overall filter performance since the reduced 3^{rd} -order ELF is still sufficient to maintain a bio-comparable roll-off slope, and also the filter now matches the cochlea selectivity using two 2^{nd} -order LPFs with heavier damping so that the instability risk is lowered and the filter

impulse response becomes more faithful with that of the real cochlea.

The transfer function of the cochlea filter channel is defined as

$$H(s) = \underbrace{\frac{(s + \omega_z)\omega_B}{s^2 + \frac{\omega_B}{Q_B}s + \omega_B^2}}_{\text{BPF}} \cdot (\underbrace{\frac{\omega_L^2}{s^2 + \frac{\omega_L}{Q_L}s + \omega_L^2}}_{\text{LPF}})^2 \cdot \underbrace{\frac{0.0692\omega_E \cdot (s^2 + 7.61\omega_E^2)}{(s + 0.524\omega_E)(s^2 + 0.455\omega_E \cdot s + 1.01\omega_E^2)}}_{\text{ELF}} (1)$$

where ω_z is the zero of the BPF, ω_B , ω_L , and ω_E represent the natural frequency of the BPF, LPF, and ELF respectively, Q_B and Q_L represent the quality factor of the BPF and LPF respectively. The transfer function of the ELF part in Eq. 1 is derived based on the design parameters for the 3^{rd} -order elliptic filter with 1 dB pass-band ripple and 40 dB stop-band attenuation [13].

The above-mentioned parameters are designed according to the physiological data in [9]:

- Q_B is set at 1 according to the -3 dB quality factor of the cochlea frequency response under strong stimulus condition (passive quality factor);
- ω_z is set equal to 0.05 ω_B according to the relationship between DC gain and peak gain under strong stimulus condition;
- ω_L and ω_E are set equal to 1.5 ω_B according to the center frequency position under weak stimulus condition (active center frequency);
- ω_B is used as a programmable parameter denoted by ω_0 that defines the filter channel's center frequency under strong stimulus condition (passive center frequency);
- Q_L is used as an adjustable parameter denoted by $1/\beta$ that defines the filter channel's quality factor under various stimulus condition (active quality factor).

Therefore, the channels transfer function is rewritten as

$$H(s) = \underbrace{\frac{(s+0.05\omega_0)\omega_0}{s^2 + \omega_0 s + \omega_0^2}}_{\text{BPF}} \cdot \underbrace{(\frac{2.25\omega_0^2}{s^2 + 1.5\beta\omega_0 s + 2.25\omega_0^2})^2}_{\text{LPF}} \cdot \underbrace{\frac{0.104\omega_0 \cdot (s^2 + 17.1\omega_0^2)}{(s+0.786\omega_0)(s^2 + 0.682\omega_0 \cdot s + 2.26\omega_0^2)}}_{\text{ELF}}. (2)$$

Eq. 2 can be interpreted as the cascade of four bi-quad sections and an integrator as shown in Fig. 3. The block diagram in Fig. 3 can be conveniently implemented using the G_m-C circuit shown in Fig. 4.

B. Filter Tuning Block

The filter tuning blocks model the function of the OHCs in real cochlea. There are two sources of tuning control loop in the real cochlea: the instantaneous control from the nonlinear

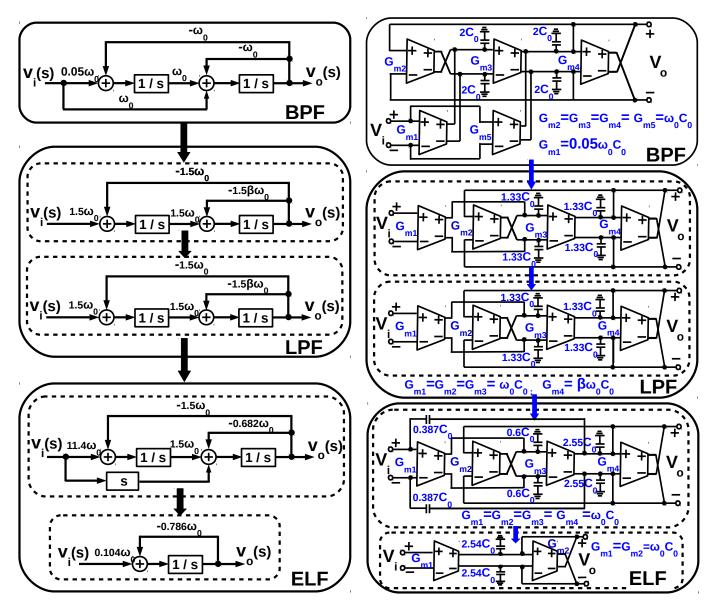


Fig. 3. Block diagram of the cochlea filter transfer function.

