Design, Fabrication & Evaluation of a Biomimetic Filter-bank Architecture
For
Low-power Noise-robust Cochlear Implant Processors

By
Guang Yang

A thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy and the Diploma of Imperial College London

Department of Bioengineering,
Imperial College London

London, September 2012
To my parents Yang JunMing and Wang XianJu
Declaration of Originality

I hereby declare that this thesis and the work presented herein were originated from me. All information derived from the work of others has been acknowledged in the text and referenced appropriately.
Abstract

This thesis presents a new bio-inspired filterbank-architecture termed “OZGF-with-LI” for the signal processing in cochlear implants (CIs), and its ultra-low-power analog very-large-scale-integration (VLSI) implementation. The OZGF-with-LI system provides a potential solution to the noise susceptibility of current CI users by simulating the lateral inhibition (LI)—a biological spectral-enhancement mechanism that may partly account for the high noise robustness of the human auditory system but is missing in current CIs—as a soft local “winner-take-all” link between different frequency regions in the input spectrum. It performs multi-channel syllabic compression via automatic gain control (AGC) while preserving well the spectral contrast and hence the original spectral features of the system input—an advantage over the compression scheme used in current CIs, which degrades spectral contrast and thus impairs the ability of CI users to pick out the spectral peaks (features) as the identity of various speech sounds embedded in noise. Two perceptual tests via acoustic simulation of CIs verified the benefits of the proposed system: the simulated LI provided a substantial benefit for listening to speech presented in noise. The proposed architecture was explicitly designed to be amenable to low-power analog VLSI implementations. This thesis reports such a silicon integrated-circuit (IC) prototype of the OZGF-with-LI system fabricated in the commercially available 0.35μm AMS CMOS process, with a power consumption of 28μW and an input dynamic range of 92dB. This system has potential for use in fully implantable CIs of the future, which have very stringent requirements on the power consumption for signal processing.
Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Emmanuel (Manos) Drakakis. It is he who taught me how to think creatively and constructively, who motivated me to keep an eye on the bigger picture while guiding me to remain focused, and who always shared his wisdom and knowledge with me. This thesis, as the fruit of my PhD research, would have not been possible without his guidance, support and patience during the last four years.

– Manos, you’ve earned a friend for life – “Mr. Cochlear Implant” (you gave this cool nickname to me, and I love it!).

I would like to thank all my colleagues in the Bioinspired VLSI CAS Group, a big academic family with an intellectually-rich and friendly atmosphere. I am indebted to Henry Man Den Ip, Konstantinos Glaros, Evdokia Kardoulaki, Alex Yue, Metha Kongpoon, Itir Koymen, Konstantinos Papadimitriou, Panayiotis Georgiou and Andreas Procopiou for making my time here one of the happiest experiences in my life. I am also grateful to our past members—Andreas G. Katsiamis, Yan Huang, Aggelos Lazaridis, Thanut Tosanguan, Amir Eftakhar and Anoop Walia—for introducing and welcoming me to this family. I am proud to have been part of this very warm family.

– Guys, the bits and pieces of my life here alongside you have been deeply rooted in my heart and memory. Let’s keep in touch in the future.

Last but not least, I own my deepest gratitude to Mr. Yang JunMing and Mrs. Wang XianJu – my dear parents. Since I began my study abroad journey six years ago, their love, their support and encouragement—from thousands of miles away—have always been constant sources of my confidence and power to get over any difficulty during this new life experience.

– Mom and Dad, I am very proud to be your son, and hope my six-year study journey here does not disappoint you. I always miss you both!
Table of Contents

Abstract ......................................................................................................................... - 4 -
Acknowledgement ......................................................................................................... - 5 -
Table of Contents ......................................................................................................... - 6 -
List of Figures ................................................................................................................ - 9 -
List of Tables ................................................................................................................ - 16 -
List of Abbreviations .................................................................................................... - 17 -

Chapter 1
Introduction .................................................................................................................. - 19 -
1.1 Motivation ............................................................................................................. - 19 -
1.2 Aspects of Normal Hearing .................................................................................... - 20 -
1.3 Deafness and Cochlear Implant (CI) ................................................................. - 23 -
1.4 Present CI Processing Strategies ....................................................................... - 28 -
1.5 Recent Advances and Future Possibilities ......................................................... - 34 -
1.6 Thesis Overview .................................................................................................... - 38 -

Chapter 2
OZGF-with-LI: A Biomimetic Filterbank for Noise-robust Speech Processing in Cochlear Implants ........................................................................................................ - 46 -
2.1 Introduction .......................................................................................................... - 46 -
2.1.1 Causes of the CI susceptibility to noise ............................................................ - 46 -
2.1.2 Improvement of the CI performance in noise ................................................... - 48 -
2.1.3 Efforts in this work ........................................................................................... - 50 -
2.2 Architecture Design ............................................................................................... - 51 -
2.2.1 One-Zero-Gammatone-Filter (OZGF) channels ............................................. - 52 -
2.2.2 Frequency-dependent $Q$-adaptive cross-coupled AGCs ............................... - 54 -
2.3 Simulation Experiments and Results ................................................................... - 58 -
2.3.1 Compression without Spectral Enhancement ................................................... - 62 -
2.3.2 Compression combined with Spectral Enhancement ....................................... - 65 -
2.3.3 Two-tone suppression (2TS) ............................................................................ - 72 -
2.3.4 Benefits for CI Processing ............................................................................... - 77 -
2.4 Potential Use in Cochlear Implants: A new Speech Processing Strategy of increased Bio-realism ........................................................................................................ - 81 -
2.5 Further Discussions and Conclusions ................................................................. - 84 -

Chapter 3
Speech Recognition Evaluation for the OZGF-with-LI System .......................... - 93 -

3.1 Introduction ........................................................................................................... - 93 -

3.2 General Methods ................................................................................................... - 94 -

3.2.1 Test Materials .................................................................................................. - 94 -

3.2.2 Cochlear Implant Simulator ............................................................................ - 94 -

3.2.3 Listeners .......................................................................................................... - 96 -

3.2.4 Procedure ....................................................................................................... - 97 -

3.3 Experiment I: OZGF vs. Cascaded Bandpass Biquads ..................................... - 98 -

3.3.1 Algorithm and Parameter Setting .................................................................... - 99 -

3.3.2 Results and Discussion .................................................................................. - 102 -

3.4 Experiment II: AGC Coupling ON vs. Coupling OFF ..................................... - 103 -

3.4.1 Algorithm and Parameter Setting .................................................................... - 103 -

3.4.2 Results ........................................................................................................... - 104 -

3.5 General Discussion and Summary ..................................................................... - 106 -

Chapter 4
Ultra-low-power Analog VLSI for the OZGF-with-LI System ............................. - 112 -

4.1 Introduction .......................................................................................................... - 112 -

4.2 System Overview ................................................................................................... - 113 -

4.3 Log-domain Class-AB OZGF Channels ............................................................. - 115 -

4.3.1 Log-domain Structure and Pseudo-differential Class-AB Design ............... - 116 -

4.3.2 Filter Synthesis in Log-domain State-space .................................................. - 119 -

4.3.3 OZGF Filterbank ........................................................................................... - 125 -

4.4 Coupled Channel AGCs ..................................................................................... - 128 -

4.4.1 Quasi- Full-wave Rectification using Geometric Mean Splitter (GMS) ...... - 131 -

4.4.2 The LPF Smoothing ....................................................................................... - 134 -

4.4.3 AGC-coupling network and circuits ............................................................... - 136 -

4.4.4 Integrated $I \rightarrow V$ and operational transconductance amplifier (OTA) ..... - 139 -

4.4.5 AGC Simulation and Measured Results ....................................................... - 141 -

4.5 Chip Measurements ............................................................................................ - 147 -

4.5.1 Optimized Device Sizing and Layout ............................................................. - 148 -

4.5.2 Measurement Setup ....................................................................................... - 152 -

4.5.3 Frequency Response ...................................................................................... - 153 -

4.5.4 Linearity Performance ................................................................................... - 161 -

4.5.5 Input Dynamic Range (DR) ........................................................................... - 163 -

4.5.6 Single-tone Test ............................................................................................. - 165 -

4.5.7 Complex-tone Test ......................................................................................... - 167 -

4.6 Summary and Conclusion ................................................................................... - 172 -
Chapter 5

OZGF-with-LI: Some Likely Next Steps .......................................................... 177

5.1 Introduction ................................................................................................................. 177
5.2 Programmable OZGF-with-LI based Processor ................................................... 177
5.3 Reconfigurable and Hybrid Channel Architecture ................................................. 180
  5.3.1 Programmable filter order and biomimetic phase response ......................... 180
  5.3.2 Hybrid OZGF Channel ...................................................................................... 184
5.4 Enhanced Lateral Inhibition ...................................................................................... 188
5.5 Conclusion ................................................................................................................... 192

Chapter 6

Conclusion ......................................................................................................... 195

Appendix A:

Acoustic Experiment – Summary/Protocol ...................................................... 202
List of Figures

Figure 1.1: A simplified diagram (not to scale) of the human ear; from Loizou [1].................. - 20 -
Figure 1.2: Frequency-place map (in Hz) on the BM (snail-shaped); from Loizou [1]. High frequencies are mapped to the places near the base (since they produce maxima there), whereas the low frequencies are mapped to the places near the apex. .............................................................. - 22 -
Figure 1.3: BM responses versus location (in mm) for different stimulus (pure tones of 300, 200, 100, and 50Hz); the stapes is one of the three ossicles in the middle ear and attached to the base of the cochlea; the BM is about 35mm long; from Békésy [7]............................................................. - 22 -
Figure 1.4: BM responses versus frequency of the stimuli tone for different locations (in mm) (the stimuli amplitude is fixed); from Békésy [7].......................................................................................... - 23 -
Figure 1.5: Block diagram of the normal auditory pathway (left panel; adapted from [9, 10]) and that rebuilt with a cochlear implant (CI) (right panel). The primary functions are indicated for each block in the normal auditory pathway. ............................................................................. - 24 -
Figure 1.6: A typical cochlear implant (CI) using the continuous interleaved sampling (CIS) processing strategy, whose details are given in Section 1.4; from Loizou [1]. The top panel shows the major components of the CI while the bottom panel describe how the microphone input (the syllable “sa”) is transformed into stimuli via the CIS processor. .............................................................................. - 25 -
Figure 1.7: Block diagram of continuous interleaved sampling (CIS) processing strategy (LPF: low-pass filter; EL-\(i\) the \(i^{th}\) electrode). The upper panel describes the architecture implementing CIS, while the lower panel illustrates the interleaved form of stimulus pulses sent for the modulation with the signal envelopes................................................................. - 29 -

Figure 2.1: Block diagram of the proposed OZGF filterbank with the cross-coupled AGC scheme modelling the LI. Weighting factors W are used to adjust the extent of the AGC coupling. ................................................................................................................................. - 52 -
Figure 2.2: The OZGF frequency response of order \(N = 4\) and with \(Q\) ranging from 0.75 to 10. The frequency axis is normalized to the natural frequency.................................................. - 53 -
Figure 2.3: Block diagram of the 4th-order OZGF channel \(i\) with cross-coupled AGC (BP: bandpass; LP: lowpass)........................................................................................................................... - 58 -
Figure 2.4: A computational model of the \(Q\) control law in AGC with input to every stage transfer characteristics .................................................................................................................. - 60 -
Figure 2.5: The plots of the parametric AGC transfer characteristic with varying the parameters in (3). The bold curves correspond to the initial setting for the simulation \((I_{\text{tail}} = 20000, I_0 = 1/0.9, I_{\text{control}} = 0.9)\).................................................................................................................. - 61 -
Figure 2.6: The parametric effects of multi-channel compression.................................................................. - 64 -
Figure 2.7: Input (dotted lines) and output spectra (continuous lines) of $u$. The AGC coupling ON case corresponds to $W_i = W_{il} = 0.4$, $W_\phi = 0.1$, $K_c = 1.12$ (1dB). The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 0.5$, $K_c = 1$ (0dB). .................................................. - 66 -

Figure 2.8: Output spectra of $u$ with varying the ratio of $W_{il}/W_\phi$. The AGC coupling ON case corresponds to $W_i = W_{il} = 0.4$, $W_\phi = 0.1$, $K_c = 1.12$ (1dB) and $W_i = W_{il} = 0.14$, $W_\phi = 0.28$, $K_c = 1$ (0dB) respectively. The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 0.5$, $K_c = 1$ (0dB). .................................................. - 66 -

Figure 2.9: The parametric effects of spectral enhancement.................................................. - 67 -

Figure 2.10: Input spectra (dotted lines) and output spectra (continuous lines) of $u$ when $w_i = 0.1\omega_0$ (shorter time constant). The AGC coupling ON case corresponds to $W_i = W_{il} = 0.4$, $W_\phi = 0.1$, $K_c = 1.12$ (1dB). The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 0.5$, $K_c = 1$ (0dB). .................................................. - 69 -

Figure 2.11: Input spectra (dotted lines) and output spectra (continuous lines) of $u$ when $w_{i,z} = 0$, $w_{i+1} = W_{il}$, $w_{i+2} = 0$, $w_{i+3} = W_\phi$. The AGC coupling ON case corresponds to $W_i = W_{il} = 1.4$, $W_\phi = 0.35$, $K_c = 1$ (0dB). The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 1.1$, $K_c = 1$ (0dB). .................................................. - 70 -

Figure 2.12: Input spectra (dotted lines) and output spectra (continuous lines) of $u$ with 32 OZGF channels ($w_{i,z} = w_{i+2} \neq 0$). The AGC coupling ON case corresponds to $W_i = W_{il} = 0.8$, $W_\phi = 0.2$, $K_c = 2$ (6dB). The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 0.5$, $K_c = 1$ (0dB). .................................................. - 70 -

Figure 2.13: Input spectra (dotted lines) and output spectra (continuous lines) of $u$ corresponding to two different coupled AGC schemes. The summation-coupling case corresponds to $W_i = W_{il} = 0.4$, $W_\phi = 0.1$, $K_c = 1$ (0dB). The RMS-coupling case corresponds to $W_i = W_{il} = 1.25$, $W_\phi = 0.35$, $K_c = 2.24$ (7dB). .................................................. - 71 -

Figure 2.14: Input spectra (dotted lines) and output spectra (continuous lines) of $u$ when a cascade of four bandpass biquad filters ($N=4$, $\omega_z = 0.1\omega_0$) is used instead of OZGF. The AGC coupling ON case corresponds to $W_i = W_{il} = 0.8$, $W_\phi = 0.5$, $K_c = 1.6$ (4dB). The AGC coupling OFF case corresponds to $W_i = W_{il} = 0$, $W_\phi = 0.5$, $K_c = 1$ (0dB). .................................................. - 72 -

Figure 2.15: Observation of Two-tone Suppression in the spectrum of $i$................................. - 73 -

Figure 2.16: The tone-to-tone suppression with the filterbank of OZGF and BP-cascade ($N=4$). In both cases, $W_i = W_{il} = 0.4$, $W_\phi = 0.1$. The suppressor is inputted with a fixed amplitude of 0dB and a frequency of 1097Hz (corresponding to the peak location) while the input amplitude of the suppressed tone is fixed to be 0 dB. The two-tone FFT of the architecture output is plotted as the suppressed-tone frequency varies................................. - 74 -

Figure 2.17: The tone-to-tone suppression (in the OZGF-filterbank) as the probe-tone frequency is varied for different values of the weighting factors. In the $W_i = W_{il}$ case, $W_{il} = W_i = 0.4$ and $W_\phi = 0.1$; In the $W_i > W_{il}$ case, $W_{il} = 2.4$, $W_i = 0.4$ and $W_\phi = 0.1$................................. - 75 -
Figure 2.18: The tone-to-tone suppression (in the BP-cascade filterbank) as the probe-tone frequency is varied for different values of the weighting factors. In the $W_{ii} = W_{ii}$ case, $W_{ii} = W_{ii} = 0.4$ and $W_{i0} = 0.1$; in the $W_{ii} > W_{ii}$ case, $W_{ii} = 2.4$, $W_{iL} = 0.4$ and $W_{i0} = 0.1$; in the $W_{ii} < W_{ii}$ case, $W_{ii} = 0.4$, $W_{iL} = 2.4$ and $W_{i0} = 0.1$.

Figure 2.19: The tone-to-tone suppression (in the BP-cascade filterbank) as the probe-tone frequency is varied for different values of the weighting factors. The AGC coupling-on case corresponds to $W_{iL} = W_{ii} = 0.4$, $W_{i0} = 0.1$. The AGC coupling-off cases corresponds to $W_{iL} = W_{ii} = 0$.

Figure 2.20: The tone-to-tone suppression (in the BP-cascade filterbank) as the suppressor-tone amplitude is varied for different suppressor frequency $W_{0}$. The suppressed-tone frequency $f_{s}$ is fixed at 2300Hz. All the cases have $W_{iL} = W_{ii} = 0.4$, $W_{i0} = 0.1$.

Figure 2.21: Bar-charts of maximum output of each channel versus channel number for the vowel input /u/: (a) AGC is disabled by setting the same $Q$ ($Q = 1.6$) for every channel with $I_{ad} = \omega_{0}/1.6$ and $I_{a} = 1000$, no compression and spectral enhancement being performed; (b) The AGC coupling is disabled ($W_{ii} = W_{iL} = 0$, $W_{i0} = 0.5$, $K_{c} = 1$), the $Q$ range for every channel being 0.8 to 15.6; (c) The AGC coupling is active ($W_{ii} = W_{iL} = 0.4$, $W_{i0} = 0.1$, $K_{c} = 1.12$), the $Q$ range for every channel being 0.8 to 15.6.

Figure 2.22: Spectrogram-like plots for the word “blue” illustrating the clarifying effect of the AGC coupling strategy. In the top figure, AGC is disabled by setting the same $Q$ ($Q = 1.6$) for every channel with $I_{ad} = \omega_{0}/1.6$ and $I_{a} = 1000$. In the middle figure, the AGC cross-coupling is disabled ($W_{ii} = 0$, $W_{iL} = 0$, $W_{i0} = 0.3$, $K_{c} = 1$), the $Q$ range for every channel being 0.8 to 15.6. In the lower figure, the AGC cross-coupling is active ($W_{ii} = 0.25$, $W_{iL} = 0.25$, $W_{i0} = 0.1$, $K_{c} = 1$), the $Q$ range for every channel being 0.8 to 15.6.

Figure 2.23: Maximum outputs of each channel versus channel number for the vowel input /u/ in Gaussian white noise of different SNRs (in 50ms, during which the synthetic vowel was repeated 5 times). The un-processing case corresponds to $Q = 1.6$ for each channel. The AGC coupling-on case corresponds to $W_{iL} = W_{ii} = 0.3$, $W_{i0} = 0.2$ and $Q = 0.8$ to 15.6 for each channel. The parameters were chosen to have $F_{1}$ with the same amplitude in both cases. The first six plots show the progressive loss of $F_{1}$ and the following six plots show the progressive loss of $F_{2}$. The bottom plot illustrates that the recognition of $F_{2}$ (local contrast) can be further improved with a larger ratio of $W_{ii}(W_{iL})$ to $W_{i0}$.

Figure 2.24: Input spectra (dotted lines) and output spectra (continuous lines) of the vowel /u/. The AGC coupling ON case corresponds to $W_{iL} = W_{ii} = 0.4$, $W_{i0} = 0.1$, $K_{c} = 1.12$. The AGC coupling OFF case corresponds to $W_{iL} = W_{ii} = 0$, $W_{i0} = 0.5$, $K_{c} = 1$. In both cases, $I_{ad} = \omega_{0}/1.5$ and $I_{a} = 1.2$ so that the $Q$-range of every channel is 1.5 to 8.3. The resulting DR of $Q$ (i.e., 15dB) is much smaller than the one initially set (i.e., 26dB) while the AGC coupling can still effectively improves the recognition of all the formants.
Figure 3.1: Long-term average spectrum (LTAS) of CRM sentences (top panel) and the corresponding speech shaped noise spectrum (bottom panel) with SNR = 12dB for clarity. The LTAS was derived by averaging the superimposed spectra across CRM sentences. 

Figure 3.2: An overview of the noise-excited envelope vocoder used to simulate the CI processing.

Figure 3.3: The graphical user interface (Matlab-based) for CRM tests.

Figure 3.4: A piece-wise approximation of the BM frequency response defined by Rhode [10] with the parameters reported in various studies [9-13]; adapted from Katsiamis et al. [14]

Figure 3.5: A comparison of the OZGF and the cascaded bandpass biquads in frequency response (adapted from [3, 5]).

Figure 3.6: Speech intelligibility results for the three SNRs with the bandpass cascade vs. the ones with the OZGF, scored in percent correct. Bars represent the mean and ±1 standard error across the eight listeners.

Figure 3.7: The spectrograms for the CRM sentence “Ready Arrow, go to blue eight now.” with the AGC coupling OFF and ON.

Figure 3.9: Speech intelligibility results for the three SNRs with AGC Coupling OFF vs. ON, scored in percent correct. Bars represent the mean and ±1 standard error across the fourteen listeners.

Figure 3.8: OZGF $S3$ slope versus stage $Q$ for different $N$. From $Q = 2.6$ to 10, the variation in $S3$ does not exceed 2% for each $N$.

Figure 3.10: $N$-channel pre-processing followed by an $M$-channel noise excited envelope vocoder for the CI simulation. $M$ is matched with one of the channel numbers used in current cochlear implants, e.g., 8 or 16, and $N > M$.

Figure 4.1: Overall architecture of the analog OZGF-with-LI filterbank. For details on each 4th order OZGF channel together with its coupled AGC, see Figure 4.2.

Figure 4.2: Block diagram of a single 4th order OZGF channel with its coupled AGC block (BP – bandpass; LP – lowpass; FWR – full wave rectifier; LPF – low pass filter; the index of coupled channels ranges from i-2 to i+2).

Figure 4.3: Block diagram of a 4th order pseudo-differential Class-AB OZGF channel, driven by a geometric mean splitter (GMS). The required bias signals are denoted by $I_0$, $I_Z$ and $I_Q$ for the OZGF as well as $I_{BLANC_GMS}$ for the GMS. The vertical arrows between the two Class-A filter branches (each branch contains one BP- and three LP- biquads) indicate some form of coupling for ensuring Class-AB operation.

Figure 4.4: Log-domain filter’s companding-structure.

Figure 4.5: The Bernoulli Cell (CMOS version).

Figure 4.6: An overview of log-domain state-space approach for log-domain filtering.
Figure 4.7: Log-domain Biquad Synthesis in a pseudo-differential Class-AB topology based on the state-space equations in Table 4-I [5]. Note the downward/upward feeding of output currents ($I_{OUT,\text{u}}$ and $I_{OUT,\text{l}}$) between the ‘upper’ and the ‘lower’ branches in topology, which corresponds to the linear and nonlinear cross-coupling terms involved in the steady-state description of the system. The circuits not shown here for clarity are the PMOS and NMOS cascode current mirrors which serve all biasing current sources and implement the subtraction operation on the above output currents. All device sizes and capacitors in the topology are displayed in Table 4-II in Section 4.45.

Figure 4.8: Simulated frequency responses ($Q = 5$) of the 16 OZGF channels of which centre (peak) frequencies are equally spaced along a logarithmic axis ranging from 250Hz to 4000Hz.

Figure 4.9: Simulated frequency responses ($Q = 1$) of the 16 OZGF channels of which centre (peak) frequencies (CF) are equally spaced along a logarithmic axis ranging from 205Hz to 3280Hz.

Figure 4.10: Simulated CF distribution across the 16 channels corresponding to the peaks in Figure 4.8 and 4.9 with varying the bias current $I_0$ from 1.25nA to 20nA uniformly on the logarithmic x-axis.

Figure 4.11: Block diagram of the coupled AGC circuits within each channel (the subscripts $i$, $i+1$, $i+2$, $i-1$ and $i-2$ are channel indexes).

Figure 4.12: The GMS with balanced architecture used as the global input conditioner (formed by weakly inversed PMOS transistors).

Figure 4.13: The indicative GMS output waveforms given by (20) and (21).

Figure 4.14: The GMS output versus input given by (21) (i.e., the output summation). The dotted line indicates the limit as the DC biasing current $I_{GMS,\text{bias}}$ approaches zero.

Figure 4.15: The GMS used as a quasi-full-wave rectifier (formed by weakly inversed PMOS transistors) in the AGC. The single-ended (rather than balanced) architecture was used since it fits a differential OZGF channel output.

Figure 4.16: The 1st-order log-domain LPF for smoothing in the AGC [22].

Figure 4.17: AGC coupling network across channels.

Figure 4.18: The coupling circuit for four different sets of weighting within each channel-AGC.

Figure 4.19: The integrated $I_{\text{to-V}}$ followed by the OTA.

Figure 4.20: Simulated waveforms generated at different stages of the uncoupled AGC (via $S_0$) for an input signal of $m = 20$. $I_{CP}$ is not shown in the figure since it is equal to $I_{ED}$ in the uncoupled case.
Figure 4.21: Measured frequency responses of the LPF in the AGC with varying $I_{LPF}$. The corresponding corner frequencies (at -3dB) are 6Hz, 13Hz, 27Hz, 57.5Hz and 100Hz, which are linearly related to the used values of $I_{LPF}$. 

Figure 4.22: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with varying $I_{tail}$.

Figure 4.23: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with varying $I_{th}$.

Figure 4.24: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with varying $I_{0_{control}}$.

Figure 4.25: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with a fixed minimum $I_Q$ via varying $I_{0_{control}}$ and $I_{th}$ simultaneously. The combinations $(I_{0_{control}}, I_{th})$ giving $K = 0.4$ are indicated in the zoomed plot.

Figure 4.26: A die photo of the 5-channel OZGF-with-LI system chip.

Figure 4.27: The ten pairs of the matched transistors (‘upper’ and ‘lower’) which constitute the Class-AB BP-biquad, corresponding to the schematic shown in Figure 4.7 (the unnecessary $M_{11}$ is removed when implementing the lossy BP-biquad). The I/O terminals, the biasing current sources $I_o$ and $I_z$, and the AGC output $I_Q$ as well as the two capacitors (denoted by $C$), are respectively connected to the source or the drain terminal of some of these transistors.

Figure 4.28: The common-centroid layout of a PMOS transistor pair. Each transistor (A or B) is divided into eight segments (i.e., fingers). The shaded portions on the illustrative pattern indicate the shared source terminals.

Figure 4.29: Measurement setup for the OZGF-with-LI system with channel outputs multiplexed to a single $I$-to-$V$ converter (the multiplexer not shown). The two current buffers or external current sources are connected to the global GMS preceding the OZGF through off-chip routing. The biasing currents $I_{bias}$ of the current buffers should correspond to the peak amplitude of $I_{IN+}$ (or $I_{IN-}$) so as to ensure the Class-A operation of the current buffer.

Figure 4.30: The gain-tunability of the 4th-order OZGF channel ($i = 9$) with the maximum peak gain at around 50dB ($Q = 5$). Each trace corresponds to a different input (and hence output) strength.

Figure 4.31: Low-frequency-tail tunability of the OZGF channel (via varying $I_z$).

Figure 4.32: A comparison of open-loop and closed-loop ($S_0$) frequency responses.

Figure 4.33: A comparison of frequency responses with AGC coupling ON ($S_1$) and OFF ($S_0$).

Figure 4.34: Measured frequency responses of the fabricated and CF-calibrated five channels with $\lambda_1$ for $Q = 5$ (upper) and $Q = 1$ (lower).