Fig. 4. $G_m - C$ circuit implementation of the cochlea filter transfer function.

behavior of OHC and the slow feedback control from the brain, which are respectively modeled as the local control and remote control signals in Fig. 1. In the local control loop of the cochlea, the OHC performs both receptor and effector functions; it transduces the BM-induced ciliary deflection into receptor potentials which then generates nonlinear mechanical feedback via somatic shape changing [14]. As observed in the physiological measurement, the OHC nonlinear motility contributes to approximately 40 dB peak gain variation [9]. In the remote control loop, the brain sends top-down feedback control signals to mediate the OHC behavior via efferent nerve fibers which is the underlying principle of humans periphery-level selective attention capabilities [15], [16].

To model the tuning control of the real cochlea, we design a filter tuning block that consists of an amplitude detector (AD), a multiplexer, and a tuning decision (TD) unit. As shown in Fig. 5, the AD which emulates the receptor function of the OHC extracts the signal amplitude from the output of the BPF

stage, and the TD which emulates the effector function of the OHC generates the corresponding β value for the LPF stages according to the output of the multiplexer which selects between the AD output (local control) and the remote control signal.

In this design, the signal amplitude is estimated according to the BPF output. There are two reasons for this setting: firstly, the auditory masking phenomenon observed in psychoacoustical experiments suggest the sensitivity adaptation of the biological cochlea is performed according to the intensity of the relevant in-band signals [17]; secondly, this setting is based on the feed-forward structure which avoids the potential latency problems due to the delay of multiple filter stages.

The AD consists of a squaring circuit and a DC extractor which respectively can be implemented using a CMOS multiplier and the same circuit for the ELF (3^{rd} -order elliptic filter). The remote control signal consists of the select part and the amplitude part. The select signal controls the multiplexer:

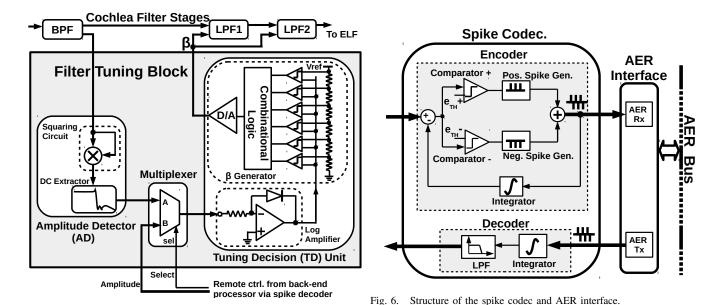


Fig. 5. Structure of the filter tuning block.

when it is off, the AD output is fed to the tuning decision unit to perform the local control; when it is on, the amplitude part of the remote control signal is selected instead so that the filter channel can remotely controlled by a back-end processor. The TD unit consists of a logarithmic amplifier and a β generator. The logarithmic amplifier compresses the detected amplitude signal into log scale. The β generator incorporates a flash ADC architecture which compares the log-scale amplitude signal with 7 threshold voltage level, and the comparison output are converted into corresponding β values via a combinational logic unit and a D/A converter. We assume the sound input dynamic range is over 100 dB which is in accordance with the physiological experiment data in [9], and the 7 β values are generated corresponding to the input intensity (detected amplitude) as shown in Table. I. The simulated peak gain listed in the table suggests that the proposed filter tuning block is able to compress the signal by maximally over 40 dB which matches the physiological data.

TABLE I. MAPPING BETWEEN INPUT INTENSITY AND β VALUES

Input Intensity (dBFS)	β	Simulated Peak Gain (dB)
≥ 0	1	1.86
-20 ∼ 0	0.61	8.04
-40 ∼ −20	0.39	14.9
-60 ∼ −40	0.24	22.3
-80 ∼ −60	0.15	30.2
-100 ∼ −80	0.092	38.3
<-100	0.057	46.5

C. Spike Codec

In the proposed silicon cochlea system, the spike codec consists of a spike encoder which models the neural receptor function of the inner hair cells (IHC) and a spike decoder which models the efferent innervation on the OHCs. In engineering point of view, the spike codec functions as the interface between the analog signal processing domain (cochlea filterbank and filter tuning blocks) and the back-end processor. Unlike the the A/D and D/A converters in conventional electronic systems, the spike codec converts signals between the

continuous-time continuous-amplitude (CTCA) analog form and the continuous-time discrete-amplitude (CTDA) spike events. Apart from the bio-mimicking purpose, the use of spike codec instead of A/D and D/A converters is also on account of at least two aspects of advantages in engineering perspective. Firstly, its asynchronous event-driven nature means the sampling rate is associated with the input signal activity which results in lower power consumption and better resource utilization compared with conventional synchronous sampling systems [18]. Secondly, the spike codec interface the analog signals with the CTDA-domain processors (such as the spiking neural networks [19], [20]) which prospectively combines the advantage of aliasing avoidance in continuous-time systems and the advantage of insensitivity to component tolerances, mismatches, and noise in discrete-amplitude systems [21].

The spike codec we employed is based on the asynchronous delta modulation principle and its block diagram is illustrated within Fig. 6. In the encoder part, the input signal is tracked by the reconstructed feedback signal which is generated by integrating the biphasic spike outputs. The tracking error is compared with the positive and negative error thresholds by a pair of comparators. A positive spike is generated if the error exceeds the positive threshold and a negative spike is generated if the error exceeds the negative threshold. The positive and negative spikes respectively result in an incremental and decremental tracking step at the feedback integrator output, and the tracking with input signal is maintained if the tracking step matches the error threshold of the comparators. In the decoder part, the biphasic spikes are integrated in the same way as that in the decoder, and their high-frequency harmonics are further removed by a low-pass filter to reconstruct the original analog signal. More details about the structure and the circuit implementation of the spike codec can be found in [18].

This spike coding scheme is selected to construct the silicon cochlea due to its two advantages. First, as demonstrated in the VLSI hardware measurement results [18], a speech signal that is converted into spikes by the spike encoder can

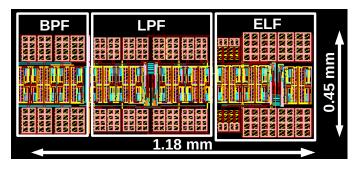


Fig. 7. Layout of a cochlea filter channel implement using SMIC $0.18 \mu m$ CMOS process.

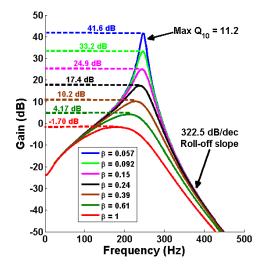


Fig. 8. Simulated frequency response of the cochlea filter channel.

be fully recovered by the spike decoder, which means the coding scheme well reserves the signal information. Besides, the measured results in [18] also proved that this coding scheme has intrinsic weighted sum computation capabilities, which means it brings an extremely convenient solution for the cochlea channel frequency weightings and the commonly used weighted summation operation in the back-end artificial neural networks.

IV. SIMULATION RESULTS

A. Circuit Simulation of a Cochlea Filter Channel

We have implemented the Gm-C cochlea filter channel shown in Fig. 4 in VLSI. Fig. 7 shows the layout of the channel. The circuit is biased with appropriate current so that the expected ω_0 is approximately 175 Hz. The total power consumption for the channel is as low as 4.45 μ W. We simulated the circuit in Cadence Spectre using the SMIC 0.18 μm CMOS process parameters. Fig. 8 shows the simulated frequency response of the filter channel with various β values. The peak gains are respectively $3\sim 5$ dB lower than the expecting values listed in Table. I due to circuit non-idealities, but the variation range is still over 40 dB. The maximum Q_{10} factor reaches over 10 by virtue of the modified filter architecture while the roll-off slope steepness maintains over 300 dB/dec.