Figure 4.35: Tunability and gain adaption of the response over the frequency range corresponding to that of the simulated sixteen channels ($i = 1$~$16$).

Figure 4.36: Across-chip offsets (15 chips) for $i = 9$ with the two extreme $Q$ values (max. and min.).
Figure 4.37: Across-chip offsets (15 chips) for the other four channels with the two extreme $Q$ values.

Figure 4.38: Plots of measured adaptive peak gain (at CF) and the corresponding THD vs. channel input for different channel indexes ($i = 1$ and $i = 9$).

Figure 4.39: Plots of measured adaptive peak gain (at CF) and the corresponding THD vs. channel input for the four different weighting schemes ($S_0$–$S_3$) used for the AGC coupling.

Figure 4.40: Single-tone output amplitudes vs. input ones of neighbouring channels ($i = 7, 8, 10, 11$) normalized with respect to Channel $i = 9$.

Figure 4.41: The complex-tone input and its FFT spectrum.

Figure 4.42: Formation of a complex tone (Tone A), represented by the continuous line, and its complementary counterpart with 180° phase difference (Tone B), represented by the dotted line, from the five sinusoids whose frequencies and amplitudes are given in Table 4-VI. The amplitudes of Tone A and Tone B have been normalized to be within -1~1.

Figure 4.43: Output RMS levels (window length = 500ms) vs. channel normalized with respect to the maximum.

Figure 4.44: Normalized output RMS levels vs. input for different channels.

Figure 4.45: Increase in the across-channel contrast ($S_1$–$S_3$ relative to $S_0$) averaged over the input RMS levels.

Figure 5.1: Programmable OZGF-with-LI based CI processor.

Figure 5.2: Basic current-mode DAC (5-bit) for programmable biasing.

Figure 5.3: OZGF slope $S_2$ versus stage $Q$ for different filter order $N$.

Figure 5.4: Improved OZGF cascade architecture with programmable order $N$.

Figure 5.5: (a) Distributed and hybrid computation [2]; (b) Improved OZGF with-LI filterbank using hybrid channel architecture.

Figure 5.6: A 2-bit A/D/A: (a) the three threshold currents ($I_{T1}$, $I_{T2}$ and $I_{T3}$) and their resulting four attractor states ($I_{L1}$, $I_{L2}$, $I_{L3}$ and $I_{L4}$); (b) a possible circuit implementation.

Figure 5.7: Improved implementation of weighting factors in (23) for saving silicon area.

Figure 5.8: TL-based implementation of the RMS coupling (a CMOS version of the counterpart in [4, 5]. The double arrow indicates that the sign of $I_{ED,Neigh}$ in (30) and (31) can be either positive or negative.

Figure 5.9: The root-sum-of-powers circuit implementing (34) (a CMOS version of the counterpart in [5, 6]).
List of Tables

Table 1-I: An overview of processing strategies (stimulation form & advantage) in current use for CIs and the manufacturers that adopt them respectively [11] ................................................................. - 31 -
Table 2-I: Variations in $I_{\text{act}}$ and $I_{\text{p}}$ as shown in Figure 2.9 and the corresponding $Q$-ranges ($DR = Q_{\text{max}}/Q_{\text{min}}$) ............................................................................................................................................... - 67 -
Table 3-I: Transfer functions used in Experiment I – OZGF vs. Cascaded Bandpass Biquads ....- 101 -
Table 4-I: Log-domain state-space-based synthesis of pseudo-differential Class-AB biquads [5, 11] .............................................................................................................................................................. - 122 -
Table 4-II: Device sizing and parameter setting ................................................................. - 144 -
Table 4- III: Measured performance of the OZGF-with-LI system ........................................ - 148 -
Table 4- IV: Measured noise floor for the fabricated five channels........................................ - 164 -
Table 4-V: Measured noise floor for the four weighting schemes and the open-loop case.. - 164 -
Table 4-VI: The five pure tones for synthesis of a complex tone.............................................. - 169 -
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>ACE</td>
<td>Advanced Combination Encoder</td>
</tr>
<tr>
<td>A/D</td>
<td>Analog-to-Digital</td>
</tr>
<tr>
<td>AFE</td>
<td>Analog Front End</td>
</tr>
<tr>
<td>AGC</td>
<td>Automatic Gain Control</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ASIC</td>
<td>Application-Specific Integrated Circuit</td>
</tr>
<tr>
<td>BC</td>
<td>Bernoulli Cell</td>
</tr>
<tr>
<td>BJT</td>
<td>Bipolar Junction Transistor</td>
</tr>
<tr>
<td>BM</td>
<td>Basilar Membrane</td>
</tr>
<tr>
<td>BP</td>
<td>Band Pass</td>
</tr>
<tr>
<td>BPF</td>
<td>Band-Pass Filter</td>
</tr>
<tr>
<td>CF</td>
<td>Centre Frequency</td>
</tr>
<tr>
<td>CI</td>
<td>Cochlear Implant</td>
</tr>
<tr>
<td>CIS</td>
<td>Continuous Interleaved Sampling</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>CRM</td>
<td>Coordinate Response Measure</td>
</tr>
<tr>
<td>D/A</td>
<td>Digital-to-Analog</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital-to-Analog Converter</td>
</tr>
<tr>
<td>DAPGF</td>
<td>Differentiated All-Pole Gammatone Filter</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DR</td>
<td>Dynamic Range</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processing/Processor</td>
</tr>
<tr>
<td>EAS</td>
<td>Electrical and Acoustical (combined) stimulation</td>
</tr>
<tr>
<td>ED</td>
<td>Envelope Detector/Detection</td>
</tr>
<tr>
<td>ELIN</td>
<td>Externally-Linear-Internally-Nonlinear</td>
</tr>
<tr>
<td>ESS</td>
<td>Exponential State Space</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>FS</td>
<td>Fine Structure</td>
</tr>
<tr>
<td>FSP</td>
<td>Fine Structure Processing</td>
</tr>
<tr>
<td>FWR</td>
<td>Full-Wave Rectifier/Rectification</td>
</tr>
<tr>
<td>GMS</td>
<td>Geometric Mean Splitter</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HiRes</td>
<td>HiResolution®</td>
</tr>
<tr>
<td>HiRes 120</td>
<td>HiResolution® with the Fidelity 120™ option</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IDN</td>
<td>Input Distribution Network</td>
</tr>
<tr>
<td>IHC</td>
<td>Inner Hair Cell</td>
</tr>
<tr>
<td>ILD</td>
<td>Interaural Level Difference</td>
</tr>
<tr>
<td>ITD</td>
<td>Interaural Time Difference</td>
</tr>
<tr>
<td>KCL</td>
<td>Kirchhoff’s Current Law</td>
</tr>
<tr>
<td>2TS</td>
<td>Two-Tone Suppression</td>
</tr>
<tr>
<td>LDSS</td>
<td>Log-Domain State-Space</td>
</tr>
<tr>
<td>LI</td>
<td>Lateral Inhibition</td>
</tr>
<tr>
<td>LP</td>
<td>Low Pass</td>
</tr>
<tr>
<td>LPF</td>
<td>Low-Pass Filter</td>
</tr>
<tr>
<td>max</td>
<td>Maximum</td>
</tr>
<tr>
<td>MUX</td>
<td>Multiplexer</td>
</tr>
<tr>
<td>NH</td>
<td>Normal Hearing</td>
</tr>
<tr>
<td>NMOS</td>
<td>N-channel MOSFET</td>
</tr>
<tr>
<td>OHC</td>
<td>Outer Hair Cell</td>
</tr>
<tr>
<td>Op-amp</td>
<td>Operational Amplifier</td>
</tr>
<tr>
<td>OTA</td>
<td>Operational Transconductance Amplifier</td>
</tr>
<tr>
<td>OZGF</td>
<td>One-Zero Gammatone Filter</td>
</tr>
<tr>
<td>OZGF-LI</td>
<td>One-Zero Gammatone Filterbank with Lateral Inhibition</td>
</tr>
<tr>
<td>PMOS</td>
<td>P-channel MOSFET</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square value</td>
</tr>
<tr>
<td>SFG</td>
<td>Signal-Flow-Graph</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SPEAK</td>
<td>Spectral Peak</td>
</tr>
<tr>
<td>SS</td>
<td>State-Space</td>
</tr>
<tr>
<td>SRT</td>
<td>Speech Recognition Threshold</td>
</tr>
<tr>
<td>ST</td>
<td>Scala Tympani</td>
</tr>
<tr>
<td>THD</td>
<td>Total Harmonic Distortion</td>
</tr>
<tr>
<td>TL</td>
<td>Translinear</td>
</tr>
<tr>
<td>VLSI</td>
<td>Very Large Scale Integration</td>
</tr>
<tr>
<td>WI</td>
<td>Weak Inversion of MOSFET</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

Over the past 30 years, cochlear implant (CI) have developed from a device that was thought impossible for speech recognition and useful only for sound perception to an established clinical device for restoring partial hearing to deaf people. At present, there have been more than 100,000 CI users worldwide. Despite great progress, there remains much room for improvement in the CI performance. The best-performance CI users still do not hear as well as normal-hearing (NH) people, especially in noisy listening environments – what they often encounter in daily life. While understanding speech well (in quite), CI users have poor ability to perceive music and other more complex sounds. In addition, the signal processors in current CIs heavily rely on digital technology (e.g., DSP via ASIC processors) and thus have a high power consumption compared with the micropower systems of modern analog technology. This impedes the feasibility of fully implantable (thus invisible) CIs of the future, which have very stringent power consumption requirements.

On the other hand, the human auditory system—with the cochlea as a biological front end—is known to have superb performance, e.g., an amazingly large dynamic range (DR) of 120dB, a very low power consumption (just microwatts of power) and a high robustness to noise even in comparison with state-of-the-art speech-recognition systems and front ends. Unfortunately, current CIs provide only a very crude approximation to the biological processing of sounds*; some important biological mechanisms—which likely account for the high performance of the auditory system—are therefore partly or entirely missing in CIs. The research described in this thesis attempted to reproduce one

*It is surprising to see that some CI users can still perform well in speech-recognition tasks with such crude representations of sounds.
of the biological mechanisms—which may partly account for the high noise robustness of the normal auditory system—through a novel filterbank architecture for the speech processors in CIs, which is amenable to low-power analog VLSI implementation. In other words, the research aimed at providing a potential and quite natural solution to the noise susceptibility of current CI users, which at the same time has potential for use in the next-generation CIs fully implanted inside the human body. The research also investigated the possibility of building from transistors such a CI processor using the proposed architecture. The research underpins the development of high-performance CI processors that are 1) noise-robust and 2) fully-implantable.

To put the reader into context, we begin with a brief review of some important aspects of normal hearing as well as the present state and future direction of CIs. At the end of this chapter is an overview of this thesis, where we summarize our research efforts presented respectively in the following chapters.

1.2 Aspects of Normal Hearing

![Figure 1.1: A simplified diagram (not to scale) of the human ear; from Loizou [1]](image)

As depicted in Figure 1.1, the human ear consists of the external, the middle, and the inner ear. In normal hearing, sound waves in air are picked up by the pinna, a visible part of the external ear, and then funneled to the eardrum via the ear canal. The external ear
amplifies frequencies selectively in a manner related to the location of sounds (due to its anatomical structure)—a direction-dependent transfer function—which provides important cues for sound localization. It also provides protection of the eardrum and the middle ear beyond against direct injury. The ear drum, which separates the middle ear from the external ear, vibrates when sound waves travel through the ear canal and hit on it. Its vibrations cause the movement of a chain of small bones (called ossicles) attached at one end to it in the middle ear. The movement induces pressure oscillations (waves) in the fluids filled in a snail-shell-like cavity called the cochlea, which is located in the inner ear and attached to the other end of the ossicular chain. Thus, the middle ear couples acoustic energy to the cochlear fluids—it provides impedance matching between air and fluid.

The fluid pressure oscillations in the cochlea lead to the travelling-wave-like motion of the *basilar membrane* (BM), a membrane dividing the cochlea lengthwise and having non-uniform width and stiffness along its length. The BM is narrow and stiff near its base (close to the middle ear) while it is wide and flexible near its apex (another end). These graded properties give rise to a spatial distribution of displacement maxima for different frequencies along the BM: each frequency (pitch) will create a maximum displacement at a particular place on the BM. This *frequency-place* mapping is referred to as the *tonotopic organization* (see Figure 1.2) and exploited by our auditory system to encode frequency. Another way of its description is that each place on the BM is most responsive to a particular frequency. Correspondingly, there are two common ways of representing the BM displacements: (1) plots of the displacement amplitudes versus the distance from the BM’s base (the distance-axis is convertible to frequency-axis in terms of the frequency-place map), in response to different pure-tone input stimuli (see Figure 1.3); (2) plots of the amplitudes versus the stimulus frequency (while the stimulus amplitude is fixed), at different BM places (see Figure 1.4).

The above two representations can be regarded as the transfer functions of the cochlea for a specific input stimulus and for a specific BM place respectively. It is interesting to see in Figure 1.3 and Figure 1.4 that the cochlea transfer function is an asymmetric bandpass-like function: the filter shape is broader at the lower-frequency side (low-frequency tail) than at the higher-frequency side (roll-off). Furthermore, the function was found to be level-dependent (nonlinear)—its gain and shape change with input sound levels; specifically, for high input levels, it becomes broader, and its peak gain and peak
frequency (i.e., the most responsive frequency) becomes lower, and vice versa for low input levels [2-6]. It is the cochlea nonlinearity that considerably boosts the dynamic range of the human ear – an amazingly wide range of 120dB.

Figure 1.2: Frequency-place map (in Hz) on the BM (snail-shaped); from Loizou [1]. High frequencies are mapped to the places near the base (since they produces maxima there), whereas the low frequencies are mapped to the places near the apex.

Figure 1.3: BM responses versus location (in mm) for different stimulus (pure tones of 300, 200, 100, and 50Hz); the stapes is one of the three ossicles in the middle ear and attached to the base of the cochlea; the BM is about 35mm long; from Békésy [7].
Besides the amplitude responses, the phase responses (level-dependent) also vary along the BM, and the phase lag increases as the stimulus frequency increases (see the lower panel of Figure 1.3). These relations between amplitude/phase and frequency/place are accounted for by the distributed filtering performed by the BM. In the cochlea, BM responses are sensed by the receptor cells attached to the top of the BM—the inner hair cells (IHCs) and the outer hair cells (OHCs)—the sensory hairs (called stereocilia) of these cells are bent in accordance with the BM displacements occurring at their locations. The bending of the IHC stereocilia gives rise to changes in the amount of neurotransmitter released at the base of the IHC, initiating the firing of the adjacent auditory neurons in a manner that reflects events at the BM. The resulting impulses, which contain information about the acoustic signal, are then transmitted to the brain via the auditory nerve. Thus, the IHCs act as a signal transducer that converts mechanical vibrations into neural impulses. On the other hand, the bending of the OHC stereocilia produces electrically-driven changes in the length (known as electromotility) of the OHCs, which in turn exert forces on the BM, reducing the damping of the BM motion when the sound input is weak and until the signal reaching the brain is strong enough. In this way, the OHCs act as a feedback amplifier that accounts for the foregoing cochlear nonlinearity which is compressive. All the above hearing process is summarized in Figure 1.5 (see the left panel).

1.3 Deafness and Cochlear Implant (CI)

The most common cause of deafness is the loss of hair cells in the cochlea due to genetic defects, or certain diseases (e.g., meningitis), or certain drug treatments (e.g., streptomycin), etc., which in turn causes degeneration of adjacent auditory neurons [1, 8]. The auditory system therefore fails to transform acoustic pressure waves to neural
impulses. A cochlear prosthesis (via cochlear implant (CI)) is based on the idea of bypassing part of the impaired auditory pathway including the missing hair cells and directly stimulates the surviving neurons in the auditory nerve via electrodes (see Figure 1.5).

A CI device mainly consists of (see Figure 1.6):

- A microphone for sensing sound and converting sound into an electrical signal.
- A speech processor for decomposing the signal into its frequency components (via a filterbank with bandpass channels) and creating a set of electrical stimuli.

Figure 1.5: Block diagram of the normal auditory pathway (left panel; adapted from [9, 10]) and that rebuilt with a cochlear implant (CI) (right panel). The primary functions are indicated for each block in the normal auditory pathway.
- A transmission system to transmit the electrical stimuli from the external speech processor to the implanted electrodes via a radio frequency (RF) link.
- An implanted electrode array (multiple electrodes) inserted into the scala tympani (ST), one of the three chambers along the length of cochlea, by surgeon, and thus placed close to the auditory nerves targeted for stimulation.

Figure 1.6: A typical cochlear implant (CI) using the continuous interleaved sampling (CIS) processing strategy, whose details are given in Section 1.4; from Loizou [1]. The top panel shows the major components of the CI while the bottom panel describe how the microphone input (the syllable “sa”) is transformed into stimuli via the CIS processor.

Figure 1.5 compares the normal auditory pathway for acoustic hearing and the pathway rebuilt with a CI for electric hearing. To minimize the performance difference between the two requires that sufficient aspects of normal hearing (auditory functions) are accounted for/reproduced by the CI. For instance, the distributed filtering (via the BM) together with the gain control (via the OHC) is lumped into the function of the speech processor in the CI; the sound collection and localization (via the middle ear) as well as the mechano-electric transduction (via the IHC) are accounted for by the microphone – a
mechano-electric transducer having directional characteristics, whereas the air-to-fluid impedance match (via the middle ear) is not required for the CI. A CI with high bio-fidelity is desirable because it may push the electric hearing closer to normal (acoustic) hearing.

The use of multiple electrodes in current CI devices is motivated by the tonotopic organization of the BM. When electrodes are implanted as an array, different electrodes placed along the length of the BM may excite different subpopulations of auditory neurons nearby; since the electrodes get stimuli of different frequencies from different channels, the electrical stimulation of the corresponding auditory neurons is frequency-dependent. Specifically, the electrodes near the base and the corresponding neurons are stimulated with high-frequency signals, whereas those near the apex are stimulated with low-frequency signals. This biomimetic, tonotopic “tuning” of neurons provides important cues indicating the presence of sounds with different frequencies, i.e. encoding frequency. On the other hand, sound level, which is perceived as loudness, can be coded or controlled by the amplitude of the stimulus (current signal), which affects the number of neurons activated (when stimulated); that is, the higher the amplitude, the larger the number of neurons activated, and thus the sound is perceived as louder, and vice versa when the amplitude is lower.

There are three factors affecting the efficiency of frequency coding in CIs [1, 11]:

(1) The number and location of the electrodes along the length of the cochlea. The number of electrodes affects the place resolution for coding frequency (i.e., spectral resolution in CIs). In principle, the larger the number, the better the resolution, but the actual resolution is constrained by the following two factors (2) and (3). In addition, the depth of electrode insertion is limited by the anatomical structure of the ST, e.g., the decreasing lumen from the base to the apex, and is typically much less than 30mm*, e.g., 18-26mm, as opposed to the cochlea’s total length (about 35mm). In an obstructed ST (e.g., due to bone growth), the allowable depth can be even shallower. This limit of insertion depth may result in mismatches between frequency-place maps reproduced via CIs and the original (ideal) one used by the normal auditory system; that is, the frequency content of the acoustic signal analyzed by a CI does not match (well) with the place along the cochlea (BM) in terms of the normal tonotopic map.

* 30mm is the maximum depth achievable in current CIs.
(2) The number of surviving auditory neurons that can be stimulated. The neuronal survival ranges from sparse to substantial, and the pattern of survival is generally not uniform along the cochlea – more neurons or less ones survives somewhere, which degrades an ideal tonotopic representation of frequency. Thus, electric hearing performance largely depends on the stimulation occurring in the regions where a large number of neurons survive (spectral resolution is better there). In addition, the survival situation (both the survival number and pattern) vary from patient to patient. Thus, CIs need to be fitted \((\text{channel-by-channel})\) to patients individually for optimizing their hearing performance.

(3) Spatial spread of excitation during stimulation, or channel interaction. That is, the stimulus at one electrode stimulates not a single but several sites of neurons, which impairs the \textit{spatial specificity} of stimulation and hence spectral resolution. Channel interaction is largely due to overlaps (interferences) between the electric fields from adjacent (or more distant) stimulating electrodes, or more specially, the vector summation of the electric fields in the overlapping region. Because of the interferences, current CIs support no more than 4–8 independent stimulus sites (effective channels) \([12, 13]\), although their electrode arrays consist of a total of 12–22 electrodes.

Among the above three factors (limitations), CIs cannot do much about the second one because it is etiology-based (despite patient-fitting), while the first and the third may be improved respectively by a) optimizing the frequency-place maps in CIs by fully or partially restoring \(^*\) them to the matched ones \([14-16]\), in a \textit{patient-specific} manner (since different patients can have very different depths of electrode insertion), and b) placing electrodes closer to the targeted neural structures for decreasing electric field overlaps to some extent. Despite this, significant improvement may well need strong support from \textit{fundamentally} new designs of electrodes yielding greatly increased number of effective (independent) channels. Channel interaction—as a major limitation of present CIs—its minimization has motivated the birth of \textit{Pulsatile} processing strategies adopted by all the present CIs in widespread clinical use for transforming microphone inputs into stimuli. These processing strategies are presented in the following section.

\(^*\) The restoration can be implemented via manipulating the bandpass filter for each channel.
1.4 Present CI Processing Strategies

This section presents briefly how acoustic signals picked up by the microphones of CIs are transformed into a set of electrical stimuli – this part of design is commonly referred to as the processing strategy. The most widely used processing strategy is the continuous interleaved sampling (CIS). The operation of CIS is described in Figure 1.7: the signal is first pre-emphasized and then processed by a bank of bandpass channels where the envelopes of the filtered signals are extracted via rectification followed by low-pass filtering (LPF), then compressed via nonlinear (e.g., logarithmic) input-output mapping, and subsequently employed to modulate trains of biphasic electrical pulses (current signals), which in turn are delivered to appropriate electrodes as stimuli.

CIS got its name from the sampling (via pulse modulation) of envelope signals which is continuous and “non-overlapping-in-time” (or interleaved) across channels. Indeed, this can be regarded as the most important feature of CIS, since it effectively reduce channel interaction by avoiding stimulating electrodes simultaneously (sequentially instead). In fact, this feature has been shared by all the present processing strategies in widespread use. The interleaved form of the pulse representation for CIS is illustrated in the lower panel of Figure 1.7, where the “Simulation cycle (frame)” from the 1st electrode to the Nth one is equal to the “1/pulse-rate” on each electrode.

An important processing stage of CIS is compression, which is actually necessary for all CIs and all the CI processing strategies because the dynamic range (DR) of acoustic input signals is much wider than that of electric hearing of patients (the former is up to about 100dB, whereas the latter is about only 3–20dB). In clinical use, the function of compression in CIs is to map the acoustic signal of a wide DR into a narrow electrical (current) range between the threshold and the most comfortable level of electric hearing. Compression in CIs can be implemented via nonlinear mapping functions (i.e., known as instantaneous compression) as shown in the figure, or via automatic gain control (AGC), or both (in this case, AGC is typically placed before the filterbank, i.e., broadband AGC, whereas nonlinear mapping is performed within each channel). Before the compression stage is the envelope detection via a rectifier followed by a LPF whose cutoff frequency is typically set at 200Hz or higher to cover the fundamental-frequency* range (i.e., pitch range) of a human voice, e.g., 120Hz for a standard male voice (the processing attempts

*Fundamental frequency (denoted $F_0$ and commonly referred to as pitch) of a speech signal is the rate of vibrations of vocal cords producing speech.
to preserve in the stimuli this portion of speech spectral content important for speech perception).

Figure 1.7: Block diagram of continuous interleaved sampling (CIS) processing strategy (LPF: low-pass filter; EL-\(i\): the \(i\)th electrode). The upper panel describes the architecture implementing CIS, while the lower panel illustrates the interleaved form of stimulus pulses sent for the modulation with the signal envelopes.

Table 1-I provides an overview of present CI processing strategies in widespread clinical use. As mentioned earlier, they all present stimulus pulses that are non-overlapping in time across channels, whereas the differences among them mainly lie in the manner of delivering the interleaved pulses to appropriate electrodes. Specifically, unlike the CIS
strategy where each electrode will be stimulated once per cycle of stimulation, the “N-of-M” type strategies, including the advanced combination encoder (ACE) and the spectral peak (SPEAK) strategy, deliver the interleaved stimulus pulses only to the N electrodes corresponding to the N channels with the highest envelope amplitudes, among an array of M electrodes (M channels).

Selectively stimulating the “loudest” electrodes among many per cycle can preserve the most important aspects of the acoustic signals while allowing for higher pulse rates at the selected electrodes, resulting in higher temporal resolution for better representation of fine temporal variations on the envelope. Nevertheless, the trade-off between the pulse rate and the preservation of information still depends on the practical value of the parameter N. A larger N tends to makes the preservation better by decreasing the possibility of the omissions of those channels containing significant information during the selection, but also results in lower pulse rates and hence lower temporal resolution as noted previously, and vice versa. The SPEAK strategy attempts to optimize the trade-off by making N variable from one stimulation cycle to the next, rather than fixing its value like ACE. Specifically, the value of N per cycle in SPEAK is made adaptive to the level and spectral composition of the microphone input signal; for instance, N becomes larger for broadband spectra while lower for spectra with limited spectral content. This strategy essentially implements an adaptive stimulation rate for preserving both spectral and temporal information.

In summary, the N-of-M type strategies can be regard as variations of CIS, where electrode stimulation is not only sequential as in CIS, but also selective (adaptively or not) per cycle. Another close variation of CIS is the HiResolution® (HiRes) strategy, whose stimulation rates and cutoff frequencies of the envelope detectors are relatively higher than those used in the standard CIS. It provides higher temporal resolution because of faster stimulation, and in particular, its higher cut-off frequencies for the envelope detection allow more phase information, or “fine structure” (FS) information to be preserved in the modulation waveforms. In a signal, FS refers to the high-frequency portion (carrier) modulated by the slow varying signal envelope, which can be extracted from the original signal through the Hilbert Transform; FS information is carried in the instantaneous phase, or instantaneous frequency (i.e., the first derivative of the phase).

† Thus, FS information is also referred to as “FM information” (FM – frequency modulation) in some psychoacoustic literatures.
Table 1-I: An overview of processing strategies (stimulation form & advantage) in current use for CIs and the manufacturers that adopt them respectively [11].