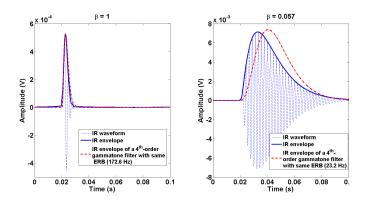


Fig. 9. Simulated impulse response (IR) of the cochlea filter channel under extreme conditions

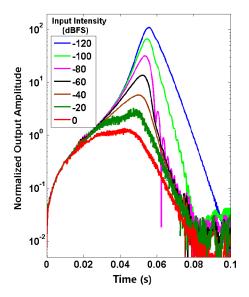
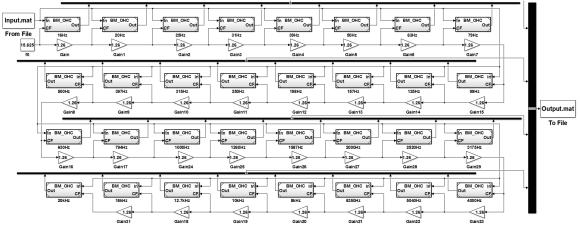


Fig. 10. Normalized filter output envelopes in response to chirp signals with various intensities.

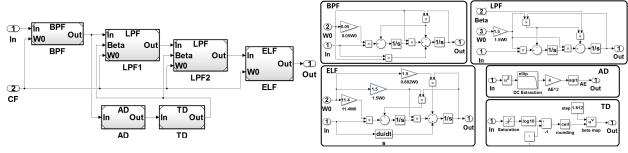
The circuit impulse responses (IR) under maximum and minimum β conditions are shown in Fig. 9. The IR envelopes are compared with those of a 4^{th} order gammatone filter under the same equivalent-rectangular-bandwith (ERB) conditions. The ERB values of the cochlea filter channel are calculated according to the frequency response shown in Fig. 8. The gammatone filter has been popularly used to model the biological cochlea filter responses due to its close similarity with the IR of the cochlea [22]. Here Fig. 9 shows that the IR of our proposed filter generally matches that of the gammatone filter. Moreover, the proposed filter is superior to the gammatone filter in the way that it has a simple form of Laplace-domain pole-zero description and is easy to implement using analog VLSI circuit as introduced in Section III-A.

B. Adaptive filter tuning with local control

The model of the filter tuning block is simulated together with the filter channel in MATLAB Simulink. The local control loop in the multiplexer of filter tuning block is selected so that the β value is adaptively adjusted by the TD unit according to the in-band signal intensity. The passive center frequency is set at 1000 Hz and a chirp signal that sweeps from 1 to



(a) Top-level architecture of the 32-channel model



(b) Details of a cochlea channel

Fig. 11. MATALB Simulink model of a 32-channel silicon cochlea (SC).

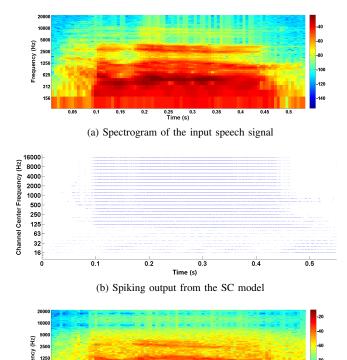
3000 Hz with various intensities from -120 to 0 dBFs are applied at the input. The signal envelopes are extracted from the filter output and are respectively divided by the input and converted to dB scale for normalization. The results are shown in Fig. IV-A which proves that the gain and selectivity of the filter adaptively varies according to the input intensity and the adaptation of the filter contributes to approximately 40 dB gain compression as expected. Notably, the roll-off slopes of under -100 dB and -120 dB conditions are more gentle than the others because of the slower settling of the center frequency signal.

C. Spiking output of the silicon cochlea system in response to audio signals

We build in MATLAB Simulink software a silicon cochlea (SC) model that consists of 32 channels with center frequencies distributed from 15.6 Hz to 20 kHz in the one-third octave scale. A male speech segment is applied at the input, and the spike outputs from the silicon cochlea model is shown in Fig. 12b where both the positive and negative spikes are represented as dots for ease of illustration. The spectrogram of the reconstructed speech signal from the spike outputs using the spike decoder is shown in Fig. 12c, which generally replicates that of the original input shown in Fig. 12a.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a silicon cochlea system that behaves faithfully as the real cochlea. The system consists of multiple cochlea channels and each channel has three basic building blocks, the cochlea filter, the filter tuning block and



(c) Spectrogram of the signal recovered from the SC output spikes

Fig. 12. Response of the SC model to a male speech saying 'Hello'.

the spike codec. Both the cochlea filter and the spike codec have been proven in analog VLSI chip results [11], [18]. We are now implementing the circuits for the filter tuning block and will then integrate the building blocks on the same chip. Eventually we will integrate the silicon cochlea chip with spiking neural networks to perform more specific audio signal processing tasks.