<table>
<thead>
<tr>
<th>Processing Strategy</th>
<th>Stimulation form</th>
<th>Motivation /Major Advantage</th>
<th>CI Manufacturer</th>
</tr>
</thead>
</table>
| Continuous Interleaved Sampling (CIS) [32]    | ▪ Stimulus pulse trains interleaved in time (nonsimultaneous) across electrodes. | ▪ Minimizing channel (electrode) interaction                                    | ▪ MED-EL Medical Electronics GmbH*  
▪ Cochlear Ltd†  
▪ Advanced Bionics Corp‡ |
| N-of-M [33] & Advanced Combination Encoder (ACE) [34] | ▪ Channel selection scheme – only N out of M (N < M) channels (and associated electrodes) with the largest energy (maximum envelope amplitudes) are selected for stimulation per cycle. (The two strategies have essentially identical designs, and fall under a category of the “N-of-M” type strategies). | ▪ Higher pulse rate at the selected electrode and hence better temporal resolution | ▪ MED-EL Medical Electronics GmbH (N-of-M)  
▪ Cochlear Ltd (ACE) |
| Spectral Peak (SPEAK) [35]                   | ▪ “N-of-M” type with adaptive N (allowable to vary from one cycle to next according to the spectral composition of the input signal). | ▪ Adaptive stimulation rate for preserving both spectral and temporal information | ▪ Cochlear Ltd |
| HiResolution® (HiRes) [36]                   | ▪ A close variation of CIS (up to 16 channels) with relatively high stimulation rates and cut-off frequencies for the envelope detection. | ▪ Improved temporal resolution and better representation of “fine structure” (FS) information | ▪ Advanced Bionics Corp |
| HiRes with the Fidelity 120™ (HiRes 120) [37] | ▪ A variation of HiRes with “virtual channels” constructed by simultaneous stimulation of adjacent electrodes with a manipulation of the ratios of stimulus pulse amplitudes between them. | | ▪ Advanced Bionics Corp |
| Fine Structure Processing (FSP) [21, 22]     | ▪ A short group of pulses presented at the zero crossings (as opposed to the continuous manner used in CIS) in 1~4 lowest-centre-frequency channels, whose magnitudes are determined just as in CIS.  
▪ Using CIS for other remaining higher frequency channels. | ▪ Increased transmission of FS information | ▪ MED-EL Medical Electronics GmbH |

* Innsbruck, Austria.  
† Lane Cove, Australia.  
‡ Valencia, California.
Over the last several years, there has been increasing attention paid to the representation of FS information in CIs because of its suggested importance for speech (especially in noise) and music perception in conditions of low spectral resolution (e.g., no more than 4–8 effective channels in current CIs) [17-22]. In the envelope-based strategies described so far, FS information is mostly limited to the cutoff-frequency range for the envelope detection, which is typically 200–400Hz or even higher in HiRes; within this low frequency range, FS information is preserved well in the modulation waveform, whereas at higher frequencies the information is largely discarded at the envelope-extraction stage. Nevertheless, some FS (or fine-frequency) information may also be conveyed by a “channel-balance” cue to pitches between the centre frequencies (CFs) of adjacent channels; that is, more available pitches and stimulus sites are obtainable for each bandpass range via simultaneous or fast sequential stimulation of adjacent electrodes, which may allow CI users access to relatively small differences in frequency of sound [11, 23]. The practical values of these intermediate pitches can be controlled by adjusting the relative strength (amplitude) of stimulus pulses at the corresponding adjacent electrodes. Such a channel-balance cue may be inherent in the CIS and other envelope-based strategies providing sequential stimulation [24, 25], while there is a new strategy explicitly designed for exploiting the channel-balance cue.

This strategy is called the HiRes with the Fidelity 120™ option (HiRes 120), which is a variation of HiRes (and CIS) and equipped with so-called “virtual channels” [23, 26-31] (i.e., channels that contain intermediate “virtual” pitches) between adjacent electrodes. HiRes 120, as indicated by its name, provides 120 sites of stimulation (120 channels including “virtual” ones), where every 8 stimulus sites are allocated to a different bandpass range among 15 ones, corresponding to 8 “subbands” within that range, and are produced using 8 different amplitude ratios of stimulus pulses delivered to the two adjacent electrodes simultaneously. The differences between HiRes 120 and CIS are 1) using multiple sites, rather than a single site, for each bandpass range and 2) stimulating two adjacent electrodes simultaneously, rather than a single electrode, every stimulus update (whereas stimulus pulses are interleaved across channels and associated sites just as in HiRes and CIS). At each update, HiRes 120 selects one out of the eight amplitude ratios for each bandpass range to create a site whose corresponding subband carries the largest energy at that time (the selection of the ratio is therefore spectra-maxima based as in N-of-M type strategies).

* The term “current steering” used in some of these literatures refers to the concept of “virtual channels”.

- 32 -
“Fine structure processing” (FSP) is another processing strategy explicitly designed for better representing FS information with CIs. The idea behind this strategy is that FS information may be represented by the timing of pulse presentations within channels. In FSP, such a timing (temporal) code is captured at the positive-going zero crossings of the output signals of one channel or up to four ones with the lowest CFs, and captured by presenting at each zero crossing a short sequence of pulses with a length related to the upper corner frequency of the respective bandpass channel. CIS processing is preserved in the remaining channels with higher CFs, and pulse presentations for all the channels are non-overlapping-in-time as in CIS. A significant difference between FSP and CIS is that FSP’s pulse presentations for low- to mild-frequency channels are triggered by the zero crossings and thus not continuous as in CIS. FSP may allow CI users better access to FS information because the repetition rate of the pulse sequences is not fixed but instantaneous, and equals the instantaneous frequency of the bandpass signal – the carrier of FS information, as noted previously.

There have been several evidences [21, 22, 37, 38] to support the benefits/advantages of the HiRes 120 and the FSP strategies for the transmission of FS information in CIs and for speech and/or music perception via CIs, and further evaluations of the two strategies are in progress. On the other hand, it still remains unclear—with present envelope-based processing strategies—how much FS information is presented and more importantly, perceived (utilized) actually by CI users; probably, these strategies transmit only a small amount of FS information (below the cut-off frequency of the envelope detection). A well-known difficulty/limitation shared by all the present processing strategies is the lack of channel independence caused by channel interaction, although this issue has been partly alleviated by means of nonsimultaneous stimulation (overlaps in the electric fields still exists). Because of this issue, the intended high spectral resolution and hence better representation of FS information via HiRes 120 may be degraded; in fact, the quite large number of stimulus sites and corresponding pitches provided by HiRes 120 does not guarantee a large number of effective (independent) channels for the CI using this strategy. In the next section, we will present two recent advances in the CI design and performance as well as some possibilities for further improvements.
1.5 Recent Advances and Future Possibilities

In recent years, CI designers have made two major steps forward, each of which has provided large improvements in speech perception performance. The two efforts are as follows [11, 39]:

(1) Combined electrical and acoustical Stimulation (EAS) for patients with some residual hearing at low frequencies. The idea behind this strategy is that a cochlear prosthesis (i.e., a CI system) should preserve well and make full use of any residual function. Studies [40, 41] have suggested that if the residual hearing of a patient, normally at low frequencies, was preserved following implantation, then acoustical stimulation occurring at the cochlea’s apical positions’ (i.e., the residual hearing region) would complement electrical stimulation provided by a CI at more basal positions. The preservation of the residual hearing can be achieved by surgical techniques (e.g., a deliberate short electrode insertion to avoid the damage of the apical part of the cochlea and the damage of the residual hair cells there), drug therapies, and special electrode designs (e.g., thin and highly flexible electrodes). On the other hand, the improvement of the residual hearing can be provided by a hearing aid acoustically amplifying low frequencies.

Compared with either electrical or acoustical stimulation alone, EAS, which combines the two, provides large improvements in speech perception in quite or noisy listening conditions [39, 40, 42-50]. EAS also provides large improvements in music perception compared with electrical stimulation alone† [43, 46, 49-52]. The improvements provided by EAS may arise from a more natural and accurate representation of low frequency information via acoustical stimulation because of the normal or nearly normal auditory pathway preserved for these frequencies, as opposed to a crude representation of information from electrical stimulation over the same frequency region. In particular, FS information in the low-frequency region may be represented in a more natural manner so that this information can be largely or fully perceived and utilized by EAS patents.

(2) Use of bilateral CIs for electrical stimulation of both ears. This strategy may reproduce (partially) the interaural time difference (ITD) and the interaural level difference (ILD) as sound localization cues, since bilateral CI users can make use

---

* Apical positions of a cochlea (BM) correspond to low frequencies in terms of the frequency-place map.
† Music perception performance is very similar between EAS and acoustical stimulation alone.
of stimuli from both sides of their head. ITD is the difference in time between the arrivals of a sound at the two ears respectively, which indicates the probable sound source direction. ILD, the sound level difference between the two sides, is another important cue for sound localization and caused by the “head shadow” (i.e., a sound from one side is obstructed, or more specifically, absorbed and diffracted by the head so that its amplitude decreases when it reaches another side—a shadowed region for this sound).

Unsurprisingly, bilateral CIs provide improved sound localization abilities, compared with unilateral implants where such abilities, in fact, are totally or almost missing [53-60]. In addition, the “head shadow” effect reproduced via bilateral CIs provides a large benefit for understanding noisy speech in cases when the target speech and competing noise are spatially separated [39, 53, 54, 56, 61, 62] (such cases are often encountered in daily life); in fact, this purely physical effect improves speech-to-background ratios at either ear. Another potential benefit of bilateral CIs is an increased number of effective channels via doubling or nearly doubling the number of stimulus sites (electrodes), compared with unilateral implants.

The above two are new approaches, whose further optimization for better performance is possible and in progress. There are also other possibilities for improvements in the CI design and performance, such as: 1) new design and placement of electrode arrays for a larger number of effective channels in CIs, 2) improving present processing strategies or developing new ones for transmitting increased FS information and representing it in a perceivable way, and 3) a closer bio-mimicking of the signal processing that occurs in the normal cochlea, as opposed to a very crude approximation provided by current CIs to the biological processing (e.g., all the present processing strategies perform linear filtering, while the BM in the normal cochlea performs level-dependent and hence nonlinear filtering, as mentioned in Section 1.2) [11, 19, 39].

In the near future, CIs will be fully implantable; that is, all the external components of a present CI system (e.g., the microphone and the speech processor) will be implanted inside the human body. The only external component of such a system will be a remote controller for programming the implanted part. An apparent benefit provided by fully implantable CIs is that the users will be indistinguishable in appearance from normal hearing (NH) people, which can boost their self-confidence and improve third-party
attitudes to them (such improvements has been witnessed with those state-of-the-art hearing aids which are placed inside the ear canal and thus invisible or nearly invisible). In addition, fully implantable CIs have fewer limitations to daily activities of the users. For instance, the users can engage in various sports like running and swimming while they don’t need to worry about the fall-off of these devices or the damage caused by water (unfortunately, such difficulties are often encountered by current CI users who have active lifestyles and thus affect the efficiency of CIs in daily life).

From the perspective of sound perception performance, the benefits provided by fully implantable CIs include:

(1) Restoring access to the directional amplification function of the pinna—the “antenna” part of the external ear (noted in Section 1.2)—by implanting the microphone inside the ear canal as in state-of-the-art “in-the-canal” hearing aids, which likely results in improved sound localization abilities.

(2) Eliminating the limitations in the data bandwidth of the RF link transmitting the processed microphone input to electrodes, since this link is not required for the commutation between the internal speech processor and electrode array in a fully implanted CI system. These limitations can lead to restrictions of available types and rates of stimuli from the external speech processor in current CIs [11]. On the contrary, the elimination of these limitations allows for higher temporal resolution for representing the stimuli, which may improve the CI performance.

To be fully implantable requires special efforts to minimize the power consumption and silicon area for a CI system. Owing to advances in the design of CI electrodes—which have allowed placing electrodes closer to the target neural structures—the power necessary to stimulate the cochlea (via electric current) has been reduced, compared with the past design, so that it can be comparable to or less than that consumed by the speech processor for signal processing. Unfortunately, current CI speech processors rely heavily on the use of digital technology (e.g., DSP via ASIC processors); this indeed provides some desirable features such as programmability and robustness, but also makes these processors suffer from relatively high power and area costs paid—compared with analog solutions—when the precision required at the processor output is low * [63, 64] — a typical case in CI applications where a CI processor’s channel bandwidth is a few kHz at

* However, digital systems are more efficient (lower power and area costs) than analog ones when high output precision is required, because analog approaches need to pay high costs to achieve high precision.
most (already sufficient) and a patient’s acceptable DR is only 3–20dB. As a result, batteries in present state-of-the-art digital CI processors need recharge every day or every two days. This is certainly undesirable for fully implantable CIs, which have very stringent requirements on the power consumption for signal processing. On the other hand, the requirement of completely reprogramming DSPs as in the past CI systems, whose processing strategies (algorithms) were in their infancy and under development, has been eliminated with the maturing of CI processing strategies (in other words, for an algorithm known to work, it is unnecessary to attain so high programmability).

The above facts have motivated multiple efforts [65-69] in recent years to use low-power analog signal processing for CIs. For instance, a low-power CI processor proposed in [66] consumes only 211 µW—one-twentieth of the power consumption in current CI processors—while providing 7-bit output precision for each channel. This allows for a 30-year operation of this processor on a 100mAh battery rechargeable 1000 times (via wireless) without needing surgery to replace the battery [66, 70]. A significant feature of this processor is that the power-efficient analog processing is performed prior to digitalization (i.e., analog-to-digital, or A/D), rather than the A/D-then-DSP design adopted in current CI processors; the analog preprocessing can first extract meaningful information while filtering out that unnecessary, reducing the amount of information needing to be digitized, so that the following digitalization is allowed to be low-precision and low-speed. In contrast, current CI processors digitize and then process much more data than is needed, at high precision and at high speed, which can account for their relatively high power costs.

Unlike digital approaches—where transistors are used as simple switches so that a lot of power is consumed during transistors’ switching on or off—analog technology treats transistors’ physical properties (or specifically transistors’ physical relationship to voltage and current) as an asset and well exploits these properties; this is where this technology’s power efficiency comes from. Our work attempted to capture this efficiency with analog solutions for CI processing, and the proposed processor architecture is intended for use in next-generation fully implantable CIs. The architecture is presented in Chapter 2 and its low-power analog circuit implementation is described in Chapter 4.

In conclusion, the future of CIs is bright: the possibilities described in this section—as well as others not presented here—for further improvements in CIs may eventually let
CI users have appearances indistinguishable from NH people and have hearing abilities comparable to NH people.

1.6 Thesis Overview

This thesis is organized into chapters that present the aspects of this work:

- Chapter 2: “OZGF-with-LI: A Biomimetic Filterbank for Noise-robust Speech Processing in Cochlear Implants”. This chapter details the design and simulation of a newly proposed filterbank-architecture for noise-robust, high-biofidelity CI processors. The architecture was inspired by a well-known biological mechanism underlying spectral enhancement in the auditory system called lateral inhibition (LI) – a biological local “winner-take-all” link between neurons and associated frequencies. By simulating this mechanism, the architecture provides a CI pre-processing strategy which naturally enhances spectral prominences in speech, thereby making spectral features of speech more robust to masking effects of noise. In addition, each channel of the architecture implements a bio-mimetic transfer function called One-Zero-Gammatone-Filter (OZGF) [71] – a robust foundation for modelling various auditory data. Therefore, the “OZGF-with-LI” architecture can provide a closer bio-mimicking of the cochlear processing.

This chapter presents various MATLAB simulation results that illustrate the workings and benefits of the architecture and particularly show how parametric variations affect these results. Several alternative topologies were also investigated and compared in terms of performance and complexity; this effort was intended to achieve a good compromise between the two aspects of the architecture. As a consequence, the proposed architecture is amenable to low-power analog VLSI. The further evaluation and VLSI implementation of the OZGF-with-LI system are presented respectively in the following Chapter 3 and Chapter 4.

- Chapter 3: “Speech Recognition Evaluation for the OZGF-with-LI System”. This chapter considers the following question: how will the OZGF-with-LI pre-processing strategy actually affect speech intelligibility in a noisy environment? To answer this question, two perceptual tests were conducted based on the use of a noise-excited envelope vocoder acoustically simulating important aspects of CIs

---

*The term “pre-processing” refers to the signal processing performed before the pulse modulation in a complete CI processing strategy.
with an OZGF-with-LI based filterbank. This chapter details all aspects of the tests, including the speech materials used for testing, the software vocoder which process these materials, the listeners, the test procedures and the data analysis, etc. The test results revealed that the simulated LI mechanism provided a substantial benefit for listening to speech in noise.

- Chapter 4: “Ultra-low-power Analog VLSI for the OZGF-with-LI System”. This chapter details the design, fabrication and testing of an ultra-low-power analog OZGF-with-LI system employing subthreshold, or weak-inversion (WI), current-mode circuits (some prior knowledge regarding the behavior of MOS transistors in weak-inversion is assumed). This analog system is actually a VLSI-counterpart of the computational model presented in Chapter 2; this chapter describes thoroughly how this VLSI counterpart was built in a compact manner. The special techniques used in the VLSI implementation include Log-domain, Class-AB and syllabic companding via AGC. The scope of this chapter covers not only how to exploit them as low-power wide-DR solutions but also the essential ideas behind these techniques, which are presented in a brief and condensed review. The testing of the fabricated IC prototype demonstrated the ability of the analog OZGF-with-LI system to enhance spectra while consuming very low power and small silicon area.

- Chapter 5: “OZGF-with-LI: Some Likely Next Steps”. This chapter describes some likely possibilities for further improvements in our current design, which can be included in future work. These possible efforts attempt to push the present OZGF-with-LI system closer to a practical cochlear prosthesis, where programmability is regarded as critical for dealing with a potential high variation in performance across CI users, one of the known problems remain to be solved for current CIs; another purpose is an even closer mimicking of the biological cochlea processing. The end result is an ultra-low-power OZGF-with-LI based CI processor with enhanced programmability, robustness and bio-fidelity, which has potential for use in the next-generation CIs that are fully implantable inside the human body.

- The last chapter (Chapter 6) summaries my research contributions to date and describes an eventual goal of what can be done next.
References


[26] B. S. Wilson, D. T. Lawson, M. Zerbi, and C. C. Finley, "Speech processors for auditory prostheses: Virtual channel interleaved sampling (VCIS) processors-


Chapter 2

OZGF-with-LI:
A Biomimetic Filterbank for Noise-robust Speech Processing in Cochlear Implants

2.1 Introduction

Current cochlear implants (CIs) can benefit a majority of patients to achieve speech understanding in quiet listening conditions, but the performance deteriorates rapidly in the presence of competing noise. In general, CI listeners require much high signal-to-noise ratios (SNRs) to match the performance of normal-hearing (NH) listeners on speech recognition tasks in noise [1-4]. Zeng, et al. [3] found that speech recognition thresholds (i.e., SRTs, defined as the SNR that produces 50% correct recognition) of CI listeners are about 14dB higher than that of NH listeners in noise. The CI susceptibility to noise has become one of key problems that remain to be solved as it significantly affects the implant’s efficiency in daily listening situations, which are normally noisy. In this chapter, we proposed a potential solution which is bio-inspired. To put the reader into context, we starts with a short review on the causes of the CI susceptibility to noise and previous solutions to this issue (Section 2.1.1 & 2.1.2) and provide a brief summary on the efforts that we have made in this work (Section 2.1.3).

2.1.1 Causes of the CI susceptibility to noise

It has been suggested that an important contributor to the noise susceptibility of CI is likely to be the reduced spectral contrast (the difference between peaks and valleys in the spectrum) [4-6]. This makes the CI listeners difficult to pick out the spectral peaks in
speech spectra (i.e., formants) as the identity of various speech sounds and more affected by the noise filling in the valleys [7-11].

Current CIs apply compression to fit a large acoustic amplitude range into patients’ narrow electrical dynamic range, which makes sound audible while maintaining listening comfort. Specifically, each channel of the CI system contains a compression block to map its output signal into the current range between the threshold and the most comfortable level. Despite the ease of customizing the extent of compression for each channel, multi-channel compression degrades spectral contrast [6, 12-15] via an asymmetric amplification over channel-specific frequency regions: weak channels with the frequency regions corresponding to spectral valleys are strongly amplified so that they are concurrently audible with weakly amplified intense ones (corresponding to spectral peaks). This causes a smearing of spectral information, which is represented by the differences among the channel output levels and likely serves as important cues for coding of spectral shape in CIs [5, 6, 16].

Current CIs have not implemented the suppressive and inhibitory processes, e.g. two-tone suppression and later inhibition, as occurred in the normal auditory periphery system. These processes, diminished or abolished due to cochlear damage or death, are considered as the significant mechanisms that underlies enhancement of internal compressed spectra (i.e., an increase in peak-to-valley difference) and are thought to improve SNR of spectral prominences (i.e., formants)[17-23]. The term lateral inhibition (LI) refers to a neuronal-originated inhibitory process activated by lateral distribution of neuron outputs to inhibitory synapses on neighbouring sensory neurons [24]. By the process, sensory neurons with a large response reduce their own gain along with the gain of others nearby. The LI has been found in various types of sensory system (e.g., vision and hearing). Different from the neural basis of lateral inhibition, two-tone suppression (2TS), or tone-to-tone suppression, originates in mechanical phenomena at the BM contained in the cochlea. The phenomena arise from complex interactions between the OHCs and the BM [25]. The 2TS can be demonstrated as the reduction of the neural response to one tone due to the simultaneous presence of another. The suppressive effect becomes stronger when the suppressor tone is more intense than the suppressed tone and/or they are closer to each other in frequency [26]. Thus, the 2TS serves to enhance the contrast between tones.

*Formants are where energy concentrates in the sound spectrum and contain important information for speech recognition.
As a phenomenon, the 2TS is qualitatively similar to the LI, imaging the sensory neurons and those nearby respond to two different tone-stimuli respectively. Despite originating from different mechanisms, both processes operate in a nonlinear fashion and serve to sharpen the internal representation of spectral peaks, thereby preserving spectral contrast in the auditory system, which is found to be highly compressive. From the perspective of noise reduction, through these processes, spectral prominences (peaks) have the effect of suppressing their weaker surrounding noise in the spectrum, which may partly account for the superb performance of the human auditory system in noise even compared with that of state-of-the-art speech recognition devices.

The side effect of multi-channel compression, in addition to the fact that those bio-enhancement processes are entirely missing in current CI systems, results in very limited spectral contrast for the present CI users. As a consequence, CI users typically require a larger 4-6 dB spectral contrast than normal-hearing listeners to achieve high recognition accuracy in quiet listening conditions [5]. Even higher contrast is likely required when speech is embedded in noise, as noise fills in spectral valleys between peaks. In addition, the poor spectral resolution due to the limited number and location of the electrodes and electrode interactions partially accounts for the CI susceptibility to noise [4] as well.

2.1.2 Improvement of the CI performance in noise

The side effect of multi-channel compression can be thought of as convolving the spectrum with a smoothing function. As a result, the features of the original spectrum represented by a well contrast between peaks and valleys may be smoothed to such an extent that they become imperceptible if a high level of compression is performed. This problem is exacerbated when a noise background accompanying speech fills in the valleys between the spectral peaks, thereby impairing their prominence. To counteract these undesirable effects, spectra can be enhanced to have increased peak-to-valley ratios, i.e., to be sharpened, which is known as spectral enhancement. Spectral features therefore have increased robustness to noise, which likely contributes to improved speech performance in noise.

So far, a number of different spectral enhancement approaches and solutions have been proposed [27-32] (for a review, see [33]), which process the short-term spectrum by modifying the spectrum in a nonlinear manner altering spectral contrast as desired (e.g., spectral expansion), or based on the convolution with an enhancement function (spectral-filtering approaches). Although improvements via spectral enhancement were reported in
terms of speech intelligibility and subjective ratings of speech quality, Moore [33] has pointed out that the present spectral enhancement techniques has not been evaluated in real-time wearable devices, and the evaluating subjects have therefore no opportunity to experience spectrally enhanced stimuli under more realistic listening conditions. The lack of the real-time evaluation is likely because a good compromise involving quality and complexity of hardware implementation has not yet been reached, which impedes the real-time implementation of spectral enhancement. Most spectral enhancement solutions proposed so far are based on power-hungry digital approaches and thus not suitable for wearable devices, where power consumption requirements are stringent. It is also worth noting that these proposed enhancement techniques were tested with hearing-impaired listeners, rather than CI users, who have in general poorer spectral resolution due to the limited number and location of the electrodes as well as interactions between them. It is thus possible that spectral enhancement might deliver more benefits to CI users than to hearing-impaired listeners. This is supported by Loizou and Poroy’s finding [5] that the increased benefits of enhanced spectral contrast can be obtained from decreasing initial spectral resolution.

The real-time implementation and real-world uses of spectral enhancement requires explicit efforts to achieve a desirable compromise involving quality and complexity. Also, these techniques need to be amenable to power-efficient implementations. One of such efforts was presented in Turiccheia and Sarpeshkar’s work [34], where they proposed a bio-inspired companding strategy for increasing spectral contrast. This strategy simulates the 2TS via relatively broadband compression followed by more frequency selective expansion. Due to the simulated 2TS, spectral peaks have effects of suppressing their surroundings, thereby increasing contrast in the spectrum. In their work, the companding strategy was explicitly designed to be suited for low-power analog VLSI implementations. The strategy was evaluated [35, 36] with CI users and normal-listening subjects listening to acoustic simulation of CI processing; recognition of words in sentences in both group improved between 10 and 20 percentage points.

Another earlier effort was presented in the speech processor designed by Ifukube and White in 1987 [37]. This processor models the LI in a spatial-convolution manner: the channel outputs are weighted and coupled in terms of a LI function, so as to provide a soft local *winner-take-all* enhancement of spectral patterns. The parameters of the LI circuit were chosen to maximize the difference among five spectral patterns of Japanese
vowels. In particular, the processor was implemented in analog IC, and every active IC was made of CMOS operation amplifiers (op-amp) for power savings. Despite the spectral enhancement performed, the LI circuit did not show a promising ability to improve the vowel discriminations in the CI subject testing. Ifukube and White [37, 38] speculated that the benefits of modelling the LI might be severely counteracted by the strong channel interactions caused by the simultaneous stimulation used in their CI processor.

Instead of enhancing spectral features in noise, many other researchers have attempted to apply various forms of noise-reduction algorithms to reduce deleterious effects of background noise directly [39-49]. In general, these schemes are based on the use of a single microphone, or multiple ones to create a highly directional characteristic. The former, normally based on digital signal processing (DSP), has been shown to produce modest but significant improvements in intelligibility [42-44]. A common issue, however, for single-microphone noise reduction schemes, is the need to estimate the spectrum of background noise; that is, a prior knowledge of noise is required. This can be done during non-speech activity, e.g., via spectral subtraction*, which, however, has proved extremely difficult to be error-free. On the other hand, noise reduction schemes based on the use of two or more microphone can offer considerably larger benefits in speech intelligibility. By adding a second directional microphone or more, these schemes can exploit spatial information due to the relative position of the emanating sounds, and thus best suit the situation in which the target sound and background are spatially separated – a typical case in everyday life [45-49]. Multi-microphone noise reduction techniques are particularly attractive for bilateral CIs since most of these devices nowadays are fitted with either one microphone in each of the two ears or two microphones in ear.

2.1.3 Efforts in this work

This work proposes a novel filterbank architecture termed “OZGF-with-LI” for noise-robust CI processors, which models the LI across channels for spectral enhancement and performs multi-channel compression simultaneously. In the presence of the LI, multi-channel compression is performed without degrading spectral contrast, and increased spectral contrast is available. A potential solution to the CI susceptibility to noise is therefore provided. It is worth noting that spectral enhancement is a nature consequence of our trying to simulate the LI phenomena seen in the auditory system, rather than

* A well-known technique for noise reduction by subtracting the estimated spectrum of noise from that of the noisy speech
realized by processing the short-term spectrum to alter the contrast therein directly as attempted by many other researchers. Thus, spectral enhancement presented in this work is performed without a prior knowledge of both speech and noise, which tends to create a heavy computational requirement and error issues. In addition, the new filterbank architecture was explicitly designed to provide a robust foundation for modelling various auditory data, by implementing a biomimetic transfer function called One-Zero-Gammatone-Filter (OZGF) [50] within each bandpass channel.

Correspondingly, we present in this Chapter a new CI pre-processing strategy based on the OZGF-with-LI architecture, which provides a closer bio-mimicking of the signal processing that occurs in the normal cochlea compared to the present CI processing strategies. Our work essentially underpins the development of future CIs of increased bio-realism. Other efforts in this work include achieving a good compromise between performance and complexity of hardware implementation and ensuring the suitability to future low power analog VLSI implementations. These efforts are highly desirable for real-time implementation and evaluation of the new architecture as well as its potential use in the next-generation fully implantable CIs.