REFERENCES

- [1] R. F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 7, pp. 1119-1134, Jul. 1988.
- [2] S. Mandal, S. M. Zhak and R. Sarpeshkar, "A bio-inspired active radio-frequency silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 44, no. 6, pp. 1814-1828, Jun. 2009.
- [3] S.-C. Liu, A. van Schaik, B. A. Minch and T. Delbruck, "Asynchronous binaural spatial audition sensor with 2 × 64 × 4 channel output," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 4, pp. 453-464, Aug. 2014.
- [4] T. J. Hamilton, C. Jin, A. van Schaik and J. Tapson, "An active 2-D silicon cochlea," *IEEE Trans. Biomed. Circuits Syst.*, vol. 2, no. 1, pp. 30-43, Mar. 2008.
- [5] B. Wen and K. Boahen, "A silicon cochlea with active coupling," *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 6, pp. 444-455, Dec. 2009.
- [6] H. Finger and S.-C. Liu. "Estimating the location of a sound source with a spike-timing localization algorithm", Proc. IEEE Int. Symp. on Circuits and Systems, 2011, pp. 2461-2464.
- [7] C.-H. Li, T. Delbruck and S.-C. Liu, "Real-time speaker identification using the AEREAR2-event-based silicon cochlea", *Proc. IEEE Int. Symp.* on Circuits and Systems, 2012, pp. 1159-1162.
- [8] T. Harczos, A. Chilian and P. Husar. "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: the SAM coding strategy," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 4, pp. 414-425, Aug. 2013
- [9] 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-nervefiber responses in chinchilla," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, pp. 11744-11750, Oct. 2000.
- [10] A. G. Katsiamis, E. M. Drakakis and R. F. Lyon, "A biomimetic, 4.5 μ W, 120+ dB, log-domain cochlea channel with AGC," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 1006-1022, Mar. 2009.
- [11] S. Wang, T. J. Koickal, A. Hamilton R. Cheung and L. S. Smith, "A bio-realistic analog CMOS cochlea filter with high tunability and ultra-steep roll-off," *IEEE Trans. Biomed. Circuits Syst.*, 2014. doi: 10.1109/TBCAS.2014.2328321.
- [12] S. Wang, T. J. Koickal, A. Hamilton, E. Mastropaolo, R. Cheung and L. Smith, "A floating active inductor based CMOS cochlea filter with high tunability and sharp cut-off," *Proc. IEEE Symp. on Circuits and Systems*, 2013, pp. 193-196.
- [13] A. B. Williams and F. J. Taylor, Electronic filter design handbook, New York: McGraw-Hill, 1995.
- [14] P. Dallos, "The active cochlea" J. Neurosci., no. 12, pp. 4575-4585, Dec. 1992.
- [15] W. B. Warr and J. J. Guinan, "Efferent innervation of the organ of corti: two separate systems," *Brain. Res.*, vol. 173, pp. 152-155, 1979.
- [16] M.-H. Girad, L. Collect, P. Bouchet, and J. Pernier, "Auditory selective attention in the human cochlea," *Brain. Res.*, vol. 633, pp. 353-356, 1994.
- [17] H. Fletcher, "Auditory patterns" Rev. Mod. Phys., vol. 12, pp. 47-65, 1940.
- [18] L. Gouveia, and T. J. Koickal, and A. Hamilton, "An asynchronous spike-event coding scheme for programmable analog array," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 5, pp. 791-799, 2011.
- [19] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri, "Neuromorphic electronic circuits for building autonomous cognitive systems," *P. IEEE*, vol. 102, no. 9, pp. 1367-1388, 2014.
- [20] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: a mixed-analog-digital multichip system for

- large-scale neural simulations," *P. IEEE*, vol. 102, no. 5, pp. 699-716, 2014.
- [21] Y. W. Li, K. L. Shepard, and Y. P. Tsividis, "Continuous-time digital signal processors," *Proc. IEEE Int. Symp. Asynchronous Circuits Syst.*, pp. 138-143, 2005.
- [22] R. D. Patterson, "The sound of a sinusoid: spectral models," J. Acoust. Soc. Amer., vol. 96, pp. 1409-1418, Sept. 1994.