The rest of this chapter is organized as follows. Section 2.2 describes the proposed OZGF-with-LI filterbank architecture. The essential ideas behind this architecture are also presented. Section 2.3 presents the MATLAB simulation results that illustrate the workings and benefits of the proposed architecture and particularly show how parametric variations affect these results. Moreover, several alternative topologies were investigated and compared in terms of performance and complexity. Section 2.4 discussed detailedly the potential use and advantages of the new architecture in CIs. Section 2.5 gives further discussions on parameter optimization and conclusions to the work presented in this chapter.

2.2 Architecture Design

Figure 2.1 shows the proposed filterbank architecture where each channel implements a transfer function called One-Zero-Gammatone-Filter (OZGF) [50] (whose details are given later) together with a compression scheme (via AGC) to regulate the filter quality factor $Q$. The AGC block takes its rectified and weighted inputs from the output of its corresponding channel $i$ but also from those of the neighbouring channels (i.e., channel $i-2$, $i-1$ and $i+1$, $i+2$), thereby simulating the LI. It is worth noting that the channel outputs
are full-wave rectified before being combined to form the AGC input so as to avoid the possible phase cancellations caused by their channel-related phase differences.

2.2.1 One-Zero-Gammatone-Filter (OZGF) channels

The OZGF was proposed by Katsiamis et al. [50] for modelling in silicon the frequency-domain behaviour of the biological cochlea. The mathematical description of the OZGF is:

\[ H_{OZGF}(\omega) = \frac{K(s + \omega_z)}{[s^2 + \frac{\omega_0}{Q} s + \omega_z^2]^N} \quad (K = \omega_0^{2N-1} \text{ for dimensional consistency}) \quad (1) \]

Where \( \omega_0 \) is the natural (or pole) frequency; \( \omega_z \) is the zero frequency; \( Q \) is the quality factor; \( N \) is the filter order.

Figure 2.2 shows the frequency response of the 4\(^{th}\)-order \((N = 4)\) OZGF. By setting \( \omega_z = 0.1\omega_0 \) in (1), the low frequency tail of the response has a DC gain of -20 dB. The
response peak increases and shifts to ‘right’ in frequency when the $Q$ value is increasing from 0.75 to 10, as indicated roughly by the dashed line.

![The OZGF Frequency Response](image)

Figure 2.2: The OZGF frequency response of the order $N = 4$ with $Q$ ranging from 0.75 to 10. The frequency axis is normalized to the natural frequency.

The OZGF filtering was incorporated “defaultly” in the proposed filterbank architecture because:

(a) OZGF provides a robust foundation for modelling cochlea transfer functions involving a variety of auditory data shown in Figure 2.2 [51-55] – the passband asymmetry (represented by a broad low-frequency tail and a sharp high-frequency roll-off), level-dependent filter gain and shape, level-dependent shift of the most-responsive frequency (i.e., centre frequency (CF)) and linear low-frequency tail for frequencies well below the centre frequency. Because of the asymmetry in the filter shape, cochlea transfer functions are regarded to provide a good trade-off between temporal resolution and spectral resolution – the former is improved by the broad low-frequency tail while the latter is improved by the sharp high-frequency roll-off. Its level-dependent properties can be simulated by simply regulating the $Q$ values of the OZGF according to the channel output levels. Furthermore, we could fit the
OZGF response to various physiological measurements by customizing the N and Q values in (1) for each channel with different centre frequencies, as described in [50].

(b) OZGF delivers simple parameterization and ease of hardware implementation in analog VLSI. As depicted by Figure 2.2, OZGF is gain-adjustable with only one parameter Q, which also controls the filter shape. The filter function can be simply synthesized by cascades of biquadratic sections, since it is possible to decompose the Nth-order OZGF into a lossy bandpass (BP) biquad (two-pole, one-zero) and N-1 cascaded low-pass (LP) biquads as shown in (2).

\[ H_{OZGF}(s) = \frac{K(s + \omega_c)}{[s^2 + \frac{\omega_b}{Q}s + \omega_b^2]^N} \]

\[ = \frac{K_1(s + \omega_c)}{s^2 + \frac{\omega_b}{Q}s + \omega_b^2} \times \frac{K_1}{[s^2 + \frac{\omega_b}{Q}s + \omega_b^2]^{N-1}} \] (2)

where \( K = K_1 \times K_2 \), \( K_1 = \omega_0^{2N-1} \) and \( K_2 = \omega_0 \) to preserve dimensional consistency and to facilitate implementation.

2.2.2 Frequency-dependent Q-adaptive cross-coupled AGCs

The AGC scheme exhibits the frequency-dependent property as it acts in feedback to regulate the Q of each filter, thereby operating only for passband frequencies. As speech spectra fall off at high frequencies, high frequency emphasis is performed automatically via the channel-specific AGC scheme: spectral content at high frequencies is strongly amplified so that it become concurrently audible with that weakly amplified at low frequencies; it is allowed to pass through more sharply tuned bandpass filters (due to higher Q) and thus to be less affected by that at low frequencies, e.g., upward spread of masking of the first formant on the second and of the second formant on the third. The improved audibility (due to high gain) and improved frequency selectivity (due to high Q) at high frequencies may provide CI listeners (binaural implanted) with a better use of the “head shadow” effects that increases at high frequencies, leading to improved speech
performance in background noise by “shadowing” one of the ear from the full intensity of the noise source on the opposite side of that ear [57-59].

Spectral content at low frequencies, on the other hand, can have a better temporal representation due to relatively broadband (i.e., lower $Q$). Temporal resolution is likely not so important for high frequencies in terms of the CI perception since the CI listeners perceive primarily the low-frequency temporal features of electrical stimuli delivered to a single electrode, and are found to be able to best detect temporal modulation with modulation frequencies below 300Hz [60]. The proposed frequency-dependent and $Q$-adaptive AGC scheme can therefore provide an efficient compromise among gain distribution, frequency selectivity and temporal resolution across different frequency regions. It also produces a close intertwining of both filtering and compression as observed in the auditory system, rather than the artificial separation of filtering and compression in present CI systems.

Since $Q$ strongly correlates with the filter phase response (for a detailed mathematical description, see [50]), it is possible to simulate the level-dependent differences in phase response at different places along the BM by customizing the $Q$-regulation for each OZGF channel that maps to a different BM segment. It may therefore improve the sound-level coding in CIs which is limited by the severely degraded dynamic range that is approximately five orders of magnitude less than that of normal hearing. Heinz et al. [61] have suggested that the normal human auditory system may use the level-dependent phase cues along the BM as a partial solution to the dynamic-range problem (worse for CI listeners) that significantly affects level discrimination. This is based on some previous studies [62, 63]: such phase cues continue to encode changes in stimulus level at high level while firing rates of the majority of auditory-nerve fibers are already saturated. Current CI systems encode sound level by stimulating current or pulse width. Moore [64] has pointed out that no current CI systems reproduced level-dependent phase cues.

While performing compression, the AGC blocks are cross-coupled to simulate the LI in such a way that the filter gain of each channel is affected by neighbouring channels’ outputs. As a result, when the output of any channel with intense spectral content is fed into the AGCs of its surrounding weak channels, the gains of these weak channels will be reduced largely, whereas the gain of the intense channel is less affected. In other words, the channels with intense spectral content are enhanced relative to their surrounding weak channels, and spectral contrast across channels is therefore increased.
It is worth noting that coupling channel outputs directly as shown in the CI speech processor designed by Ifukube and White [37] can simulate the LI as well and thus perform spectral enhancement naturally. However, care must be taken to preserve channel-related cues while coupling different passbands directly so that the place mechanism for encoding frequencies in CIs can operate properly. To avoid this difficulty, we opted for the coupled AGC scheme, which is also motivated by the fact that the nonlinear mechanisms involved in the cochlear processing, for the purpose of speech analysis, can be adequately accounted for by lumping them into the action of the AGC [65]. It is also worth noting that the proposed architecture was not intended to replicate the ‘exact’ biological operations of the LI—a quite complex and expensive (in hardware) task—but to rebuild a missing bio-realistic link between bandpass channels, which exhibits a soft local winner-take-all property with an efficient compromise between quality and complexity of hardware implementation.

Previously, researches [65-69] have attempted to incorporate various forms of coupled AGC into travelling-wave cochlea models consisting of filter-cascade stages (for a review, see [70]), changing the stage gain adaptively, separately, but not independently. In 1982, Lyon proposed a VLSI-compatible computational model of filtering, detection and compression in the cochlea [65]. The DSP-based model incorporated a coupled-AGC compression network of which coupling weights were determined by the target level and time constant of the filter stage, to accomplish a local gain adaption following each filter cascade stage. After six years, Lyon’s succeeding work with Mead—“An Analog Electronic Cochlea”—updated the concept of his coupled AGC to be Q-adaptive and mimic closely the nonlinear behaviour of cochlea in such a way that when a signal of particular frequency and amplitude travels along the filter cascade, each filter stage will adjust its $Q$ value so that a small variation in the stage $Q$ can produce a large overall gain change (known as pseudo-resonance); with the coupled AGC, a large detected signal at one filter (BM) stage is able to reduce the gain at nearby stages [66]. Although Lyon presented this idea in their work, he implemented merely the BM filtering stages without AGC in analog VLSI.

Based on Lyon’s silicon BM filtering implementation, Fragnière et al. in 1997 [68] proposed an analogue VLSI model of an active cochlear with a feedback closed-loop $Q$-adaptive AGC scheme. They found that to build up a pseudo-resonance of which gain is locally controllable at the output of a particular stage, the AGC scheme including a $Q$-
control loop must be coupled in such a way that the output from any filter stage in the cascade controls the $Q$ of a stage whose centre frequency (CF) is $1/6 \sim 1/3$ of an octave lower than the CF of that particular stage. Their work essentially proposed a spatial-distribution of the feedback gains along the cascade via coupled AGC. Distributed gain control was also realized as an important aspect in Sarpeshhkar et al.’s silicon cochlea where a weighted-averaging unit was coupled to each of the energy-detected outputs from different filter taps, producing a $Q$-control signal for the local filtering [69]. Their cochlear model was shown to be able to exhibit a variety of nonlinear and active effects seen in the biological cochlea (e.g., the 2TS).

A specific effort on reproducing the 2TS accurately in the travelling-wave cochlear model was presented in Kates’ work in 1995 [67]. This work investigated the ability of his previously proposed cochlear model [35], which is based on a 1-D transmission line, to reproduce the 2TS. In the cochlear model, a $Q$-control law was specified in AGC that the control signal was taken as the largest peak detected over a region extending from three sections above to three sections below each frequency tap location. Kates found that with the coupled AGC scheme, the cochlear model was adequate to reproduce the 2TS when the suppressor was located higher in frequency than the suppressed one, but inadequate vice-versa. He therefore added at each filter tap an extra gain stage that operates with a saturating nonlinearity to compensate the insufficient low-frequency suppression. This made his cochlear model much more accurate in reproducing the 2TS behaviour.

Above work on coupled AGC refer seldomly or do not refer to spectral enhancement. We therefore clarified and detailed in our work the role of coupled AGC in spectral enhancement with a variety of simulation results (see Section 2.3). Moreover, motivated by the works of Kates and Sarpeshhkar et al. and the qualitative similarity between the 2TS and LI, we also attempted in this work to investigate whether the 2TS would be an emergent effect of our trying to simulate the LI. The coupled AGC scheme presented in this work was implemented in a bank of filters, rather than in a filter cascade, in order to make it suited for current CI devices, all of which are filterbank-based, and avoid several inherent issues involved in a filter-cascade implementation, e.g., the susceptibility to noise and offset accumulation.
2.3 Simulation Experiments and Results

We implemented the proposed architecture as shown in Figure 2.1 with \( N = 4 \) (i.e., the 4th order OZGF). The filters have pole frequencies (i.e., \( f_0 = 2\pi\omega_0 \)) logarithmically spaced between 250 and 4000 Hz* across the 16 channels†. The zero frequencies were chosen to scale with these pole frequencies (i.e., \( \omega_z = K\omega_0 \)). For all the simulation experiments, we set \( K = 0.1 \) so that each OZGF has a DC gain of -20dB as shown in Figure 2.2.

Figure 2.3 shows the block diagram of the implemented OZGF channel, which consists of a cascade of four 2nd-order filter stages—one bandpass biquads followed by three lowpass biquads—and the coupled feedback AGC regulating the \( Q \) value for each stage in terms of a particular \( Q \)-control law. The outputs of the channel \( i \) and its neighbouring channels at higher frequency side (i.e., channel \( i+1 \) and \( i+2 \)) and at lower frequency side (i.e., channel \( i-1 \) and \( i-2 \)) are first full-wave rectified (denoted by \(|u|\)) and then weighed using the factors \( W'_0, W'_1, \) and \( W'_L \) respectively (for simplicity, we set \( w_{i-2} = w_{i-1} = W'_L, w_{i+2} = w_{i+1} = W'_H \) and \( w_i = W'_0 \)) before they are added together to form the AGC input.

![figure 2.3](image)

Besides the OZGF, we developed a computational model of the \( Q \) control law in AGC (based on [56]) using MATLAB Simulink (see Figure 2.4):

**STAGE 1:** The input first goes through the low-pass filter (LPF), producing a quasi-DC output whose value corresponds to the peak value \( \hat{I}_{IN} \) of the full-wave rectified AGC

---

* The characteristics of the speech signal lies below 4 kHz.
† 16 is a typically number of channels for current CIs.
input. The pole of the LPF was chosen to scale with the pole frequency of the OZGF channel (i.e., \( w_f = a \omega_0 \) where \( a \) is a scaling factor).

**STAGE 2**: The output of the LPF is scaled with a factor \( I_{0,\text{control}} \) and then superimposed on a threshold \( I_s \) whose value is set to be larger than 1 for ensuring a positive value generated by the following logarithmic processing.

**STAGE 3&4**: The resulting signal is then processed by logarithmic and hyperbolic functions in succession, and finally multiplied by a gain factor \( (I_{\text{tail}}) \).

The final AGC output is \( \frac{\omega_0}{Q} \) rather than \( Q \) itself, which means that a large output lead to a small \( Q \) value. The mathematical description of the above \( Q \)-control law (i.e., the overall AGC input-output transfer characteristic) is

\[
\frac{\frac{\omega_0}{Q}}{Q} = I_{\text{tail}} \tanh \left[ \ln \left( \frac{I_{\text{IN}}}{I_{0,\text{control}}} + I_s \right) \right]
\]  

(3)

In terms of (3), the \( Q \)-control law is quasi-logarithmically compressive as depicted by Figure 2.4 (d), and thus has the ability to compress a potentially wide dynamic range (DR) at the channel output into a small \( Q \)-range. The parametric dependence of this control law is illustrated by Figure 2.5:

(a) \( I_s \) and \( I_{\text{tail}} \) together control the actual compression level. An increase in \( I_s \) or a decrease in \( I_{\text{tail}} \) results in an increase in the compression level. The two parameters determine the lower limit via \( I_{\text{tail}} \tanh \left[ \ln \left( I_s \right) \right] \) \( (I_{\text{IN}} = 0) \) and the upper (saturation) limit \( I_{\text{tail}} \) of the transfer characteristic and thus set a specific \( Q \) range.

(b) \( I_{0,\text{control}} \) controls the actual shape of the characteristic within the range determined by \( I_s \) and \( I_{\text{tail}} \). A decrease in \( I_{0,\text{control}} \) leads to an increase in sensitivity of the AGC transfer characteristic.

(c) For simplicity of the implementation, \( I_{\text{tail}} \) is scaled with \( \omega_0 \) to ensure the same \( Q \) range for every channel. Therefore, (3) can be re-written by substituting \( I_{\text{tail}} = m \omega_0 \) (\( m \) is a scaling factor), which gives

\[
\frac{\omega_0}{Q} = m \tanh \left[ \ln \left( \frac{I_{\text{IN}}}{I_{0,\text{control}}} + I_s \right) \right]
\]  

(4)

* The AGC sensitivity is defined as the rate at which the output grows when the input is increased.
Then, the $Q$ range is specified as:

$$Q_{\text{max}} = [m \tanh (\ln I_{th})]^{-1} \, \text{and} \, Q_{\text{min}} \approx m^{-1} \quad (5)$$
Figure 2.5: The plots of the parametric AGC transfer characteristic with varying the parameters involved in (3). The bold curves correspond to the initial setting for the simulation ($I_{sat} = 20000$, $I_d = 1/0.9$, $I_{0,control} = 0.9$)
Note that in (5) $Q_{\text{min}}$ can never reach to $m^{-1}$ practically as the upper limit of a hyperbolic tangent function cannot become equal to unity. Despite this, a very large signal can lead to a $Q$-value approaching this limit.

The initial settings of the AGC parameters for the following simulation experiments were as follow: $I_{\text{tail}} = 1.25\omega_0$ (m = 1.25), $I_{\text{th}} = 1.05$, $I_{\text{0,control}} = 0.95$, $w_i = 0.025\omega_0$ ($\alpha = 0.025$), $\omega_z = 0.1\omega_0$ ($K = 0.1$); pole frequencies (i.e., $f_p = 2\pi\omega_0$) are logarithmically spaced between 250 and 4000. The $Q$-range for each channel was 0.8~15.6 (i.e., 25dB of DR). Based on these initial settings, the effects of compression and spectral enhancement performed by the proposed architecture are investigated with varying each of these parameters individually in the following experiments.

2.3.1 Compression without Spectral Enhancement

In this experiment, $W_L$ and $W_H$ were set to be zero (i.e., no AGC coupling) to investigate the effects of multi-channel compression without spectral enhancement—a typical case in current CIs. The input to the architecture was chosen to be a synthetic vowel /u/ with the pitch $F_0$ at 100 Hz, the first formant $F_1$ at 300 Hz, the second formant $F_2$ at 900 Hz, and the third formant $F_3$ at 2200 Hz. All the channel outputs were combined to form a final single output of the filterbank for spectral analysis; this operation was repeated in all the following experiments.

Figure 2.6(a)-(d) shows the output spectra and illustrates the parametric dependence of compression with varying the weighting factor $W_0$, and the AGC parameters (i.e., $I_{\text{th}}$, $I_{\text{tail}}$ and $I_{\text{0,control}}$). For clarity, the harmonics in the spectrum were joined with lines. The bold lines represent the input spectra and the fine lines represent the compressed spectra. It can be observed that a decrease in any of the parameters $W_0$, $I_{\text{th}}$ and $I_{\text{tail}}$ or an increase in $I_{\text{0,control}}$ leads to flattened output spectra with the progressively lost formants, whereas varying these parameters in a contrary direction renders the output spectrum closer to the original input spectrum preserving the well resolvable formants. The explanations for this observation (corresponding to Figure 2.6 (a)-(d) respectively) are as follows:

(a) $I_{\text{th}}$ determines the DR of the $Q$, namely, that $Q_{\text{max}} / Q_{\text{min}} = [\tanh (\ln I_{\text{th}})]^{-1}$. A high $I_{\text{th}}$ results in a narrow DR of the $Q$ and a small $Q$ for each channel and thus multi-channel compression amplifies different frequencies with similar small gains, producing a spectrum close to the original input one. On the contrary, a low $I_{\text{th}}$

---

The first three formants (from lower frequencies to higher) are commonly denoted as $F_1$, $F_2$ and $F_3$ respectively and carry sufficient information for the recognition of voiced sounds (e.g., vowels).
results in a wide DR of $Q$ and hence strong multi-channel compression flattening the input spectrum.

(b) $I_{\text{tail}}$ determines the actual $Q$ value within the DR for a given $I_0$ since it determines $Q_{\text{min}}$ as shown in (5), where $m$ is the scaling factor. A low $I_{\text{tail}}$ (i.e., small $m$) results in high $Q$ values within the DR available for each channel, thereby improving spectral resolution but also rendering the spectral valleys well resolvable to be enhanced relative to the formants, namely, degrading spectral contrast. On the contrary, a high $I_{\text{tail}}$ (i.e., large $m$) results in low $Q$ values and thus the AGC across channels operates for a wide range of frequencies, providing similar gains to these frequencies and preserving the contrast in the original spectrum.

(c) $I_{0,\text{control}}$ determines the sensitivity of the AGC transfer characteristic as illustrated by Figure 2.5 (bottom). A high $I_{0,\text{control}}$ results in a low AGC sensitivity and thus prevent the AGC input from being mapped close to the upper limit (saturation level). On the contrary, a very small $I_{0,\text{control}}$ may force the AGC to operate almost in saturation and thus reduce the effectiveness of compression together with its side effect.

(d) The weighting factor $W_0$ also determines the AGC sensitivity given that the un-weighted channel output is viewed as the AGC input, since it functions as a gain factor similar to $1/I_{0,\text{control}}$ at STAGE 1 (see Figure 2.4).

With (a)-(d) in mind, the effects of the multi-channel compression performed by the OZGF-with-LI system depend on the sensitivity of the AGC input-output transfer characteristic (given that the un-weighted channel output is viewed as the AGC input), the DR of the $Q$ and the corresponding $Q_{\text{min}}$. With appropriate parameter settings, multi-channel compression can provide an efficient high-frequency emphasis, i.e., enhancing $F_3$ relative to $F_2$, $F_2$ relative to $F_1$. However, it also enhances the spectral valleys relative to the peaks (i.e., formants), thereby degrading spectral contrast. A solution to this issue is demonstrated in the following simulation experiments, where multi-channel compression and spectral enhancement are performed therein simultaneously.
Figure 2.6: The parametric effects of multi-channel compression.

(a).

(b).

(c).
2.3.2 Compression combined with Spectral Enhancement

The proposed system can perform simultaneous multi-channel compression and spectral enhancement when \( W_L = W_H \neq 0, W_o \neq 0 \), namely, that AGCs are coupled. Figure 2.7 compares the output spectra of the vowel input /u/ with the AGC cross-coupling on \((W_L = W_H = 0.4, W_o = 0.1)\) and the coupling off \((W_L = W_H = 0, W_o = 0.5)\). It can be observed that compression alone (AGC coupling OFF) flattens the spectrum and thus the second formant and the third formant are almost lost, whereas an active AGC cross-coupling sharpens the two formants, improving their recognition in the spectrum. For clarity of comparisons in spectral contrast, the weighting factors were chosen to yield approximately the same magnitude (dB) at the three formant frequencies in both cases. To facilitate this operation, a small gain \( K_c \) was applied to the output of each channel with different values in the two cases \((K_c = 1 \text{ in the coupling-off case})\). A change in \( K_c \) shifts upwardly or downwardly the whole spectrum without affecting the spectral contrast therein. The same procedures were applied to set the parameters for all the following experiments involving comparisons in spectral contrast.

Figure 2.8 shows that the improved spectral contrast degrades when the ratio of \( W_L \) or \( W_H (W_H = W_L) \) to \( W_o \) decreases. A decrease in these ratios reduces the simulated LI effect: the \( Q \) of the channel \( i \) is less affected by the outputs of the neighbouring channel \( i-2, i-1, i+1, i+2 \).
Figure 2.7: Input (dotted lines) and output spectra (continuous lines) of \(f/u\). The AGC coupling ON case corresponds to \(W_L = W_H = 0.4, W_0 = 0.1, K_c = 1.12 \text{ (1dB)}\). The AGC coupling OFF case corresponds to \(W_L = W_H = 0, W_0 = 0.5, K_c = 1 \text{ (0dB)}\).

Figure 2.8: Output spectra of \(f/u\) with varying the ratio of \(W_H (W'_H) / W_0\). The two AGC coupling-ON case corresponds to \(W_L = W_H = 0.4, W_0 = 0.1, K_c = 1.12 \text{ (1dB)}\) and \(W_L = W_H = 0.14, W_0 = 0.28, K_c = 1 \text{ (0dB)}\) respectively. The AGC coupling-OFF case corresponds to \(W_L = W_H = 0, W_0 = 0.5, K_c = 1 \text{ (0dB)}\).
Figure 2.9(a)-(d) illustrates how the parameter variations affect the quality of spectral enhancement. It can be observed that the enhancement of the formants is reduced and the output spectra (the dotted lines) become closer to the input (the bold line), when (1) $I_{\text{tail}}$ or $I_{\text{th}}$ is increased (Figure 2.9(a) and (b)), which results in a reduced DR of $Q$ with smaller $Q$ values as shown in Table 2-I, and (2) $I_{0,\text{control}}$ is decreased (Figure 2.9(c)), or the weighting factors with the same ratio of $W_{\perp1}(W_{\perp})$ to $W_{0}$ is increased (Figure 2.9(d)), which results in an increased sensitivity of the AGC transfer characteristic.

Above tunings impair the ability of the AGC coupling scheme to increase the difference among channels $Q$ values, thereby degrading the quality of spectral enhancement. To best enhance spectral contrast, we requires a relatively low sensitivity of the AGC input-output transfer characteristic, an adequately wide DR of the $Q$ and adequate $Q$ values within the DR, which have been achieved with our initial parameter settings.

<table>
<thead>
<tr>
<th>$I_{\text{th}}$</th>
<th>$1.25\omega_0$</th>
<th>$3\omega_0$</th>
<th>$10\omega_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{\text{tail}}$</td>
<td>(DR = 26 dB)</td>
<td>(DR = 26 dB)</td>
<td>(DR = 26 dB)</td>
</tr>
<tr>
<td>1.05</td>
<td>0.8~15.6</td>
<td>0.3~6.5</td>
<td>0.1~1.95</td>
</tr>
<tr>
<td>1.25</td>
<td>0.8~3.64 (DR = 13 dB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.8~1.33 (DR = 4 dB)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(b). Effect of Vane 3

(c). Effect of Vane 4

(d). Effect of Vane 5 with $\theta_i = 4$
Figure 2.10 shows that a smaller time constant (i.e., \( \tau = 1/w_1 = 1/0.1\omega_0 \) compared with initially \( \tau = 1/0.025\omega_0 \)) results in increased distortions in the spectra. It can be observed that small unwanted peaks appears more noticeably in the AGC coupling-OFF case compared with those in Figure 2.7, and an extra peak appears between \( F_2 \) and \( F_3 \) in the AGC coupling-ON case.

Figure 2.11 shows that, when the number of the neighbouring channel outputs fed into the AGC is reduced from 4 to 2 by setting \( w_{i-2} = w_{i+2} = 0 \), the frequency region at which the LI is less effective is broadened and thus the spectrum within this region (e.g., between \( F_2 \) and \( F_3 \)) is more affected by compression, which results in an extra peak.

Figure 2.12 shows that an increase in the total channel number from 16 to 32 while setting \( w_{i-2} = w_{i+2} \neq 0 \) gives a similar result, that is, an extra peak between \( F_2 \) and \( F_3 \). In both cases, enhancement becomes more local in frequency around the formants.
Figure 2.11: Input spectra (dotted lines) and output spectra (continuous lines) of /u/ when \( w_{i-2} = 0, w_{i-1} = W_L, w_{i+1} = 0, w_{i+2} = W_H, w_i = W_0 \). The AGC coupling ON case corresponds to \( W_L = W_{ij} = 1.4, W_0 = 0.35, K_c = 1 \) (0dB). The AGC coupling OFF case corresponds to \( W_L = W_{ij} = 0, W_0 = 1.1, K_c = 1 \) (0dB).

Figure 2.12: Input spectra (dotted lines) and output spectra (continuous lines) of /u/ with 32 OZGF channels \( (w_{i-2} = w_{i+2} \neq 0) \). The AGC coupling ON case corresponds to \( W_L = W_{ij} = 0.8, W_0 = 0.2, K_c = 2 \) (6dB). The AGC coupling OFF case corresponds to \( W_L = W_{ij} = 0, W_0 = 0.5, K_c = 1 \) (0dB).
The AGC cross-coupling discussed so far is a simple summation operation of the channel outputs. If a Root-Mean-Square (RMS) coupling is used instead, then more extent of spectral enhancement, in theory, is expected to be seen since this operation enhances the LI effect by adding ‘weights’ to strong channel signals. A possible concern with the resulting topology, however, is the increased complexity of hardware implementation. In fact, as shown in Figure 2.13, we may manipulate carefully the weighting factors used in the original summation-coupling scheme to attain similar or even slightly higher spectral contrast compared to the RMS-coupling scheme – a good compromise between quality and complexity is achieved.

Figure 2.13: Input spectra (dotted lines) and output spectra (continuous lines) of /u/ corresponding to two different coupled AGC schemes. The summation-coupling case corresponds to $W_l = W_H = 0.4, W_0 = 0.1, K_c = 1$ (0dB). The RMS-coupling case corresponds to $W_l = W_H = 1.25, W_0 = 0.35, K_c = 2.24$ (7dB).

Figure 2.14 shows that the AGC coupling still works efficiently if the OZGF channels are substituted with a standard cascade of BP biquads. In general, other filters may also be used and theoretically there is no such a limit of filter type for the proposed filterbank.

The AGC input is $\left[ (W_l \times ch_i)^2 + (W_l \times ch_{i-1})^2 + (W_0 \times ch_i)^2 + (W_H \times ch_{i+1})^2 + (W_H \times ch_{i+2})^2 \right]^{1/2}$ where $ch_i$ denotes the full-wave rectified output of the channel $i$. 

* The AGC input is $\left[ (W_l \times ch_i)^2 + (W_l \times ch_{i-1})^2 + (W_0 \times ch_i)^2 + (W_H \times ch_{i+1})^2 + (W_H \times ch_{i+2})^2 \right]^{1/2}$ where $ch_i$ denotes the full-wave rectified output of the channel $i$. 

- 71 -
architecture. The reasons why OZGFs were chosen in this work include its increased biological realism and its suitability of hardware implementation in analog VLSI, as discussed in Section 2.2.1.

2.3.3 Two-tone suppression (2TS)

Figure 2.15 (a) compares the output spectra with the vowel input /i/ with the AGC cross-coupling on \(W_L = W_H = 1.5, W_0 = 0.75, K_c = 1.4\) and the coupling off \(W_L = W_H = 0, W_0 = 0.9, K_c = 1\). The first formant \(F_1\) is at 300Hz, the second formant \(F_2\) is at 2300Hz and the third formant \(F_3\) is at 3000Hz. The parameters were chosen to provide the first two formants with the same spectrum magnitudes in both cases. It can be observed that the AGC coupling improves spectral contrast (especially around \(F_2\)), but also gives less gain to the weakest formant \(F_3\) compared with the compression alone. The gain to \(F_1\) is evidently suppressed probably because this formant is close to the more intense one (i.e., \(F_3\)). This is qualitative similar to the 2TS effect observed in the auditory system. In other words, the proposed architecture that simulates the LI may also
reproduce the 2TS as an emergent property. It is worth noting that the degraded high frequency (HF) emphasis ($F_3$ relative to $F_2$) can be compensated by customizing the AGC parameters for each channel to increase the amplification of high frequencies. Figure 2.15 (b) illustrates a simplified channel customization scheme for this purpose; therein $I_{tail}$ is customized to be unity merely for the 15th channel, and has the same value as initially set ($I_{tail} = 1.25$) for all the other channels. It can be observed that the enhanced $F_3$ is slightly higher than $F_2$ while both formants are well resolvable.

![ZTS Observation](image)

![ZTS Observation with HF Emphasis Compensation](image)

Figure 2.15: Observation of Two-tone Suppression in the spectrum of /i/
Based on the above observations, we performed tone-to-tone suppression experiments to further verify this emergent property: a fixed sinusoid (the suppressor tone) was inputted at 1097Hz* and with the amplitude of 0dB while varying the frequency of a second sinusoid input (the suppressed tone) with the fixed amplitude of -20dB. The output spectrum of the two tones is extracted by performing FFT and plotted to depict the profile of the tone-to-tone suppression with varying the suppressed tone frequency. Note that there are some ripples on the profiles likely due to the resonances of the 16 channel filters.

Figure 2.16 compares the profiles of the tone-to-tone suppression with the OZGF filterbank and the filterbank formed by cascades of the BP biquads when the AGC cross-coupling was active in both cases. It can be observed that the suppressed tone strength decreases in the output when it is close to the suppressor (at 1097Hz) in frequency. In addition, it is interesting to see that compared with the profile of tone-to-tone suppression in the BP-cascade case, the profile in the OZGF case clearly presents an asymmetry between the high frequency and low frequency sides of the suppressor tone; that is, the suppression is more widespread at the high frequency side than that is at the low frequency side of the suppressor. This is likely a nature consequence of the use of the OZGF, whose passband is broader at low frequency side than that at high frequency side. Due to this asymmetry, high frequencies are more affected by low frequencies than low frequencies are affected by high frequencies.

Figure 2.16: The tone-to-tone suppression with the filterbank of OZGF and BP-cascade (N=4). In both cases, \(W_L = W_H = 0.4\), \(W_0 = 0.1\). The suppressor is inputted with a fixed amplitude of 0dB and a frequency of 1097Hz (corresponding to the peak location) while the input amplitude of the suppressed tone is fixed to be 0 dB. The two-tone FFT of the architecture output is plotted as the suppressed-tone frequency varies.

* 1097Hz corresponds to the nature frequency of the 9th channel.
Figure 2.17 and Figure 2.18 show that the extent of the asymmetry in the tone-to-tone suppression profiles can be adjusted via $W_H$ and $W_L$ since the two parameters determine the actual amount of the AGC cross-coupling at high- and low-frequency side of each channel respectively. As illustrated by Figure 2.17, the extent of the asymmetry in the OZGF case is reduced when $W_H$ is larger than $W_L$. In the BP-cascade case (see Figure 2.18), when $W_H$ is larger than $W_L$, the suppression is more effective at the low frequency side of the suppressor tone; on the contrary, the suppression become more effective at the high frequency side when $W_H$ is less than $W_L$. Choosing different values for $W_H$ and $W_L$ respectively is useful for simulating the differential growth of suppression for low- and high-side suppressor as observed in the cochlea [26], although this effort seems to be unnecessary for merely spectral enhancement.

Figure 2.17: The tone-to-tone suppression (in the OZGF-filterbank) as the probe-tone frequency is varied for different values of the weighting factors. In the $W_H = W_L$ case, $W_H = W_L = 0.4$ and $W_0 = 0.1$; In the $W_H > W_L$ case, $W_H = 2.4$, $W_L = 0.4$ and $W_0 = 0.1$.

Figure 2.18: The tone-to-tone suppression (in the BP-cascade filterbank) as the probe-tone frequency is varied for different values of the weighting factors. In the $W_H = W_L$ case, $W_H = W_L = 0.4$ and $W_0 = 0.1$; in the $W_H > W_L$ case, $W_H = 2.4$, $W_L = 0.4$ and $W_0 = 0.1$; in the $W_H < W_L$ case, $W_H = 0.4$, $W_L = 2.4$ and $W_0 = 0.1$.
Figure 2.19 compares the profiles of the tone-to-tone suppression with the AGC coupling ON and OFF in a BP-cascade filterbank. It can be observed that when the AGC cross-coupling is turned off (via $W_0 = 0.1$, $W_H = W_L = 0$), the suppressed tone is efficiently enlarged due to compression and as such its output amplitude is even higher than that of the suppressor tone, resulting in a pattern similar to that in the AGC coupling-on case with a 180 degree rotation from its original orientation. In addition, the effect of compression decreases when a large value of $W_0$ (i.e., $W_0 = 0.5$) is used, which is consistent with that shown in Figure 2.6.

![Graph](image)

Figure 2.19: The tone-to-tone suppression (in the BP-cascade filterbank) as the probe-tone frequency is varied for different values of the weighing factors. The AGC coupling-ON case corresponds to $W_L = W_H = 0.4, W_0 = 0.1$. The AGC coupling-OFF case corresponds to $W_L = W_H = 0$.

Figure 2.20 plots the suppressed-tone strength in the output versus the suppressor amplitude. The suppressed tone was inputted with fixed amplitude of 0.1. It can be observed that, as the suppressor amplitude increases, the output strength of the suppressed tone becomes smaller, and suppression becomes more effective when both tones are closer to each other in frequency, which is consistent with that shown in Figure 2.16-2.19.

The above experiment results show some similarities between the simulated 2TS and that observed in the auditory system; that is, suppression become more effective when the suppressor tone is more intense than the suppressed tone or they are closer to each other in frequency. It is possible to further increase the accuracy of simulating the 2TS since various suppression profiles can be achieved by manipulating the weighting factors and customizing the AGC parameters for each channel.
2.3.4 Benefits for CI Processing

The simulation experiments presented so far combined all the channel outputs to reconstruct the processed signal, by which linear spectral analyses were performed to show the workings and the benefits of the proposed architecture. In the following experiments, the final summation operation was omitted so as to investigate the benefits provided by the new architecture explicitly for the CI processing, where each channel output is dealt with individually for electrode stimulation.

Figure 2.21 shows the output spectral patterns for a synthetic vowel /u/ input: the maximum output of every channel is plotted versus channel number. In the experiment illustrated by Figure 2.21 (a), every channel has the same $Q$ value of 1.6 by setting $I_{tail} = \omega_b/1.6$ and $I_{th} = 1000$ (i.e., using very high $I_{th}$ so that $Q_{max} = [m \tanh (\ln 1000)]^{-1} \approx m^{-1} = Q_{min}$ where $m = 1/1.6$). In this case, the input goes through the filterbank without multichannel compression and spectral enhancement. As shown in the figure, the maximum outputs of the second channel ($\omega_b = 301$Hz) and the eighth channel ($\omega_b = 912$Hz) correspond to the first formant ($F_1 = 300$Hz) and the second formant ($F_2 = 900$Hz).
respectively, although it is more difficult to identify $F_2$ than $F_1$. The third formant corresponds to the thirteenth channel since its frequency ($F_{13} = 2200$Hz) is close to the pole frequency of that channel ($\omega_0 = 2297$Hz), but cannot be resolved in the spectral pattern. The foregoing $Q$-value of 1.6 was chosen to provide the maximum output of the second channel (corresponding to $F_1$) with similar amplitudes in all the cases shown in Figure 2.22 (a)-(c).

Figure 2.21(b) and (c) shows the resulting spectral patterns with AGC cross-coupling OFF and ON respectively. In both cases, the $Q$ ranges were set to be the same as the initial setting (i.e., 0.8~15.6). The values of the weighting factors and $K_c$ were chosen to be the same as those used in the experiment illustrated by Figure 2.7, where $W_H = W_L = 0$, $W_0 = 0.5$, $K_c = 1$ in the coupling-off case and $W_H = W_L = 0.4$, $W_0 = 0.1$, $K_c = 1.12$ in the coupling-on case. Figure 2.21(b) shows that compression alone (corresponding to the coupling-off case) severely degrades local contrast of $F_1$ and $F_2$. The corresponding two channels lie buried in their surroundings and thus the recognition of $F_1$ and $F_2$ become impossible. Figure 2.21(c) shows that when the AGC coupling is present, multi-channel compression provides high-frequency emphasis ($F_2$ and $F_3$ relative to $F_1$) without degrading spectral contrast, which is consist with that shown in Figure 2.7.

![Figure 2.21](image)

Figure 2.21: Bar-charts of maximum output of each channel versus channel number for the vowel input /u/; (a) AGC is disabled by setting the same $Q$ ($Q=1.6$) for every channel with $I_{so} = \omega_0/1.6$ and $I_o = 1000$, no compression and spectral enhancement being performed; (b) The AGC coupling is disabled ($W_H = W_L = 0$, $W_0 = 0.5$, $K_c = 1$), the $Q$ range for every channel being 0.8 to 15.6; (c) The AGC coupling is active ($W_H = W_L = 0.4$, $W_0 = 0.1$, $K_c = 1.12$), the $Q$ range for every channel being 0.8 to 15.6.

Figure 2.22 compare the spectrogram-like plots of an intentionally low-quality rendition of the word “blue” with AGC disabled and $Q = 2.4$ for each channel, with the AGC cross-coupling ON ($W_H = W_L = 0.25$, $W_0 = 0.1$) and with the cross-coupling OFF ($W_H = W_L = 0$, $W_0 = 0.3$). The $Q$ ranges of each channel in the latter two cases were set to be
0.8~15.6. In this figure, a dark black color represents strong intensity. It can be observed that compression alone (corresponding to the coupling-off case) blurs the spectral representation observed in the AGC disabled case with a lot of active channels, whereas the AGC cross-coupling sharpens the spectral features evolving in time by suppressing the surrounding clutter, thus clarifying the burred spectral representation.

Figure 2.23 shows the ability of the proposed architecture to enhance noisy spectra: the input is the vowel /u/ with Gaussian white noise band-limited to 3 kHz. The figure plots the maximum outputs of each channel for different SNRs and compares the results in the AGC-disabled (or unprocessing*) case ($Q = 1.6$) and in the AGC coupling-on case ($W_H = W_L = 0.3$, $W_0 = 0.2$). In both cases, the formants are increasingly lost as the SNR decreases. Note that the $F_j$ cannot be resolved in the AGC-disabled spectral pattern even without input noise. As shown in the figure, the AGC cross-coupling effectively improves the recognition of all the formants (especially $F_2$ and $F_3$) by enhancing their

* The term “unprocessing” herein and in Figure 2.23 means no compression and spectral enhancement.
local contrast. The spectral features represented by these formants are therefore more gracefully lost as the SNR decreases. Even at 15dB SNR, when the eighth channel corresponding to $F_2$ is buried in the surrounding channels, namely, that $F_2$ is totally lost in the AGC-disabled spectral pattern, this formant is preserved to some extent in the coupling-on case. Furthermore, a greater improvement of the $F_2$ recognition at 15dB SNR can be achieved when the ratio of $W_H(W_L)$ to $W_0$ is increased by 0.3/0.1.

Figure 2.23: Maximum outputs of each channel versus channel number for the vowel input /u/ in Gaussian white noise of different SNRs (in 50ms, during which the synthetic vowel was repeated 5 times). The AGC-disabled case, or the un-processing case, corresponds to $Q = 1.6$ for each channel. The AGC coupling-on case corresponds to $W_H = W_L = 0.3$, $W_0 = 0.2$ and $Q = 0.8$ to 15.6 for each channel. The parameters were chosen to have $F_1$ with the same amplitude in both cases. The first six plots show the progressive loss of $F_3$ and the following six plots show the progressive loss of $F_2$. The bottom plot illustrates that the recognition of $F_2$ (local contrast) can be further improved with a larger ratio of $W_H(W_L)$ to $W_0$. 
2.4 Potential Use in Cochlear Implants: A new Speech Processing Strategy of increased Bio-realism

Based on the proposed architecture, we present a new CI pre-processing strategy of increased bio-realism: the speech input is first processed by the simulated cochlear transfer functions (provided by the OZGF), then compressed and spectrally enhanced via the coupled AGC scheme simulating the LI. The channel outputs are finally delivered to electrodes in the form of modulated pulses for stimulation. We present and discuss the advantages of this OZGF-with-LI based strategy as follows:

(a) Increased bio-realism. Wilson et al. [71] have suggested one of future directions of CIs is likely to provide a closer mimicking of processing in the normal cochlea. This is quite understandable since it may reduce the discrepancy between hearing though an implant and through a healthy ear. Our work apparently points in this direction. The OZGF incorporated in the filterbank provides a robust foundation for modelling various auditory data, while the coupled $Q$-adaptive feedback AGC scheme is able to simulate the LI and 2TS as well as level-dependent differences in phase response at different sites on the BM. Our future work can include fitting these bio-realistic features to physiological data from biological cochlea. This can be done by customizing the filter and AGC parameters as well as the weighting factors of AGC coupling for each OZGF channel.

It is worth noting that to best deliver the potential benefits of these bio-realistic features to CI listeners, e.g., level-dependent phased cues, may need to resort to improved electric stimulation. For instance, present CIs convey merely envelope information and fail to deliver phase information, which has proven important for perception of music, tonal language and speech in noise. Wilson et al. [72] have therefore suggested a new place coding of phase information, that is, fine tuning the sites of stimulation along the electrode array according to a fine-structure signal for each channel. The coding by place of stimulation (rather than frequency or rate of stimulation) may be done by selecting a particular electrode among many: recently, Sit et al. [73] proposed a bio-inspired asynchronous interleaved sampling (AIS) algorithm, as opposed to current standard synchronous stimulation strategies (e.g., CIS); with the AIS strategy, high-intensity channels of a filterbank were selected to be stimulated more frequently than low-intensity ones, and thus

---

* We could tailor the form to fit any present stimulation strategies, e.g., CIS.
stimulation pulses are delivered naturally at asynchronous times which have a
definite correlation with phase information within each channel. In the two
perceptual tests, Sit et al. showed the benefits of AIS for melody and speech
recognition in noise. We speculate that through the AIS stimulation, the bio-
realistic level-dependent phase cues simulated in our new strategy, along with
their potential benefits, may be perceived and utilized by CI listeners.

(b) Release of spectral contrast while performing compression. Due to the simulated
LI, compression is performed without degrading spectral contrast, and increased
spectral contrast is therefore available for CI listeners. There is compelling
evidence that the increased spectral contrast is likely to benefit CI listeners:
Loizou et al. showed that compared with normal-hearing listeners, a 4-6 dB larger
spectral contrast is required by CI listeners to identify vowels with relative high
accuracy in quiet listening condition; they also found that some of the CI listeners
could obtain significantly higher scores when spectral contrast is enhanced to
6dB with a spectral enhancement algorithm [5]. Previously, Dorman and Loizou
already showed that it was possible to improve CI users’ consonant intelligibility
by enhancing differences among channel outputs [16].

(c) Robustness to noise. From the perspective of noise reduction, spectral peaks (i.e.,
formants) can have the effects of suppressing the background noise around them
when the LI is simulated. In low SNR environments, more speech information
(represented by spectral features) can therefore be preserved and degrade slowly
when SNR is further decreased (as illustrated by Figure 2.23).

(d) Improved channel selection. The $N$-of-$M$ type stimulation strategies (e.g., ACE
and SPEAK) selects only $N$ channels with the largest outputs amplitudes from
the total $M$ channels of the filterbank ($N<M$) for electrode stimulation. This
maximum selection criterion is sensitive to the spectral distribution of the input
signal. For instance, the selection tends to be biased towards those channels
corresponding to strong or broad spectral peaks (e.g., $F_1$) but neglect other
channels that may contain important spectral content (e.g., $F_2$ and $F_3$). This issue
is likely exacerbated when multi-channel compression is performed, thereby
producing a lot of channel clusters as illustrated by Figure 2.21 and 2.22. Our LI-
based strategies essentially provide a quite natural solution: the spectral pattern
formed by channel outputs is enhanced to have less channel clusters around
spectral prominences and the selected channels can therefore match better with important spectral features. Benefits of such efforts in terms of intelligibility have been reported: Nogueira et al. [74] incorporated a psychoacoustic-masking model in the ACE strategy, in order to avoid the selection of channel clusters due to the simple maximum selection criterion and thus preserve the most meaningful components of any given audio signal; they found a mean improvement of 17% over the traditional ACE strategy in the $N = 4$ case ($M = 20$) but no significant improvement when $N$ was increased by 8.

(e) Likely reduced channel interactions and power consumptions on stimulating electrodes. The LI-based strategy likely produce a ‘sharpened’ current distribution over the simulating sites of the nerve and thus neural response to stimuli at one site is less affected by those at the other sites, counteracting channel interactions. It also likely improves the power efficiency of stimulation since channel clusters are reduced and thus less active but significant channels will be stimulated.

(f) Suitability to low power analog VLSI implementations. The proposed strategy delivers simplicity for hardware implementation, and is amenable to low power analog VLSI implementation. For instance, Katsiamis et al. [56] showed that OZGF could be efficiently synthesized by using low power log-domain biquadratic filters employing MOS transistors operating in their weak-inversion (WI) region. Their work also presented a low power current-mode approach to implement the quasi-logarithmically compressive $Q$-control in AGC. The complete system they developed, including OZGF channels together with the corresponding AGCs, achieved 120+dB of input DR (comparable to that of the human auditory system) and dissipated a mere 4.46μW of static power. Thanks to its low power nature, the new CI pre-processing strategy has potential use in the next-generation CIs that are fully implantable and thus have very stringent power consumption requirements for signal processing.

It is worth noting that the proposed architecture also has potential for use in hearing aids and speech-recognition front ends if the channel outputs are recombined as shown in

---

* It describes masking effects take place in a healthy auditory system and is based on numerous studies of human perception.
Section 2.3.2. Some of the potential benefits discussed so far, e.g., increased spectral contrast and robustness to noise, can be obtained therein via the OZGF-with-LI as well.

2.5 Further Discussions and Conclusions

As discussed in Section 2.3, ensuring the effectiveness of the compression and spectral enhancement simultaneously provided by the OZGF-with-LI system requires a relatively low sensitivity of the AGC input-output transfer characteristic, an adequately wide DR of the $Q$ with an adequately high $Q_{\text{min}}$. The initial setting of the AGC parameters, which has been shown to work effectively, can be further optimized. For instance, the effectiveness can be preserved (see Figure 2.24) when the DR of the $Q$, i.e., 26dB ($Q = 0.8$ to 15.6) is reduced by 15dB ($Q = 1.5$ to 8.3). A moderate DR with moderate values of the $Q$ is desirable as it leads to a better compromise between spectral and temporal resolution for each channel.

While setting the time constant, it is worth noting that a small time constant offers fast-acting compression which can rapidly track the variations in the channel output, but also results in more spectral distortions. The initial setting of time constants was based on the resulting quality of spectral enhancement and can be optimized in the future perceptual tests, where it is possible to evaluate the effects of time constants in terms of listeners’ performance.

Besides the AGC parameters, the weighting factors can be also optimized for practical use since the initial setting is simply intended to have the formants with approximately the same magnitude in different cases (e.g., AGC coupling ON and OFF) and thus the comparison in spectral contrast can be easily done by simply checking the valleys between the formants (peaks). Other efforts may include customizing the AGC parameters and the weighting factors for each channel as discussed at the beginning of Section 2.3.3, so that the quality of high frequency emphasis is preserved in the presence of the 2TS. A possible concern with such efforts is the increased complexity of hardware implementation, and a good trade-off is required therein.
In summary, our investigation on the parametric dependence, as presented in this chapter, shed light to future work on parameter optimization considering as well practical implementation issues in low power analog VLSI. In Chapter 4, we will present the detailed circuit implementation of the proposed OZGF-with-LI system, which involves setting the parameters according to certain circuit specifications. In this work, we have also investigated the effects of several variations in topology, that is, reducing the number of the neighbouring channel outputs fed to the AGC (from 4 to 2) or increasing the total number of channels (from 16 to 32) leads to more local spectral enhancement around the spectral peaks in frequency. We therefore suggest that to preserve the quality of spectral enhancement, more neighbouring channel outputs are required to be fed into the AGC when the channel number is increased. This can actually extend the use of the proposed architecture in future CIs whose spectral resolution is improved due to a largely increased number of channels (e.g. 32 channels or more). Conclusions to the work presented in this chapter are listed as follows:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure224.png}
\caption{Input spectra (dotted lines) and output spectra (continuous lines) of the vowel /u/. The AGC coupling ON case corresponds to $W_L = W_H = 0.4$, $W_0 = 0.1$, $K_e = 1.12$. The AGC coupling OFF case corresponds to $W_L = W_H = 0$, $W_0 = 0.5$, $K_e = 1$. In both cases, $I_{\text{cut}} = \omega / 1.5$ and $I_s = 1.2$ so that the $Q$-range of every channel is 1.5 to 8.3. The resulting DR of $Q$ (i.e., 15dB) is much smaller than the one initially set (i.e., 26dB) while the AGC coupling can still effectively improves the recognition of all the formants.}
\end{figure}
The proposed architecture simulates the LI to enhance compressed spectra, and thus multi-channel compression can be performed without degrading spectral contrast.

The architecture has the ability to enhance noisy spectra by suppressing background noise around the spectral peaks (i.e., formants), thereby providing a potential solution to the CI susceptibility to noise.

The architecture uses the OZGF transfer function for each channel, which provides a robust foundation for modelling cochlear transfer functions, and is able to simulate the LI and the 2TS (as an emergent property) as well as level-dependent differences in phase response along the BM. The novel OZGF-with-LI based strategy, therefore, provides a closer bio-mimicking of the signal processing that occurs in the normal cochlea.

The architecture is amenable to low-power analog VLSI implementations, facilitating its potential use in the next-generation CIs, which are fully implantable.
References


Chapter 3

Speech Recognition Evaluation for the OZGF-with-LI System

3.1 Introduction

This chapter deals with the speech recognition evaluation of the proposed OZGF-with-LI based strategy in terms of its two following design aspects:

(a) The OZGF function for spectral analysis. The purpose herein was to compare intelligibility obtained using the OZGF channel to constitute the filterbank of a cochlear-implant (CI) simulator with that obtained using a standard cascade of bandpass biquadratic filters, in noisy listening conditions. Both filter types have been exploited in recent work towards ultra-low power signal processing for CIs [1-5]. By comparison of the two, we attempted to investigate whether a filter type of increased bio-realism (e.g., OZGF) could contribute to the improvement of speech intelligibility in noise.

(b) The lateral inhibition (LI) mechanism for spectral enhancement. This work was also intended to assess, in terms of speech intelligibility, the effects of the AGC coupling scheme that simulates the LI. For comparison, the AGC coupling-OFF case was used as the reference. In fact, the purpose herein was to investigate whether the side effect of multi-channel compression, i.e., reduced spectral contrast, could be efficiently compensated by adding the bio-realistic spectral enhancement mechanism.

Correspondingly, two perceptual experiments are involved in this work, measuring speech intelligibility of normal-hearing (NH) listeners in a steady-state background noise. Before being presented to the listeners, both speech and noise were processed by a noise-excited envelope vocoder used for the acoustic simulation of CIs in certain aspects, e.g., the limited spectral resolution and the lack of FS information [6, 7]. This chapter has the following organization. Section 3.2 provides an overview of the methods used for the
evaluation. Section 3.3 and Section 3.4 respectively describe the two experiments, present and discuss the evaluation results. Section 3.5 summarizes the present findings and provides a general discussion.

3.2 General Methods

In general, the evaluation methods used in this work involves the following four aspects.

3.2.1 Test Materials

The speech materials used for the tests were derived from a recording of the Coordinate Response Measure (CRM) corpus with eight British talkers [8]. The CRM corpus consists of simple sentences of the form “READY <call sign>, GO TO <color><digit> NOW”, and provides eight call signs (“Arrow”, “Baron”, “Charlie”, “Eagle”, “Hopper”, “Laker”, “Ringo” and “Tiger”), four colors (red, blue, green, white”) and eight digits (1 to 8). It was originally intended to measure speech intelligibility with a noise masker but also shown to be very useful in multi-talker listening tasks due to its call-sign-based structure. In our experiments, listeners were asked to identify the color (one of the four) and digit (one of the eight) combinations contained in the sentences, and only sentences with the call sign “Arrow” were chosen for use so that listeners could naturally ignore the call sign, resulting in relatively little decision-making and working-memory demands during the listening task. The speech-shaped noise was used and constructed by filtering white noise with the long-term average spectrum of the CRM sentences, smoothed using a Hann function (See Figure 3.1). Prior to any processing, the noise and speech were always mixed together as the stimuli, and the noise level was varied to give different signal-to-noise ratios (0, -3, and -6 dB SNR). All the stimuli (before and after any processing) were calibrated to have the same root mean square (RMS) level.

3.2.2 Cochlear Implant Simulator

Acoustic simulation of CI was implemented via a noise-excited envelope vocoder in the following manner. The input acoustic signal was first filtered into 16 frequency bands (channels), using OZGFs or a cascade of bandpass biquads (as the analysis filters). After filtering, the amplitude envelopes were extracted from each band with a half-wave rectification followed by lowpass filtering using a second-order Butterworth filter whose cutoff frequency is at 320 Hz. The extracted envelopes were each used to modulate a broadband white noise, after which the modulated noises were spectrally limited by the same bandpass filters used for filtering the input stimuli (as the synthesis filters). All filtered
noise bands were then added back together to form the acoustic simulation of a sixteen-channel CI. This simulation makes use of temporal envelope cues in speech, but removes FS information [6], which is consistent with the present enveloped-based CI processing strategies. Figure 3.2 summarizes the above process in a flow chart. Note that the AGC with coupling ON or OFF was incorporated in the vocoder for the second experiment (see details later in Section 3.4).

![Figure 3.1: Long-term average spectrum (LTAS) of CRM sentences (top panel) and the corresponding speech shaped noise spectrum (bottom panel) with SNR = 12dB. The LTAS was derived by averaging the superimposed spectra across CRM sentences.](image)
Figure 3.2: An overview of the noise-excited envelope vocoder used to simulate the CI processing.

3.2.3 Listeners

Eight NH volunteers participated in the first experiment (Section 3.3), and fourteen participated in the second one (Section 3.4). No subject participated in more than one experiment. In each experiment, half of the listeners were female, and the other half were male. All listeners were native English speakers, with the majority being undergraduate and postgraduate students from the Imperial College of Science, Technology and Medicine, UK.
3.2.4 Procedure

All the experiments were conducted in an anechoic chamber, and provided listeners with a graphical user interface (GUI) containing 4 colors and 8 digits displayed as buttons on the computer screen (see Figure 3.3). The listeners were asked to click on the button corresponding to the color and digit coordinate contained in the sentence they heard.

![Figure 3.3: The graphical user interface (Matlab-based) for CRM tests.](image)

All the tests comprised the three different SNR ratios of 0, -3 and -6 dB. In the first experiment, the SNR ratios were each presented with the OZGF and with a cascade of bandpass biquads, while in the second experiment each was presented with the AGC coupling ON and with coupling OFF. Considering the CRM sentences were spoken by 8 different talkers, there are in effect a total of 48 conditions for each experiment. These conditions were each repeated five times in a randomized manner for each listener, resulting in a list of 240 items. For each of the condition list, one sentence was selected randomly from the 256 CRM ones (4 colors × 8 digits × 8 talks) and presented to the listener then via headphones (Sennheiser) at a comfortable volume level which was kept constant in all experiments. After all the tests, the listeners’ responses were scored off-line according to the correctness of identifying the color-digit coordinates.
Each experiment provided listeners with practice trials before the actual test in order to acclimatize them to the output sound of cochlear implant simulator. During this phase, sentences were initially presented without noise and then at the SNR of 3 dB, while feedback was provided by displaying the correct GUI buttons. No feedback was given during the test following this practice phase. A formal protocol (showed and explained to listeners) is given in Appendix A.

3.3 Experiment I: OZGF vs. Cascaded Bandpass Biquads

Cochlea transfer function, refer to the frequency response of the basilar membrane (BM), is known to be neither purely lowpass nor purely bandpass, but an asymmetric bandpass function sharply tuned [9-13]. Rhode [10] defined a Bode-plot (see Figure 3.4) to characterize the BM frequency response with three slopes ($S1$, $S2$ and $S3$) as well as two break points (at $\omega_Z$ and $\omega_{CF}$) at which the straight lines cross. Depending on the centre frequency (CF) $\omega_{CF}$ and the input sound intensity, $\omega_Z$ ranges between 0.5~1 octave below $\omega_{CF}$, $S1$ and $S2$ range between 6~12 dB/Oct. and 20~60 dB/Oct. respectively, $S3$ is lower than -100 dB/Oct. and can even be close to -300 dB/Oct., while the excess gain is higher than 17dB [9-13]. From the engineering point of view, $S1$ corresponds to a 1st- or 2nd- order high pass response and $S2$ to at least a 4th- (up to 10th-) order one, while $S3$ to at least a 17th-order low pass response. It seems that we need a bandpass function which is bio-realistically asymmetric and of high order, in order to best grasp most of the cochlea’s frequency-domain behaviour, especially the very steep slope at high-frequency side of center frequency ($S3$). For instance, Katsiamis et al. [14] demonstrated that Rhode’s measured BM response [9] can be approximated by an 8th- order Differentiated All-pole Gammatone Filter (DAPGF) with a $Q$ of 1.44, yielding the slope $S3$ of -100 dB/Oct. and a peak gain of 28 dB. The DAPGF transfer function is almost the same as the OZGF except that its zero is fixed at DC (i.e., $\omega_Z = 0$). Instead, the OZGF provides flexibility to vary the zero’s location on the real axis and thus provides a more accurate frequency response to be fitted to various physiological measurements.

The foregoing ‘sharpness’ in cochlear processing is thought to contribute significantly to the superb frequency selectivity (or spectral resolution) of the human auditory system. On the other hand, people with hearing loss found difficulty in understanding speech in noise as the foregoing ‘sharpness’ diminished due to the cochlear damage, resulting in impaired spectral resolution and reduced spectral contrast after the cochlear filtering [15-21]. Thus, a valuable question will be whether a bio-realistic ‘sharpness’ in bandpass
filtering would in effect increase the speech intelligibility in noise if it is exploited in CIs? We tried to find an instructive answer for this by comparing the intelligibility obtained using the OZGF—which mimics the biological sharpness of filtering—with that obtained using a cascade of standard bandpass biquads. To add value to this study from the practical realization point of view, and to make it extendable to a real-time realization, the filters used for the comparison were adapted from two recent engineering efforts on low power analog VLSI implementation of cochlear processing [3, 5].

![Figure 3.4](image_url)

Figure 3.4: A piece-wise approximation of the BM frequency response defined by Rhode [10] with the parameters reported in various studies [9-13]; adapted from Katsiamis et al. [14].

### 3.3.1 Algorithm and Parameter Setting

Table 3.1 details the filter transfer functions used for this experiment. The order of the OZGF was set to be 4 (i.e., \( N = 4 \)), as in the active 4th-order OZGF channel proposed by Katsiamis et al. [5]: the channel with its AGC mechanism was able to accommodate 120dB+ of input dynamic range (comparable to that of the human auditory system) at 3kHz and dissipated a mere 4.46μW of static power. The OZGF used in this study was in effect an AGC-free version of their work. Note that \( N = 4 \) means the OZGF can be decomposed into a lossy bandpass biquad (i.e., a two-pole one-zero transfer function) and three following identical lowpass biquads – see (2) in Section 2.2.1 and Figure 2.3. In
other words, the 4\textsuperscript{th}-order OZGF channel was essentially an 8\textsuperscript{th}-order cascaded filter structure.

On the other hand, the bandpass (BP) filter used for the comparison was a cascade of two biquadratic filters, or biquads. Such filters had been exploited in recent work towards low-power analog CI processors [1-4], e.g., the one proposed by Sarpeshkar \textit{et al}. [3]: the filters in this processor achieved a 65-dB dynamic range with a power consumption of 5.4\(\mu\)W at 5- to 10-kHz and 112nW at the 100- to 200-Hz; this processor operates with a typical \(Q\) value of 4, and its \(Q\) is programmable up to 10. In our experiment, the stage \(Q\) for each biquad was set to be 2.6, giving an overall \(Q\) of 4 as in that CI processor, and we also used the same stage \(Q\) value for the OZGF. The centre frequencies and the peak gains of the OZGF and of the cascaded biquads were equated as well. A comparison of the resulting OZGF and cascade bandpass biquads in the frequency domain is shown in Figure 3.5. Observe that the OZGF is more sharply tuned than the cascade bandpass biquads, thereby providing better spectral resolution. The slopes \(S2\) and \(S3\) of the OZGF are about 30 dB/Oct. and -50 dB/Oct. respectively while the two parameters of the cascaded biquads are about 24 dB/Oct. and -16 dB/Oct. respectively. A total of sixteen OZGF channels or cascaded biquads were used to constitute the analysis and synthesis filterbank in the vocoder, whose centre frequencies were logarithmically spaced between 97 Hz and 4860 Hz.

It is worth noting that the above two low-power engineering efforts deliberately deviated from the ideal transfer functions we adopted (listed in Table 3-I) in a modest manner so as to solve some practical issues in the analog VLSI implementations of these filters while the performance was not significantly scarified. Fortunately, our \textit{software} implementation did not suffer from these issues, which can be completely neglected in realizing the ideal filter transfer functions.
Table 3-I: Transfer functions used in Experiment I – OZGF vs. Cascaded Bandpass Biquads

<table>
<thead>
<tr>
<th>Filter Type</th>
<th>Transfer Function</th>
</tr>
</thead>
</table>
| OZGF \((N = 4)\)                   | \[
\frac{K(s + \omega_s)}{(s^2 + \frac{\omega_0}{Q} + \omega_0^2)^N}
\] \(N = 4\)  
\(K = \omega_0 2^{N-1}\) for dimensional consistency  
(3)                                      |
| BP Filter-cascade \((N = 2)\)      | \[
\frac{K \cdot s^N}{(s^2 + \frac{\omega_0}{Q} s + \omega_0^2)^N}
\] \(N = 2\)  
\(K = \left(\frac{\omega_0}{Q}\right)^N\) for dimensional consistency  
(4)                                      |

Figure 3.5: A comparison of the OZGF and the cascaded bandpass biquads in frequency response (adapted from [3, 5]).
Figure 3.6: Speech intelligibility results for the three SNRs with the bandpass cascade vs. the ones with the OZGF, scored in percent correct. Bars represent the mean and ±1 standard error across the eight listeners.

3.3.2 Results and Discussion

The results, scored in percent correct and averaged across the eight listeners, are shown with the bandpass cascade and with the OZGF at the three SNRs tested (i.e. 0, -3, and -6 dB SNR) in Figure 3.6. A repeated-measures analysis of variance (ANOVA), with filter-type (the OZGF or the bandpass cascade) and SNR as fixed factors, revealed that merely SNR was statistically significant \([F(2, 14) = 115.89, p<0.001]\), whereas using the OZGF more sharply tuned has no significant effect on intelligibility \([F(1, 7) = 0.78, p=0.406]\).

This finding indicates that the normal-hearing listeners via the acoustic stimulation of CIs cannot benefit from the sharp-tuning in filter shape, at least with present parameter settings. The potential benefits discussed earlier may be counteracted by other effects: for instance, more narrowly tuned channels passed less background noise but meantime less speech information, especially when the total number of channels is small (e.g., 16 channels in this case), and some important information may be therefore missing after such filtering; despite better frequency selectivity, the listeners may suffer from a loss of temporal resolution due to the narrower passbands. It is worth noting that although the sharp tuning of the OZGF has no significant effect on intelligibility according to the
present results, it is desirable from the perspective of spectral enhancement, facilitating the OZGF-with-LI system to resolve spectral peaks and valleys. Further support to this view can be found in our previous simulation of the coupled AGC scheme, which has been presented in the last chapter.

3.4 Experiment II: AGC Coupling ON vs. Coupling OFF

In Chapter 2, we showed that the coupled AGC scheme, which simulates lateral inhibition (LI), can perform multi-channel compression while preserving spectral contrast well; on the other hand, multi-channel compression alone without spectral enhancement (i.e., AGC Coupling OFF) can severely degrade the input spectral contrast. In this experiment, we attempted to verify whether the above advantage of the coupled AGC scheme can in effect improve speech intelligibility in noisy listening conditions – more specifically, whether is it possible to improve speech performance by simulating the LI, which results in increased across-channel intensity contrasts (as depicted by Figure 2.21 and 2.22).

3.4.1 Algorithm and Parameter Setting

The AGCs were added to each analysis filter channel, acting in feedback, and coupled across channels, as described in Figure 2.1. For comparison, the tests were performed with the AGC coupling ON and coupling OFF by setting different values of the weighting factors: the former case corresponded to positive weighting factors (i.e., $W_{HH} = W_{LL} = 0.4$, $W_{0} = 0.1$); in the latter case, the weighting factors for neighbouring channels were set to be zero ($W_{HH} = W_{LL} = 0$). Other parameter settings for both cases include: $I_{slow} = \omega_0/1.5$, $I_{sh} = 1.2$, $I_{0,slow} = 0.95$, $w_0 = 0.025\omega_0$, $\omega_z = 0.1\omega_0$ giving a $Q$-range of 1.5~8.5 as used in Figure 2.24. In both cases, the OZGF analysis and synthesis channels were used, and all the stimuli after processing were adjusted to have the same RMS level. For a constant loudness level, it was likely that speech intelligibility was significantly affected by how well the original spectral features would be preserved after the processing in the two cases. Figure 3.7 compares the two resulting spectrograms of the CRM sentence “Ready Arrow, go to blue eight now”: the spectral features evolving in time were blurred with the AGC coupling OFF but preserved better with AGC coupling ON.

Note that the same AGC scheme was deliberately inhibited for the synthesis filterbank, which means the synthesis filterbank was not exactly matched with the analysis one. We did this because the high level of compression resulting from the matched synthesis and analysis (with two respect AGC schemes operating simultaneously) was found to severely
degrade the speech intelligibility even in quiet listening conditions, during a preliminary perceptual test. A trade-off was achieved by fixing the stage $Q$ value for the synthesis filterbank to be the same as that used in Experiment I, i.e., $Q = 2.6$, a relatively low value for preserving temporal resolution. This value is close to the lower-end of the $Q$-range and hence the regulated $Q$ values of the low-frequency analysis channels containing most important spectral information that determines the speech intelligibility (e.g., the first two formants $F_1$ and $F_2$). On the other hand, spectral resolution (referring to frequency selectivity) was not significantly sacrificed due to the fairly high filter order of the OZGF (i.e., $N = 4$ and hence an $8^{th}$ order filter-cascade structure). Furthermore, the OZGF preserves sharp tuning at the high frequency side of its centre frequency for different $Q$ values: the slope $S_3$ (defined in Figure 3.4) is almost constant with varying $Q$ from 2.6 to 10 for a given filter order $N$ as illustrated by Figure 3.8.

3.4.2 Results

Figure 3.9 shows the scores in percent correct averaged across the fourteen listeners, with AGC coupling OFF and ON at the three different SNR tested (i.e., 0, -3, and -6 dB). It can be seen that the coupled AGC scheme provided a noticeable improvement in performance, and the improvement was 31 percentages points averaged across SNRs. The improvement was confirmed statistically using analysis of variance (ANOVA) with AGC Coupling ON/OFF and SNR as fixed factors: both factors were significant with no interaction [SNR: $F(2, 26) = 69.93$, $p<0.001$; AGC schemes: $F((1, 13) = 275.12$, $p<0.001$].

![Figure 3.7: The spectrograms of the CRM sentence “Ready Arrow, go to blue eight now.” with the AGC coupling OFF and ON.](image)
Figure 3.8: OZGF $\delta 3$ slope versus stage $Q$ for different $N$. From $Q = 2.6$ to 10, the variation in $\delta 3$ does not exceed 2% for each $N$.

Figure 3.9: Speech intelligibility results for the three SNRs with AGC Coupling OFF vs. ON, scored in percent correct. Bars represent the mean and ±1 standard error across the fourteen listeners.
3.5 General Discussion and Summary

The present findings in the two foregoing experiments can be summarized as follows:

(1) The simulated biologically asymmetric ‘sharpness’ of tuning of the filter cannot contribute to an improvement in the speech intelligibility in a steady-state noise background, at least with present parameter settings and with only 16 channels.

(2) The simulated LI can result in a considerable and significant improvement in speech intelligibility in steady-state noise, compared with the case in which multi-channel compression was performed alone, thereby degrading the spectral contrast that is present in the input speech stimuli.

Although no benefits were observed to be present with the bio-inspired OZGF filtering, it is likely that its potential benefits have been severely counteracted by the loss of useful spectral content when using only 16 channels in our implementation. In fact, the very limited channel number combined with the sharp tuning of these channels can result in large spectral gaps within the overall passband. It therefore remains to see whether those potential benefits will emerge when more channels are incorporated in the filterbank.

We opted for a total of 16 channels in the experiments because this is a channel number widely used in current CIs, but advance in technology may lead to a larger number of channels in the future. A preliminary evaluation for such a case is possible by using an OZGF or BP-cascade filterbank which comprises more analysis channels (e.g., 32) and combines all the channel outputs before their being fed to a 16-channel CI simulator. In other words, we may resort to a separation of the pre-filtering and the noise-excited envelope vocoding (see Figure 3.10) so that the former can have more pre-processing channels than the latter while the following vocoder captures those important aspects of current CIs [6, 22, 23] with the commonly used number of channels as well as the filter type, e.g., the Butterworth filter, which exhibits maximum flatness of response within its passband.

Such an implementation using the pre-processed stimuli as inputs to the vocoder can often be seen in the evaluations of various spectral enhancement techniques, e.g., in [24, 25] where the companding-based spectral enhancement was tested with a total of 50 analysis channels. The most apparent advantage of preprocessing-then-vocoding is that we can
tune the parameters and hence performance of either of the two different processing separately. It may also facilitate our further study on some other aspects of the CI processing. For instance, to simulate the spectral-smearing effects caused by channel interactions, we may deliberately introduce a mismatch between the slope of the analysis- and the synthesis-filters within the CI simulation, i.e., increasing the latter’s slope relative to the former as described in [26], while retaining all the parameters of the OZGF or BP-cascade pre-processing. Then we can investigate whether the present ‘sharpness’ at the pre-filtering stage is able to render the noisy spectral features more robust to the simulated effects of channel interactions at the following stage. It should be noted that, however, to add such a pre-processing scheme to current CIs inevitably increase their complexity and hence is often prohibitive, whereas our present experiments resorted to its “pared-down” implementable version— as shown in Figure 3.2—which integrates the OZGF filtering into the present CI processing.

In our second experiment, the results show a noticeable and significant improvement in speech intelligibility when the AGC coupling scheme is enabled to simulate the LI effects. Also, it remains to be seen whether a further improvement may result from increasing the total number of channels in the filterbank, i.e., a better spectral resolution, which can also be realized as depicted in Figure 3.10 (corresponding to the with-LI case). It is worth noting that while conventional speech evaluation experiments involve an un-processing case (no any compression and enhancement) as the reference for comparison, it was not in consideration in our second experiment and the substitute is the AGC coupling OFF case since the primary purpose of this study is to investigate whether it is possible to compensate the side effect of the compression used in current CIs by reproducing one of biological enhancement mechanisms that are present in the human auditory system. In fact, it is not fair to compare the intelligibility of the compressed speech sounds (even if enhanced simultaneously via the simulated LI) with that obtained when listening to the uncompressed speech, since the compressed speech can sound un-natural for NH listeners, which may degrade their performance in the speech recognition tasks. It was therefore intended to avoid the potential confounding issues associated with such a comparison.

* The un-processing case is equivalent to the AGC disabled scheme shown in Section 2.3.4.
The present study is based on an *indirect* evaluation using acoustic stimulation and measuring speech intelligibility of NH listeners rather than of CI users directly. Therefore, there are obvious limits to its applicability to actual CIs, and it seems to be impractical for a CI simulator to fully account for all the differences between acoustic- and electric-hearing. Nevertheless, using acoustic simulation of CIs provide a faster and less costly way to evaluate a new CI processing strategy while capturing many important aspects of the present CI processing [6, 22, 23]. Another advantage of this approach is to release the evaluation task from some of the confounding individual differences (e.g., differences in patients’ surviving neural population) associated with a high variation in the performance of CI users. Since the present study has shown the benefits of this new CI processing strategy, it gives us confidence in testing it on real CI patients in the future.

There are still some extra efforts that can be made in a CI simulator to provide a more comprehensive evaluation on the proposed OZGF-with-LI processing strategy. Besides introducing more aspects to further approach real CIs’ performance as discussed earlier (e.g. channel interactions), the evaluation may involve examining speech intelligibility in some of the most challenging listening situations for CIs – speech is presented with maskers that are *spectrally and/or temporally fluctuating* [27-29], e.g., in multi-talker babble (resulting from concurrent competing talkers), as opposed to our present study in *steady-state* noise. Since listeners often encounter multiple talkers in daily life, it is worthwhile to study whether the present work continue to benefit them in such complex masking listening situations. On the hand, it is likely to achieve a further improvement in listeners’ speech intelligibility after optimizing the present parameter settings (e.g. the AGC time constant) which, to large extent, were based on our previous simulation work shown in

Figure 3.10: $N$-channel pre-processing followed by an $M$-channel noise excited envelope vocoder for the CI simulation. $M$ is matched with one of the channel numbers used in current CIs, e.g., 8 or 16, and $N > M$. 

- 108 -
Chapter 2. A further parameter optimization for the present OZGF-with-LI system will be primarily based on its resulting intelligibility measured in speech recognition tasks.

References


Chapter 4

Ultra-low-power Analog VLSI for the OZGF-with-LI System

4.1 Introduction

In this chapter, we report an ultra-low power analog VLSI implementation of the previously proposed OZGF-with-LI system (thoroughly described in Chapter 2) integrated in the commercially available 0.35µm AMS CMOS technology. Our efforts in VLSI were intended to realize a CI processor prototype which is suitable for use in fully CIs of the future, which have very stringent requirements on the power consumption for signal processing. We adopted an analogue solution rather than its digital counterpart because the former can provide considerable saving in both power consumption and silicon area compared to the latter, as noted in Section 1.5.

The OZGF channels within the filterbank were synthesized with log-domain circuits, which employs MOS transistors operating in their weak-inversion (or subthreshold) region, and arranged in a Class-AB topology. In addition, the biasing of each channel is dynamically adaptable to a coupled measure of its own output signal strength together with the strength of those neighboring channel outputs, which mimics the dynamics of the LI mechanism and thus enhances the differences among channel outputs (i.e., across-channel contrast). The measure is automatically performed within each channel through a low power current-mode AGC circuit evolving from the previously developed computational model shown in Section 2.3. In this chapter, we describe thoroughly how the VLSI counterpart of this model was built in a compact manner.

The above techniques, including Log-domain, Class-AB and Dynamic Biasing via AGC (referring to syllabic companding), serve as low-power wide-dynamic-range solutions of the proposed system. The scope of this chapter covers not merely how to exploit them in the VLSI implementation, but also the essential ideas behind these techniques, which are
presented in a brief and condensed review. In the sections that follow, we begin with a system overview and then go into transistor-level details on each analog building blocks of the system – within the OZGF and within the coupled AGC respectively. In particular, we present a preliminary performance evaluation of the system based on a silicon IC prototype which contains five OZGF channels together with their coupled AGC circuits.

4.2 System Overview

Figure 4.1 depicts the high-level block diagram of our current-mode analog OZGF-with-LI system. Each channel was synthesized using (a) log-domain filtering techniques [1, 2] and using (b) a pseudo-differential Class-AB arrangement [3, 4] (see Figure 4.3); that is, two log-domain Class-A filters of the ‘upper’ and the ‘lower’ branches (denoted by the subscripts \(u\) and \(l\) processes respectively each of a differential input pair and performs subtraction operation at the output \(I_{OUT} = I_{OUT}^u - I_{OUT}^l\) to form a linearly filtered version of the input.

For Class-AB operation, a signal conditioner (e.g., the GMS in Figure 4.3) is employed at the global input stage to split a bi-directional signal (denoted by \(I_{IN}\)) into a pair of complementary, unidirectional, positive ones (denoted by \(I_{IN}^u\) and \(I_{IN}^l\)) respectively \(I_{IN} = I_{IN}^u - I_{IN}^l\), which then go through a current-distribution network (IDN, formed by both PMOS and NMOS cascode current mirrors) so that their mirrored copies are allocated for different channels.

Figure 4.1: Overall architecture of the analog OZGF-with-LI filterbank. For details on each 4th order OZGF channel together with its coupled AGC, see Figure 4.2.
Figure 4.2: Block diagram of a single 4th order OZGF channel with its coupled AGC block (BP – bandpass; LP – lowpass; FWR – full wave rectifier; LPF – low pass filter; the index of coupled channels ranges from i-2 to i+2).

Similar to the IDN, various biasing currents (denoted by $I_{\text{BIAS} 1, 2, 3 ...}$) are supplied throughout the system via a current distribution network. For debugging purpose, each channel contains several follower circuits to allow a select set of analog waveforms to be visible along critical data paths within the system. The resulting visibility is potentially useful as well for the patient fitting of OZGF-with-LI based CIs of the future.

Figure 4.2 gives details on implementing each channel of a 4th order cascade (i.e., four, 2nd order transfer functions) along with the coupled AGC acting in feedback to adapt the filter $Q$ mirroring to each stage, not merely according to its own channel output but also neighboring channels’. It is worth noting that the weighting (via $W_L$, $W_H$ and $W_0$) of the AGC-internal signals proceed after both full-wave rectification (FWR) and low-pass smoothing (LPF) combined together to function as an envelope detector (ED), while in our previous computational modeling and design in Matlab the signals are weighted and coupled just after the FWR and before the LPF as shown in Figure 2.3 and 2.4. This modification, from the perspective of VLSI, is to ensure a compact transistor-level design of the ED and deliver ease of characterizing the LPF circuitry with a simple uncoupled but rectified signal. From the perspective of mathematical description, the original
AGC input-output characteristic (see Figure 2.4) is unaffected by simply interchanging the order of linear low-pass filtering and weighting.

The transistor-level details on all analog building blocks within the system, as well as essential ideas behind the choice of the foregoing low-power wide-dynamic-range design techniques (a) and (b), are presented in the following sections.

4.3 Log-domain Class-AB OZGF Channels

The synthesis of the open-loop OZGF presented in this work was adapted from Katsiamis et al. [5] and re-arranged the order of cascading filter stages in such a way that the cascade starts with a lossy BP biquadratic stage (i.e., two poles and one zero) followed by other LP stages, as opposed to the previous arrangement in [5]. Through this modification, it is feasible to realize a quasi-variable length of the cascade (i.e., a programmable filter order \( N \)) of the OZGF if output tapers are added following each LP stage and multiplexed into a single final channel output. For instance, if \( N = 2 \) (i.e. 2nd-order OZGF) is desired, we can program the multiplexer to pick out the output merely from the first LP stage just following the BP one. In Chapter 5, this idea is elaborated and further developed into a reconfigurable OZGF filterbank, whose filter slope (determined by \( N \)) and quality factor \( Q \) can vary across channels.

Figure 4.3 depicts a detailed implementation of the OZGF channel in log-domain and in a pseudo-differential Class-AB manner. As noted previously, Class-AB operation in log-domain can be based on [3]:

1) The use of a pair of log-domain Class-A filter branches, properly driven by a signal conditioner at the input (e.g., a geometric mean splitter, or GMS), which serve them respectively with two positive input signals within a differential pair \( (I_{IN} = I_{IN}^+ - I_{IN}^-) \).

2) The subtraction of their uni-directional processed outputs (i.e., \( I_{OUT} = I_{OUT}^+ - I_{OUT}^- \)) to ensure a linear relation of the resulting final channel output \( I_{OUT} \) to the original input \( I_{IN} \) which is a bi-directional signal.

It should be emphasized that the above two basis does not fully account for truly Class-AB operation and more generic criterions to meet have been introduced by Frey [4] in the state-space description of such a system. In the following subsections, we therefore review his pioneering works first, which provide insight into the log-domain signal processing
within externally linear Class-AB systems, and then go into details on the transistor-level implementation of an analog OZGF channel, which exploited these low-power wide-DR techniques.

4.3.1 Log-domain Structure and Pseudo-differential Class-AB Design

The concept of filtering in the log-domain was originally proposed by Adam [2] and later generalized in the Frey’s work [1, 6], where he presented a systematic realization of log-domain filter. Specifically, Frey imposed a nonlinear mapping—which is exponential—on the state-space description of a desired linear input-output transfer function and then interpreted the resulting set of nonlinear equations as Kirchhoff’s Current Law (KCL) nodal equations, from which he developed a novel circuit topology class for log-domain filtering, termed Exponential State Space (ESS). The attractive features of the log-domain filters proposed by Frey mainly include:

(1) They make efficient use of nonlinear elements such as transistors and thus are electronically tunable over a wide range by changing their bias facilitated with current sources.
(2) They are not limited to small-signal operation via carefully exploiting their internal nonlinearities and thus can potentially provide the extended DR under low power supply voltages.

(3) They are suitable for high-frequency operation because their structure is simple, utilizing basic transistor-blocks which are very similar to current mirrors (as opposed to operational amplifiers, operational transconductance amplifier, etc.), thereby suffering less from parasitic, and because they operates yielding very low-voltage swings at any of the nodes.

Essentially, Frey’s ESS design methodology provides insight into a generic *instantaneous companding* [7-9] operation of log-domain filters, which in turn accounts for the above advantages. As depicted in Figure 4.4, log-domain filters are explicitly designed to allow a nonlinear operation internally for large signal while maintaining a linear overall response externally by imposing an *I-to-V* logarithmic compressor applied to the input current and correspondingly a *V-to-I* exponential expander to yield the output current. As a consequence, no linearization schemes are required in the log-domain technique as opposed to other traditional linear filter design approaches such as \( g_m C \) employing linearized transconductors; this results in considerable savings in complexity. In addition, very small voltages swings are yielded at any of the nodes since the processing in the logarithmic domain is performed on compressed voltages.

Note that the compressor and the expander could be a single device respectively, e.g., a BJT or a weakly inversed MOST. In fact, the log-domain technique makes efficient use of intrinsic nonlinear (exponential) I-V characteristic of these transistors to produce *externally-linear-internally-nonlinear* (ELIN) [10] frequency shaping networks, which show promise in applications that require wide DR under low power-supply voltages.

Figure 4.4: Log-domain filter’s companding-structure [3, 4].
It is possible to further extend the DR offered by the log-domain technique without introducing too much power, by means of incorporating log-domain filters in a pseudo-differential Class-AB architecture (see Figure 4.3). Class-AB design approach is known to be able to offers a desirable trade-off among linearity, noise, and power consumption as well as DR. It combines the merits of both Class-A and Class-B design: for small signals, it exhibits the quality of Class-A approach, offering high linearity; for large signal swings, it captures the efficiency of Class-B approach (i.e., low quiescent power consumption and standing noise) while maintaining distortion low.

As discussed earlier, an externally linear Class-AB operation in log-domain is based on the use of an input splitter, followed by a pair of log-domain Class-A filters processing respectively two positive signals generated from the splitter, and a final restoration of the input-output linearity via a simple subtraction of the two filter outputs. However, merely relying on such a topology, it is impractical to ensure all the devices in the topology always carry strictly positive currents, which is required for Class-AB operation. For instance, Frey [4] demonstrated an undesirable case in a certain second-order differential lowpass filter, where the capacitor voltage became bi-directional sometimes due to the existence of overshoot. He therefore specified two generic criterions to meet in the design for truly Class-AB operation, respectively relating to the static and dynamic behavior of an externally linear system described in state-space:

(1) First, given any strictly positive static (i.e., DC) values for the input signal pair, there always exists a strictly positive dc-operating point solution for the state-variables (corresponding to currents in log-domain circuits). This means it is always possible that the derivatives of these state-variables become equal to zero (i.e., static behavior) while the variables themselves remain strictly positive.

(2) Second, after a dynamic input (strictly positive and bounded) is applied to the system, the derivative of each state-variable is strictly constrained to be positive whenever the variable itself tends to zero; this ensures that variable can never reach zero and stay strictly positive for all time.

For better understanding of above two criterions, the reader could refer to the ‘state-space representation’ column of Table 4-I regarding an example of Class-AB differential filters (i.e., a differential biquad). Observe all the state-space equations in that column, where $u$ denotes the differential input and the subscripts $u$ and $l$ denote ‘upper’ and
‘lower’ corresponding to two Class-A filter branches respectively; $y_{lp}$ and $y_{bp}$ denote the differential LP and BP outputs respectively; $g$ is a positive constant and could be defined as required in the circuit implementation; $\omega_0$ and $Q$ correspond to the pole frequency and quality factor respectively. It is not difficult to verify that the derivatives of the state-variable $x$ with different subscripts can take on the zero value while these variables and the input signals remain positive (Criterion 1), since all the coefficients of the terms (left-hand/right-hand side) excluding the derivatives are positive.

Note the role of the nonlinear cross-coupling terms $x_{ul_j} x_{tl_j}$ $(j = 1, 2)$ on the left-hand side of the equations; apparently, the last two equations (at the bottom of this column) cannot hold without the nonlinear term $x_{u2} x_{l2}$ given zero values of the derivatives. These nonlinear terms will not appear in the final derived transfer function since they cancel out when forming the outputs differentially, which ensures an overall linear input-output characteristic. On the other hand, the second criterion is guaranteed due to the fact that whenever a state-variable approaches zero, there always exists the terms on the right-hand side corresponding to a different variable that stays positive, e.g., a linear cross-coupling term (i.e., the $l$- subscripted terms in the $u$- equation, or vice versa).

It should be emphasized that the state-space formulation of a Class-AB biquad presented in Table 4-I is not the only choice. The flexibility exists, but the above two conditions must be met simultaneously in the design to ensure a truly Class-AB design. Details on the other columns of Table 4-I are provided in the next section, which describes the transistor-level synthesis of a Class-AB log-domain OZGF channel.

### 4.3.2 Filter Synthesis in Log-domain State-space

The log-domain synthesis of OZGF channels in [5] was based on the idea of the log-domain state-space (LDSS) [11] – a generic set of linear differential equations with time-dependent coefficients and state-variables nonlinearly related to currents internal to the circuit. These equations describes the dynamic behaviour of several interconnected low-level circuit elements each termed a Bernoulli Cell (see Figure 4.5) [12, 13]. The cell consists of an exponential transconductors (normally a BJT or a weakly inversed MOS transistor) and a grounded capacitor connected to the emitter or source terminal. A simple analysis as follows reveals that this element implements a nonlinear differential equation (6) of the well-known Bernoulli form, which can be linearized to yield (8) via a nonlinear substitution (7) of the form. The $n$, $C$ and $V_T$ in these equations denote
respectively the subthreshold slope parameter (typically between 1 and 2), the grounded integrating capacitor and thermal voltage (25mV @ T=300K).

\[ I_D(t) = I_s \exp \left( \frac{V_C(t) - V_C(t)}{nV_T} \right) \]  

(5)

\[ \dot{I}_D(t) - \left[ \frac{\dot{V}_C(t)}{nV_T} + \frac{n(t)}{nCV_T} \right] I_D(t) + \frac{I_D^2(t)}{nCV_T} = 0 \]  

(6)

\[ I_D(t) = \frac{1}{T(t)} \neq 0 \]  

(7)

\[ \dot{T}(t) + \left[ \frac{\dot{V}_C(t)}{nV_T} + \frac{n(t)}{nCV_T} \right] T(t) - \frac{1}{nCV_T} = 0 \]  

(8)

![Figure 4.5: The Bernoulli Cell (CMOS version).](image)

The second column of Table 4-I shows a generic set of LDSS equations described in differential form where the \( u \) and \( l \) (subscripts/superscripts) denote the ‘upper’ and ‘lower’ branch of Class-A filter respectively. The interested reader could refer to [11] for further details on how to derive these equations from an interconnection of Bernoulli Cells. Here we place an emphasis on the use of this set of equations for synthesizing log-domain circuits.
Figure 4.6 provides an overview of the LDSS-based approach and correspondingly Table 4-I details the synthesis of a log-domain Class-AB biquad using this approach [5]. This synthesis approach typically utilizes a Bernoulli-Cell interconnection (cascade) to form the “backbone” of a log-domain filter, dynamics of which corresponds to a set of LDSS equations. It performs a comparison between the LDSS equations and the state-space (SS) or signal-flow-graph (SFG) representation of a prototype system (corresponding to the desired transfer function). Such a comparison aims at extracting the design equations necessary to ensure that the LDSS modified via these equations has the same form as the prototype system, as illustrated in Table 4-I. Consequently, the LDSS after modification should give a log-domain system of the same dynamics as the required original prototype. These necessary design equations can be implemented by means of translinear (TL) circuits [14, 15]. Therefore, the final system is formed by integrating the TL circuits with the Bernoulli backbone, which is denoted by “+” in Figure 4.6.
Table 4-I: Log-domain state-space-based synthesis of pseudo-differential Class-AB biquads [5, 11]

<table>
<thead>
<tr>
<th>Class-AB-compatible state-space representation</th>
<th>Generic log-domain state-space (LDSS)</th>
<th>Necessary design equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u = u_a - u_i$</td>
<td>$I_{IN} = I_{IN^e} - I_{IN^l}$</td>
<td>$I_{OUT_{IN^e}} \propto u_a - w_1$</td>
</tr>
<tr>
<td>$y_{IP} = x_{a1} - x_{a2}$</td>
<td>$I_{OUT_{IP}} = I_{OUT^e} - I_{OUT^l}$</td>
<td>$I_{OUT_{IP^e}} \propto w_{a2} - w_2$</td>
</tr>
<tr>
<td>$y_{1P} = x_{a2} - x_{a2}$</td>
<td>$I_{OUT_{1P}} = I_{OUT^e} - I_{OUT^l}$</td>
<td>$I_{OUT_{1P^e}} \propto w_{a2} - w_2$</td>
</tr>
</tbody>
</table>

$x_{a1} + \left(\frac{\omega_0}{Q} + \omega_0 \frac{x_{a1}}{g}\right)x_{a1} = \omega_0 x_{a2} + \omega_0 u_a$  
$nCV_T \dot{w}_{a1} + u_{a1}w_{a1} = I_{IN^e}$  

$x_{a2} + \omega_0 \frac{x_{a2}}{g}x_{a2} = \omega_0 x_{a3}$  
$nCV_T \dot{w}_{a2} + u_{a2}w_{a2} = w_{a1}$  

$\dot{x}_{a2} + \omega_0 \frac{x_{a2}}{g}x_{a2} = \omega_0 x_{a3}$  
$nCV_T \dot{w}_{a2} + u_{a2}w_{a2} = w_{a1}$  

With defined $w$-state-variables:

$w_{a1} = T_a (I_{IN^e} + I_0 w_{a2}) = T_a (I_{IN^e} + I_{OUT^l})$  

$w_{a2} = T_a (I_{IN^e} + I_0^2 w_{a2}) = T_a (I_{IN^e} + I_{OUT^2})$  

$w_{a3} = T_a T_b (I_{IN^e} + I_0^2 w_{a2}) = T_a T_b (I_{IN^e} + I_{OUT^2}) = T_a w_{a1}$  

$w_{a2} = T_a^2 T_b (I_{IN^e} + I_0^2 w_{a2}) = T_a^2 T_b (I_{IN^e} + I_{OUT^2}) = T_a^2 w_{a1}$  

$w_{a2} = T_a T_b (I_{IN^e} + I_0^2 w_{a2}) = T_a T_b (I_{IN^e} + I_{OUT^2}) = T_a w_{a1}$

Note:
$I_0$ and $I_Q$ are the biasing currents as shown in Figure 4.7, which controls the pole frequency $\omega_0$ and quality factor $Q$; $T$-variables corresponds to the definition in Equation (7); the constant $g$ in the original state-space equations is defined as $I_{IP}$.
Figure 4.7: Log-domain Biquad Synthesis in a pseudo-differential Class-AB topology based on the state-space equations in Table 4-I [5]. Note the downward/upward feeding of output currents ($I_{OUT1,u}$ and $I_{OUT1,l}$) between the ‘upper’ and the ‘lower’ branches in topology, which corresponds to the linear and nonlinear cross-coupling terms involved in the steady-state description of the system. The circuits not shown here for clarity are the PMOS and NMOS cascode current mirrors which serve all biasing current sources and implement the subtraction operation on the above output currents. All device sizes and capacitors in the topology are displayed in Table 4-II in Section 4.45.

The final derived LP and BP transfer functions are (11) and (12) respectively and the corresponding circuit topology is shown in Figure 4.7, where the circles mark the Bernoulli Cells. Each Class-A biquad branch (‘upper’ or ‘lower’) contains a cascade of two Bernoulli Cells (corresponding to the filter order $N = 2$). Having in mind Figure 4.5 and the “Necessary design equations” in Table 4-I, the $u$ currents within these cells are implemented as a combination of the DC biasing currents ($I_Q$ or $I_z$) and the cross-coupled outputs. In addition, the output signals, which are defined as a linear function of the $u$-variables (see the 1” and 2” design equations in Table 4-I) are ‘sensed’ through TL,
e.g., the TL-loop formed by $M_1^u M_2^u M_3^u M_4^u M_9^u M_{10}^u$ for $I_{OUT}^u$, which gives $I_{OUT}^u = I_0 \times u_1$, where $u_1 = T_a (I_{IN}^u + I_{OUT}^u)$; other similar TL loops include the ‘lower’ version of the above loop as well as the ‘upper’ and ‘lower’ loops formed starting from the $M_3$ to $M_6$ or $7, 8$. These TL loops yield the following linear relations [5]:

$$I_{OUT}^{(a,b)} = I_0 w_{[a,b]}$$

where the definition of the $w$-variables refers to Table 4-I. Note that one zero is required to be realized in the OZGF and in [5] the simplest possible way was chosen that a DC current $I_z$ was added to the $u$ current of the Bernoulli Cell that contains the $M_5^{(a,b)}$, as illustrated by the dashed line in Figure 4.7. The correspondingly changed LDSS equations (compared to Table 4-I) are as follows and the resulting BP transfer function is shown in (13) with two poles and one zero.

$$nCV_l \dot{w}_{a2} + (I_0^2 w_{a2} + I) w_{a2} = u_{a1}$$
$$nCV_l \dot{w}_{b2} + (I_0^2 w_{b2} + I) w_{b2} = u_{b2}$$

The presence of the coefficients $\mu$ and $\eta$ causes a deviation of (13) from the ideal 2-pole 1-zero biquad (presented in Section 2.2.1) which has the same denominator as (11) or (12). However, this does not affect significantly the overall OZGF response when the order $N$ is fairly high (e.g., $N = 4$), a typical case for yielding desirable sharp roll-off at the high-frequency side of the OZGF response. This is because such a deviation occurs to merely one stage within a filter-cascade.

In term of (11)-(13), the pole frequency is determined as $\omega_0 = I_0 / nCV_γ$, the zero frequency as $\omega_z = I_z / nCV_γ$, and the quality factor as $Q = I_0 / I_Q$. Note that setting a small value of $I_z$ in $\mu$ and $\eta$ (and hence a low $\omega_z$) can further reduce the effects of the above deviation. In practice, we opted for $\omega_z = 0.1 \omega_0$ in the last two chapters to realize a DC tail of -20dB. In the next section, we describes how to manipulate these bias parameters (i.e., $I_0$, $I_z$ and $I_Q$) to form a complete 16-channel OZGF filterbank.

$$\frac{I_{OUT,P}}{I_{IN}} = \frac{I_{OUT}^P - I_{OUT}^L}{I_{IN}^p - I_{IN}^l} = \frac{(I_0/nCV_γ)^2}{s^2 + \frac{(I_0/nCV_γ)}{(I_0/I_Q)} s + (I_0/nCV_γ)^2}$$

$$\frac{I_{OUT,BP}}{I_{IN}} = \frac{I_{OUT}^b - I_{OUT}^l}{I_{IN}^b - I_{IN}^l} = \frac{(I_0/nCV_γ) s}{s^2 + \frac{(I_0/nCV_γ)}{(I_0/I_Q)} s + (I_0/nCV_γ)^2}$$
\[
\frac{I_{OUT_{P+Z}}}{I_{IN}} = \frac{I_{OUT} - I_{OUT'}}{I_{IN} - I_{IN'}} = \frac{(I_d/\mu CV_r) [(s + (I_z/\mu CV_r))]}{s^2 + \mu \frac{(I_d/\mu CV_r)}{(I_d/I_Q)} s + \eta (I_d/\mu CV_r)^2}
\]

(13)

Where \( \mu = 1 + I_d/I_Q \) and \( \eta = 1 + (I_d/I_0) (I_Q/I_0) \).

4.3.3 OZGF Filterbank

In the previous section, we reviewed the log-domain state-space technique and the filter synthesis using this technique (proposed in [5]). This section shows how we pushed the OZGF design to a higher level, i.e., a complete OZGF filterbank, as shown in Figure 4.1.

First, the topology described in Figure 4.7 was repeated to form a 4th-order biquad cascade (as proposed in Figure 4.2) with a removal of unnecessary transistors for each stage. Having (11) and (13) in mind, these transistors include the \( M_{ii}^*, M_{ii}^t \) unnecessary for a lossy BP stage and the \( M_{8}^*, M_{8}^t \) unnecessary for a LP stage. Then, the resulting cascade as a whole was repeated in parallel, yielding a bank of OZGF channels with a bias-distribution network added to specify the parameters \( I_0 \) and \( I_z \) (and hence the pole frequency and zero frequency) individually for each channel (while \( I_Q \) is controlled via the channel-AGC). The bias distribution was realized by means of groups of current mirrors in which the width of MOS transistors was carefully adjusted so that the values of the distributed biasing currents are spaced equally on a logarithmic axis, resulting in the logarithmically-scaled channels presented in Chapter 2 and Chapter 3.

Figure 4.8 and Figure 4.9 show the simulated frequency responses of a complete OZGF filterbank developed using the above circuit topology (with the capacitance \( C = 20\text{pF} \)), which consist of sixteen 4th-order channels with logarithmically spaced centre frequencies (CFs), or peak frequencies. Figure 4.10 details the corresponding CF distribution – a linear function of \( I_0 \) ranging from 1.25nA to 20nA. As the \( I_0 \) varies across channels, the two different \( Q \)-values were realized by setting \( I_Q = 0.2 I_0 \) for \( Q = 5 \) and \( I_Q = I_0 \) for \( Q = 1 \), and the \( I_z \) was kept to be equal to \( 0.1 I_0 \), yielding a fixed DC tail of -20dB (i.e., the region towards very low frequencies in these figures).
Figure 4.8: Simulated frequency responses ($Q = 5$) of the 16 OZGF channels of which centre (peak) frequencies are equally spaced along a logarithmic axis ranging from 250Hz to 4000Hz.

Figure 4.9: Simulated frequency responses ($Q = 1$) of the 16 OZGF channels of which centre (peak) frequencies (CF) are equally spaced along a logarithmic axis ranging from 205Hz to 3280Hz.
Figure 4.10: Simulated CF distribution across the 16 channels corresponding to the peaks in Figure 4.8 and 4.9 with varying the bias current $I_0$ from 1.25nA to 20nA uniformly on the logarithmic x-axis. Observe that the CFs of $Q = 1$ deviates from the ones of $Q = 5$ by approximately 20%, which corresponds to a shift of the response peak towards lower frequencies, as we have presented in Section 2.2.1. The interested reader could refer to [16] for details regarding how the CF varies with different $Q$ values. In particular, the CF becomes approximately equal to the pole frequency $\omega_0$ (or $f_0$) for high $Q$ values (e.g., $Q = 5$).

It is worth noting that the scaling of the CFs could instead resort to the adjustment of the capacitance $C$ within each channel (although it is not electrically tunable). However, to achieve a desired low-frequency response tends to require large capacitors and thus consume much chip area. Furthermore, fitting a CI to the individual patient requires customizing the electrically tunable parameters for each channel, i.e., channel-by-channel adjustment, in order to maximize the hearing benefits in cases where the patient has different amount of hearing loss at different frequencies. The tuning of the channel CFs may also occur after implantation for compensating the effects of a shallow electrode insertion to the ossified cochlea [17]. These flexibilities can be achieved if we make the basing currents digitally programmable.
The above simulations were performed with open-loop OZGF channels and the $I_Q$ was set manually for different $Q$ values. We describe detailedly in the next section how the $I_Q$ becomes adaptive to the channel outputs through AGC circuits. The presence of the AGC scheme, in addition to log-domain and Class-AB techniques, further extends the available DR of each channel since it adapts the filter’s $Q$ and hence gain for different signal levels: weak signals get significant amplification to be well above noise levels (given that they are originally close to or below the noise floor), whereas strong signals get low or no amplification for acceptable distortion levels.

### 4.4 Coupled Channel AGCs

As presented in Section 2.3, we have developed a computational model of the $Q$-control law implemented via the AGC which is quasi-logarithmically compressive as follows:

$$\frac{\omega_0}{Q} = I_{\text{tail}} \tanh \left[ \ln \left( \frac{\hat{I}_{IN}}{I_{\text{th} \text{control}}} + I_\phi \right) \right]$$

where $I_{\text{tail}}$, $I_{\theta \text{control}}$, and $I_\phi$ are constants with no physical meaning, and $\hat{I}_{IN}$ is the extracted envelope (peak) of the AGC input. With $\omega_0 = I_\phi / nCV_T$ and $Q = I_\phi / I_Q$ in mind, its VLSI-compatible form can be described as follows:

$$I_Q = I_{\text{tail}} \tanh \left[ \ln \left( \frac{I_{CP} + I_{\theta \text{control}}} {I_{\theta \text{control}}} \right) \right]$$

where $I_{\text{tail}}$, $I_{\theta \text{control}}$, and $I_\phi$ are the biasing currents while $I_{\phi} = K I_{\theta \text{control}}$ where $K$ is a factor that functions as the $I_{\phi}$ in (14) and depends on the scaling of $I_{\phi}$ with $I_{\theta \text{control}}$. In the coupled AGC scheme, $I_{CP}$ of the channel $i \text{ (index)}$ is an envelope signal derived from the output of this channel (denoted by $I_{ED,i}$) and its neighbouring ones (denoted by $I_{ED,i+1}$ and $I_{ED,i+2}$).

It has the following form:

$$I_{CP,i} = f(I_{ED,i}, I_{ED,i+1}, I_{ED,i+2}, I_{ED,i+1}, I_{ED,i+2})$$

Where $f(\cdot)$ denotes a linear or nonlinear coupling function and will be defined in Section 4.4.3. Note that the dimensional consistency achieved in (15) is unnecessary for purely computational modelling and design but must be guaranteed for VLSI design. In terms of (15), the factor $K$ together with $I_{\text{tail}}$ determines the maximum $I_Q$ and hence the minimum $Q$ corresponding to $I_{CP} = 0$, given that the DC bias for $I_{CP}$ is already lumped into $I_{\phi}$ (or $K$). On the other hand, $I_{\text{tail}}$ itself determines the minimum $I_Q$ and hence the
maximum $Q$ and $I_{o_{\text{ctrl}}}$ controls the AGC sensitivity (i.e., the rate at which the output varies with the input). These parametric dependencies are consistent with those shown in Section 2.3 (for a quick review, see Figure 2.5). Details on how they are yielded from the AGC circuits are presented in the sections that follow.

The implementation of (15) and (16) through the AGC circuits are depicted at high-level in Figure 4.11. The channel output $I_{\text{OUT},i}$ is first processed through an envelope detector that consists of a quasi-full-wave rectifier operating based on the novel use of a GMS, and a LPF that follows to smooth the output $I_{\text{SUM},i}$ of the GMS. The biasing currents for the two building blocks are $I_{\text{GMS, bias}}$ and $I_{\text{LPF}}$ respectively, and the latter controls the time constant of the LPF. The extracted envelope signal $I_{\text{ED},i}$ and those from the neighbouring channels (i.e., $I_{\text{ED},i\pm 1}$ and $I_{\text{ED},i\pm 2}$) are then combined together yielding $I_{\text{CP},i}$ via the coupling circuit that implements the foregoing function $f(\cdot)$. Simultaneously, the mirrored copies of $I_{\text{ED},i}$ are fed to the neighbouring channels for such processing as well. The coupling circuit offers four sets of the weighting factors to choose (via the switches $S_0$–$S_3$) for the constituent current signals involved in (16), and a DC bias $I_{\text{th}}$ is added to its output $I_{\text{CP},i}$ realize the threshold involved in (15).

To implement the logarithmical- and hyperbolic-functions in (15), $I_{\text{CP},i}$ is converted to a voltage signal based on the exponential $I$-$V$ characteristic of weakly-inversed MOS transistors, and this voltage signal is immediately converted back to current through an operational transconductance amplifier (OTA) (as its weak-inversion operation exhibits a hyperbolic tangent transfer characteristic), yielding the final current output $I_{Q,i}$. The $I$-$to$-$V$ circuits and OTA are integrated in a compact architecture (denoted by a single block in Figure 4.11) where the two biasing currents $I_{o_{\text{ctrl}}}$ and $I_{\text{tail}}$ control the sensitivity and upper-limit of the AGC- transfer characteristic respectively as mentioned earlier.

$I_{Q,i}$ is then scaled with $I_o$ for different channels (and hence different $\omega_0$) to ensure that all the channels in the filterbank operates with the same $Q$- range. It is worth noting that such scaling is not performed directly on the OTA tail-current $I_{\text{tail}}$ as presented in Section 2.3 (i.e., $I_{\text{tail}} = \beta \cdot \omega_0$ where $\beta$ is a scaling factor); instead, it is realized through an independent functioning block applied to the OTA output across channels (not visible in Figure 4.11), scaling $I_{Q,i}$ with a factor determined by the ratio $I_o/I_{\text{tail}}$ (corresponding to $1/\beta$), as instructed by (17) and (18). This modification, from the perspective of VLSI, is to ensure the same biasing condition for all the OTAs (and hence the same performance in principle) within different channel-AGCs. All the scaling functions mentioned in this
chapter were realized by one or more pairs of the MOS transistors with different device sizes within each pair.

\[
Q_{i(\text{max})} = \frac{I_{0i}}{I_{Q_i(\text{min})}} = \frac{I_{0i}}{I_{\text{tail}} \tanh \left[ \ln(K) \right]} \tag{17}
\]

\[
Q_{i(\text{min})} = \frac{I_{0i}}{I_{Q_i(\text{max})}} \approx \frac{I_{0i}}{I_{\text{tail}}} \tag{18}
\]

The scaled \( I_{Q_i} \) is eventually mirrored to yield four copies (denoted by four arrows), which are fed respectively to the four biquad stages that compose the 4th-order OZGF channel, regulating their \( Q \) values simultaneously and adaptively.

The above AGC scheme essentially realizes a dynamic-biasing for the OZGF channel via changing \( I_{Q_i} \), which is otherwise known as syllabic companding – a technique that potentially optimizes the power dissipation and output SNR when it is applied in analogue signal processors [18, 19]. The following sections (Section 4.4.1–4.4.4) will provide transistor-level details on each building block of the AGC scheme and finally show measurements of the compete AGC (Section 4.4.5). Device sizing, the values of the capacitor used in the LPF and all the biasing currents is shown in Table 4-II (see Section 4.4.5).

![Figure 4.11: Block diagram of the coupled AGC circuits within each channel (the subscripts \( i, i+1, i+2, i-1 \) and \( i-2 \) are the channel indexes).](image-url)
4.4.1 Quasi-Full-wave Rectification using Geometric Mean Splitter (GMS)

In the previous sections, we have presented the use of a geometric mean splitter (GMS) as the input conditioner for Class-AB operation. A standard TL-based implementation of the GMS is shown in Figure 4.12, where $I_{\text{IN}}$ is a differential input. Observe that the architecture is balanced with the two TL loops formed by $M_1$–$M_4$ and $M_5$–$M_8$ respectively and governed by (assuming all the transistors therein are matched)

$$I_{\text{IN}} = I_{\text{IN}}^\text{a} - I_{\text{IN}}^\text{l}$$

$$I_{\text{GMS\_bias}}^2 = I_{\text{IN}}^\text{a} \times I_{\text{IN}}^\text{l}$$

(19)

Based on (19), the two generated complementary signals can be expressed as follows:

$$I_{\text{IN}}^{\text{a/l}} = \frac{1}{2}(\sqrt{4I_{\text{GMS\_bias}}^2 + I_{\text{IN}}^2} \pm I_{\text{IN}})$$

(20)

In (20), both $I_{\text{IN}}^\text{a}$ and $I_{\text{IN}}^\text{l}$ are strictly positive because of the presence of the biasing term $4I_{\text{GMS\_bias}}^2$, and each of them is complementary to the other, as illustrated in Figure 4.13. Furthermore, they tend to the DC bias $I_{\text{GMS\_bias}}$ for small input swings, whereas for large input swings ($\gg I_{\text{GMS\_bias}}$), their peak values become virtually equal to that of $|I_{\text{IN}}|$ (i.e., the absolute value of $I_{\text{IN}}$), which means the DR could be arbitrarily large in principle since no bound is imposed on the peak values of these signals. Certainly, a limit for the DR exists in practice due to the fact that for TL, the PMOS transistors must be ensured to always operate in their weak-inversion (WI) regime, which specifies a maximum allowable signal swing at the input. Nevertheless, it is possible to extend the upper limit of the WI region considerably through a careful device sizing, e.g., a $\mu$A-range is obtainable with the chosen device sizes shown in Table 4-II.

According to (20), a summation of $I_{\text{IN}}^\text{a}$ and $I_{\text{IN}}^\text{l}$ gives (see the bottom plot in Figure 4.13)

$$I_{\text{SUM}} = \sqrt{4I_{\text{GMS\_bias}}^2 + I_{\text{IN}}^2}$$

(21)

Thus, $I_{\text{SUM}}$ is a positive signal with a lower limit at $2I_{\text{GMS\_bias}}$ and tend to $|I_{\text{IN}}|$, i.e., an full-wave rectified version of $I_{\text{IN}}$, when $I_{\text{IN}}$ is large enough compared to $I_{\text{GMS\_bias}}$, as illustrated in Figure 4.14 (both $I_{\text{IN}}$ and $I_{\text{SUM}}$ are normalized with respect to $I_{\text{GMS\_bias}}$). This suggests a ‘hidden’ advantage of the GMS: if the two output signals are combined in a common-mode (rather than differential) manner, the resulting topology can simultaneously act as a quasi-full-wave rectifier (since its output resembles a full-wave rectified signal) and a compressor (since it imposes a positive lower-limit, or threshold, on its output). We captured this advantage in our AGC implementation by means of a single-ended circuit.
topology shown in Figure 4.15, which contains only one TL loop formed by $M_1-M_4$ and gives equations similar to (19)-(21) (just replacing $I_{IN}^{(n,t)}$ with $I_{OUT}^{(n,t)}$). Note that a cascode current mirror comprised of $M_2,M_5,M_6,M_7$ was added to 'sense' $I_{OUT}$ and thus facilitate the summation operation (i.e., $I_{OUT}^{*} + I_{OUT}^{t}$).

![Figure 4.12: The balanced GMS (formed by weakly inversed PMOS transistors) used as the global input conditioner.](image)

![Figure 4.13: The indicative GMS output waveforms given by (20) and (21).](image)
Figure 4.14: The GMS output versus input given by (21) (normalized with respect to the DC biasing signal). The dotted line indicates the limit as the DC biasing current $I_{GMS\_bias}$ approaches zero.

Figure 4.15: The GMS (formed by weakly inversed PMOS transistors) used as a quasi-full-wave rectifier in the AGC. The single-ended (rather than balanced) architecture was adopted because it is compatible with the differential OZGF channel output.
It should be clarified that there exist topologies that can exactly implement the foregoing absolute-value function, e.g., precision full-wave rectifiers; such a high precision for this function, however, was considered uncritical to the overall functionality of our AGC, which could be roughly described as a average-level detection for $Q$-control, especially when the AGC characteristic was explicitly designed to be compressive (nonlinear) so as to accommodate a potentially wide channel-output range to the desired narrow $Q$-range. Based on this fact, the modified GMS topology can be regarded as a good compromise between precision and complexity of circuit implementation since its quasi-full-wave rectification function is realized with few non-critical components, merely one TL loop and a current mirror being needed. In practice, higher accuracy is obtainable in terms of (21) by using lower $I_{\text{GMS, bias}}$ but a trade-off is needed as the DC bias also affects the speed of the transistors in the GMS.

4.4.2 The LPF Smoothing

The GMS is followed by a 1st-order LPF to smooth the rectified signal $I_{\text{SUM}}$ and thus extract its envelope (DC component). An important consideration in the design of this LPF is to determine its time-constant value. Our previous computational work presented in Chapter 2 has shown that the proposed AGC scheme requires long time constants to ensure small spectral distortions resulted from multi-channel compression. In addition, some perceptual studies have suggested the negative effect of fast-acting* compression on listeners’ intelligibility measured in speech-recognition tasks [20, 21]. Our VLSI design therefore opted for long time constants and applied such parameter settings as chosen in Section 2.3 and 3.4. Specifically, the time constant (denoted by $\tau$) is scaled with the pole frequency $\omega_0$ of each channel, i.e., $f_c = 1/\tau = (1/40) \cdot f_0$ where $f_c$ is the corner frequency (or cut-off frequency) of the LPF, and $f_0 = 2\pi\omega_0$. Thus, high-frequency channels get shorter time constant to adapt its gain than low-frequency channels.

There is a potential difficulty that may be encountered in an attempt to obtain very large time constants for the low-frequency channels without consuming much area, since large capacitors/resistors tend to be needed in this case. Utilization of the Miller effect may help providing increased effective capacitance, but this may also increase the power consumption due to its heavy use of active components for high gain amplification. For a good compromise between area and power consumption, we opted for a simple log-domain integrator solution proposed in [22], as shown in Figure 4.16. The transistors $M_i$

* The fast action of AGC is realized via short time constants.
and $M_2$ (i.e., the input and output device) are matched, and $M_3$ together with the current source $I_{LPF}$ forms a level-shifter that shifts the gate voltage of $M_4$ (a floating transistor connected in between the two gates) relative to that of $M_4$. Mirrored copies of the LPF output $I_{ED}$ are fed to the neighbouring channel-AGCs for coupling.

All the transistors in Figure 4.16 operate in weak inversion and in saturation except that the floating transistor $M_4$ is allowed to operate in its triode regime. $M_4$ acts as a nonlinear pseudo-resistor which is electronically tunable via the level-shifter [23]. In other words, the $M_4$–$C_{LPF}$ circuit is analogous to an RC circuit of the same topology. The resulting time constant is determined as follows [22]:

$$
\tau = C_{LPF} \cdot \frac{V_T}{I_{LPF} I_{04}/I_{03}}
$$

(22)

where $V_T$ is the thermal voltage, and $I_{03}$ and $I_{04}$ represent the saturation currents of $M_3$ and $M_4$ respectively at zero gate bias, i.e., their specific current parameters.

![Figure 4.16: The 1st-order log-domain LPF for smoothing in the AGC [22.]](image)

The presence of the specific-current terms in (22) suggests that besides manipulating the capacitance and/or the DC bias, long time constants can be obtained through relative sizing of the devices. For smoothing, this is a noticeable advantage over traditional 1st-order log-domain integrators (e.g., that used in [24]) where time constants are not dependent on the specific currents. For instance, assuming all parameters of $M_3$ and $M_4$ are matched except that their aspect ratios are set as in Figure 4.16: $W_3/L_3 = 300/1.5$ and $W_4/L_4 = 8/40$, then we have $I_{04}/I_{03} = 1/1000$ and the resulting time constant $\tau$ becomes
one thousand times that obtained from adjusting the DC bias and the capacitance merely (corresponding to $I_{o4} = I_o$) in traditional log-domain integrators. We will show later in Section 4.4.5 that this compact LPF implementation, with the foregoing device sizing, can offer long time constants on the order of hundreds of milliseconds with $I_{LPF}$ within the 10nA range and a capacitor of 10pF. Our desired time constant value for the lowest-frequency channels ($f_0 = 250$Hz) is 160ms in terms of the foregoing scaling manner, while other channels need shorter time constants. In practice, even longer time constants, on the order of 1s or more, are obtainable as shown in [22].

4.4.3 AGC-coupling network and circuits

The AGC cross-coupling was realized via the network shown in Figure 4.17 where the detected envelope $I_{ED}$ is distributed across channels (with indexes $i$, $i \pm 1$ and $i \pm 2$) and subsequently processed by the coupling circuit to yield $I_{CP}$.

To implement the same manner of coupling as our previous computational design, each channel-AGC should accept the signals from its neighbouring four channels (i.e., the higher-frequency two and lower-frequency two) except the AGCs for the channels of $\omega_0$ close to upper or lower edges of the overall filterbank passband, where the number of coupling signals from even higher or lower frequency channels are less than four, e.g., only three signals for the coupling that occurs within the lowest-CF channel-AGC. Our VLSI implementation, however, ensures all the channel-AGCs get the same number of coupling inputs (i.e., five inputs). Specifically, if the coupling signals from lower or higher frequency neighbouring channels are not available, the channel-AGC gets multiple copies of the coupling signals from other neighbouring channels. For instance, as illustrated in Figure 4.17, the neighbouring signals fed to the channel $i-2$ AGC are merely from its higher-frequency side, i.e., the channel-AGCs of $i-1$ and $i$; that is, “doubled” $I_{ED,i-1}$ and $I_{ED,i}$ together with $I_{ED,i-2}$ constitute the five inputs.

The purpose of the above efforts is to have the same biasing condition applied to the internal coupled signal $I_{CP}$ across channels, which is therefore unaffected by different coupling manners but exclusively determined by the preceding GMS's bias $I_{GMS-bias}$. Note the relation of the DC bias for $I_{CP}$ to the minimum $I_0$, i.e., $I_{Qmin}$ (corresponding to zero AGC-input), in terms of (15) – the above effort actually realizes the same $I_{Qmin}$ (before its scaling with $I_0$) for all the channels. It should be clarified that there exists a number of coupling topologies alternative to that shown in Figure 4.17, which was designed to fit our five-channel fabrication where $i = 3$. 
Figure 4.17: AGC coupling network across channels.
From Neighboring Channel AGCs

Figure 4.18: The coupling circuit for four different sets of weighting within each channel-AGC.
The AGC coupling circuit is depicted in Figure 4.18 (see the previous page). The different device aspect ratios indicated therein represent relative sizing of these devices, which implements the weighting factors (i.e., $W_H$, $W_L$, and $W_0$) in the coupling function as follows (corresponding to (16)):

$$I_{CP,i} = W_H (I_{ED,i+1} + I_{ED,i+2}) + W_L (I_{ED,i+1} + I_{ED,i+2}) + W_0 I_{ED,i}$$

where the subscripts $H$ and $L$ denote the weighting for higher and lower frequency channels respectively.

For simplicity, we set $W_H = W_L$ as in Chapter 2 so that the signals from the neighbouring channel-AGCs can be combined together for weighting, as illustrated by the bold arrows in the figure. The weighted signals are multiplexed into $I_{CP}$ via the four switches $S_0$-$S_3$, implementing four different sets of weighting factors correspondingly (for convenience, we use $S_0$-$S_3$ to represent the four sets of weighting). In particular, $S_0$ corresponds to the AGC-uncoupled case ($W_H = W_L = 0$, $W_0 = 1$ and thus $I_{CP} = I_{ED}$), which was used as the reference for comparison as in Chapter 2. From $S_1$ to $S_3$, less and less ‘weights’ (i.e., decreasing $W_{H,L}/W_0$) are allocated to the neighbouring channel-AGC signals. To ensure the four different sets of weighting provide the same DC bias level for $I_{CP}$, the sum of these factors is maintained to be equal to unity, i.e., $W_H \times 2 + W_L \times 2 + W_0 = 1$. Another DC biasing current $I_{th}$ is added to the coupled signal $I_{CP}$ implementing the term $I_{CP} + I_{th}$ in (15), which is subsequently processed by the integrated I-to-V and OTA. In addition, $I_{th}$ can be negative via current subtraction so that there is more flexibility to manipulate the already existing bias level (determined by $I_{GMS,bias}$) of $I_{CP}$.

### 4.4.4 Integrated I-to-V and OTA

The logarithmic and hyperbolic tangent functions together were implemented using the simple architecture shown in Figure 4.19. The two diode-connected and weakly-inversed transistors $M_1$ and $M_2$ are connected respectively to the OTA’s differential inputs $V_+$ and $V_-$, which gives (assuming the two transistors are matched)

$$V_+ - V_- = nV_T \ln \left( \frac{I_{CP} + I_{th}}{I_{0_{control}}} \right)$$

i.e., a logarithmic I-to-V conversion to facilitate the OTA operation (in weak inversion), which subsequently yields $I_{CP}$ as follows:
By substituting (24) into (25), the resulting $I_Q$ has the same form as (15) except that a factor of $\frac{1}{2}$ is additional, which can be lumped into $I_{0,\text{control}}$ or the foregoing factor $K$ in practice.

We applied a basic OTA rather than linearized one because of the following facts:

1. The required linear range is potentially small for the compressed voltage input from the preceding stages GMS (thresholding) and $I$-to-$V$ (logarithmic).

2. It is possible to get even smaller required linear range using a larger DC bias $I_{0,\text{control}}$ according to (23).

3. The OTA operates in quasi-DC as its input $V_+$ is a highly smoothed envelope signal while $V_-$ is simply a DC bias. Thus, the input-output linearity is not strictly needed to be high.

In other words, the above facts allowed us to opt for the simplest OTA topology yielding tolerable linearity, without needing to resort to a more complicated one, e.g., source-degeneration or a interconnection of several transconductance devices [25].

![Figure 4.19: The integrated $I$-to-$V$ followed by the OTA.](image-url)
4.4.5 AGC Simulation and Measured Results

The previous sections present a very compact circuit-implementation of the coupled AGC scheme. Its device sizing and parameter settings are given in Table 4-II (except the LPF’s sizing already shown in Figure 4.16), where such information is also provided for other building blocks of the OZGF-with-LI system depicted in Figure 4.1, including the global GMS, the IDN and the biquads within the OZGF channel. Explanations of these sizing choices are given later in Section 4.5.1.

Note that the value of $I_0$ therein corresponds to that used for the ninth channel (i.e., $i = 9$ and CF = 1100Hz) in the 16-channel OZGF simulation. This value was also adopted by the reference channel in our fabricated 5-channel chip, details on which will be given in the next section. The other channels of the filterbank take logarithmically scaled copies of this value via a bias-distribution network as presented in Section 4.3.3. Such a manner of scaling across channels were also applied to both $I_Z$ for $I_Z = 0.1I_0$ and $I_{LPF}$ for the desired long time constant $\tau = 40/f_0$ (mentioned in Section 4.4.2) where $f_0$ is the pole frequency in Hz. Similarly, the two reference values used for the respective scaling of $I_Z$ and $\tau$ are listed in Table 4-II, and their corresponding distribution networks are merely copies of that used for $I_0$.

All the channel-AGCs have the same parameter settings as listed in Table 4-II except the foregoing $I_{LPF}$, which are scaled for different channels and specify the $Q$-range in the following way (in terms of (17) and (18)):

$$Q_{\text{min}} \approx \frac{I_0}{I_{\text{tail}}} = 0.83 \quad \text{and} \quad Q_{\text{max}} = \frac{I_0}{\{I_{\text{sat}}\tanh[\ln(K)]\}} = 7 \quad \text{for} \quad S_0-S_3$$

(26)

where $K = [(2I_{\text{GMS, bias (AGC)}}-I_0)/I_{0,\text{control}}]^{0.5}$ and the power of 0.5 arises in (25). Note that the two combinations of $I_{th}$ and $I_{0,\text{control}}$ shown in the table give the same $K$ and thus the same $Q_{\text{max}}$ for both the AGC-uncoupled ($S_0$) and coupled cases ($S_1-S_3$). As a consequence, approximately the same input DR for the two cases has been achieved in our final fabricated system, which will be shown later in Section 4.5.5. More examples of such combinations, which give $K = 1.1$ as in (26), are depicted in Figure 4.25.

In practice, the $Q_{\text{max}}$ in (26) will be smaller than 7 as the AGC input (i.e., the OZGF output $I_{\text{OUT}}$) can never reach zero: it should be above the noise floor and hence detectable. Specifically, in our fabricated system, the lower-limit of the AGC input signal was specified equal to 25nA according to the measured noise floor. With the parameter settings in Table 4-II, the AGC scheme maps this value to a maximum biquad $Q$ of 5.
with $I_{Q_{\text{max}}} = 1.16\text{nA}$, i.e., $Q_{\text{max}} = I_0 / I_{Q_{\text{max}}} = 5$, which corresponds to an OZGF peak gain of 50dB and hence a minimum allowable (detectable) OZGF input signal of 80pA. On the other hand, for a large $I_{\text{OUT}}$ of 400nA, the AGC gives a low $Q$ of 1 with $I_Q = I_0 = 5.8\text{nA}$, and a further increase in $I_{\text{OUT}}$ (e.g., within a $\mu$A-range) eventually results in a saturated low $Q$ of 0.83 ($I_{Q_{\text{max}}} \approx I_{\text{tail}} = 7\text{nA}$).

Figure 4.20 shows the simulated waveforms at the outputs of the AGC building blocks. The AGC input $I_{\text{OUT}}$ is a sinusoidal current signal of 400nA. Observe that $I_{\text{SUM}}$ is a quasi full-wave rectified version of $I_{\text{OUT}}$ (with a small DC bias), and $I_{\text{ED}}$ is the exacted envelope signal; the AGC output $I_Q$ is a quasi-DC signal equivalent to $I_0$ corresponding to $Q = 1$. Figure 4.21 shows the measured parametric LPF response, the corner frequency of which is electronically tunable via $I_{\text{LPF}}$ for a fixed $C_{\text{LPF}}$ of 10pF. The time constant can be as long as 167ms when $I_{\text{LPF}} = 4\text{nA}$, corresponding to the desired one (i.e., $1/\tau = f_0/40$, see Section 4.2.2) for the lowest-frequency channel ($f_0 = 250\text{Hz}$). Figure 4.22-4.25 show the measured DC responses of the coupled AGC (correspond to its quasi-DC output) that represent the $I_{CP}$-$I_Q$ transfer characteristic depicted by (15) for different settings of $I_{\text{lp}}$, $I_{0\text{,control}}$, and $I_{th}$. Table 4-II gives the initial settings for these measurements. Several observations on the plots are as follows, as well as the corresponding explanations in terms of (15):

1. $I_{\text{tail}}$ directly determines the saturating level of the characteristic as illustrated in Figure 4.22. In addition, a linear scaling of $I_{\text{tail}}$ results in a linear scaling of the $I_{CP}$-$I_Q$ characteristic: the data points (dots) measured at the same $I_{CP}$ (on the x-axis) are linearly scaled with $I_{\text{tail}}$. This is consistent with (15) in which $I_{\text{tail}}$ is multiplied by the hyperbolic tangent function and the reason why our previous computational design scaled this parameter with the pole frequency $f_0$ to ensure the same $Q$-range for different channels.

2. Compared to $I_{\text{tail}}$, changes in $I_{0\text{,control}}$ and $I_{th}$ (see Figure 4.23 and 4.24) have much smaller effects on the characteristic within the range with large $I_{CP}$ values (from hundreds of nA to 1$\mu$A); especially for a varying $I_{th}$ the characteristic is almost fixed there, which can be accounted for by the fact that $I_{th}$ is negligible for very large $I_{CP}$ in (15). Their effects, instead, are more noticeable at the $I_{CP}$ range with smaller values. Furthermore, a comparison between the effects of $I_{\text{th}}$ and $I_{0\text{,control}}$ reveal that the latter effect is more distributed over the whole range of $I_{CP}$. 

Figure 4.20 shows the simulated waveforms at the outputs of the AGC building blocks. The AGC input $I_{\text{OUT}}$ is a sinusoidal current signal of 400nA. Observe that $I_{\text{SUM}}$ is a quasi full-wave rectified version of $I_{\text{OUT}}$ (with a small DC bias), and $I_{\text{ED}}$ is the exacted envelope signal; the AGC output $I_Q$ is a quasi-DC signal equivalent to $I_0$ corresponding to $Q = 1$. Figure 4.21 shows the measured parametric LPF response, the corner frequency of which is electronically tunable via $I_{\text{LPF}}$ for a fixed $C_{\text{LPF}}$ of 10pF. The time constant can be as long as 167ms when $I_{\text{LPF}} = 4\text{nA}$, corresponding to the desired one (i.e., $1/\tau = f_0/40$, see Section 4.2.2) for the lowest-frequency channel ($f_0 = 250\text{Hz}$). Figure 4.22-4.25 show the measured DC responses of the coupled AGC (correspond to its quasi-DC output) that represent the $I_{CP}$-$I_Q$ transfer characteristic depicted by (15) for different settings of $I_{\text{lp}}$, $I_{0\text{,control}}$, and $I_{th}$. Table 4-II gives the initial settings for these measurements. Several observations on the plots are as follows, as well as the corresponding explanations in terms of (15):

1. $I_{\text{tail}}$ directly determines the saturating level of the characteristic as illustrated in Figure 4.22. In addition, a linear scaling of $I_{\text{tail}}$ results in a linear scaling of the $I_{CP}$-$I_Q$ characteristic: the data points (dots) measured at the same $I_{CP}$ (on the x-axis) are linearly scaled with $I_{\text{tail}}$. This is consistent with (15) in which $I_{\text{tail}}$ is multiplied by the hyperbolic tangent function and the reason why our previous computational design scaled this parameter with the pole frequency $f_0$ to ensure the same $Q$-range for different channels.

2. Compared to $I_{\text{tail}}$, changes in $I_{0\text{,control}}$ and $I_{th}$ (see Figure 4.23 and 4.24) have much smaller effects on the characteristic within the range with large $I_{CP}$ values (from hundreds of nA to 1$\mu$A); especially for a varying $I_{th}$ the characteristic is almost fixed there, which can be accounted for by the fact that $I_{th}$ is negligible for very large $I_{CP}$ in (15). Their effects, instead, are more noticeable at the $I_{CP}$ range with smaller values. Furthermore, a comparison between the effects of $I_{\text{th}}$ and $I_{0\text{,control}}$ reveal that the latter effect is more distributed over the whole range of $I_{CP}$.
(3) Different combinations of $I_{0\_control}$ and $I_{th}$, if carefully manipulated to ensure the same $K$ in (26), can give a family of the $I_{CP}$ versus $I_Q$ curves that converge to the data point $(I_{CP\_min}, I_{Q\_min})$, as illustrated in Figure 4.25.

(4) All the $I_{CP}$-$I_Q$ transfer curves are quasi-logarithmic as suggested by (15) and have non-zero and strictly positive lower-ends where $I_{CP\_min}$ corresponds to a DC offset of $2I_{GMS\_bias\_AGC}$, which is generated at the GMS and a constituent of the foregoing factor $K$ in (26).

In summary, the above observations demonstrate a quasi-logarithmically compressive and electronically tunable AGC transfer characteristic: its upper limit is exclusively determined by the OTA tail-current (i.e., $I_{tail}$), and its lower limit is determined by a ratio between the two DC biasing currents of the $I$-to-$V$ (i.e., $2I_{GMS\_bias\_AGC}-I_{th}$ and $I_{0\_control}$) as suggested by (26). In particular, the third observation suggests a very useful operation that given $I_{tail}$ and $I_{GMS\_bias\_AGC}$ we can manipulate the combination of the $I_{0\_control}$ and $I_{th}$ values to maintain a fixed $Q$-range while adjusting the AGC sensitivity (i.e., the rate at which the AGC output $I_Q$ grows with the coupled input $I_{CP}$).

Figure 4.20: Simulated waveforms generated at different stages of the uncoupled AGC (via $S_0$) for an input signal of $m = 20$. $I_{CP}$ is not shown in the figure since it is equal to $I_{ED}$ in the uncoupled case.
Table 4-II: Device sizing and parameter setting

<table>
<thead>
<tr>
<th>Topology</th>
<th>1DN</th>
<th>GMS (global)</th>
<th>Biquads (OZGF)</th>
<th>GMS (AGC)</th>
<th>OTA</th>
<th>Coupling Circuit*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(W/L)PMOS</td>
<td>300µm/1.5µm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W/L)NMOS</td>
<td>60µm/8µm</td>
<td>15µm/4µm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>1.8V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{OZGF}$</td>
<td>20pF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{LPF}$</td>
<td>10pF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_0$</td>
<td>5.8nA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_Z$</td>
<td>580pA (i.e., $I_Z = 0.1I_0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{ref}$</td>
<td>7nA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{GMS,\text{bias}}$</td>
<td>20nA (global); 30nA (AGC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{LPF}$</td>
<td>18nA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_0$</td>
<td>33.5nA ($S_0$); 36nA ($S_1, S_2, S_3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{0,\text{control}}$</td>
<td>21nA ($S_0$); 19nA ($S_1, S_2, S_3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The dimensions refer to the minimum device sizes in Figure 4.18.

Figure 4.21: Measured frequency responses of the LPF in the AGC with varying $I_{LPF}$. The corresponding corner frequencies (at -3dB) are 6Hz, 13Hz, 27Hz, 57.5Hz and 100Hz, which are linearly related to the used values of $I_{LPF}$. 
Figure 4.22: Measured parametric $I_{CP}-I_Q$ transfer characteristic with varying $I_{tail}$

Figure 4.23: Measured parametric $I_{CP}-I_Q$ transfer characteristic with varying $I_{th}$
Figure 4.24: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with varying $I_{0,\text{control}}$.

Figure 4.25: Measured parametric $I_{CP}$-$I_Q$ transfer characteristic with a fixed minimum $I_Q$ via varying $I_{0,\text{control}}$ and $I_0$ simultaneously. The combinations $(I_{0,\text{control}}, I_0)$ giving $K = 0.4$ are indicated in the zoomed plot.
4.5 Chip Measurements

Figure 4.26 shows a die photo of the proposed OZGF-with-LI system (five-channel) fabricated in a standard 0.35\textmu m AMS 2P/4M CMOS process. The fabricated 5 channels correspond to the 7\textsuperscript{th} \textasciitilde 11\textsuperscript{th} ones within the 16-channel filterbank presented in Section 4.3.3. Table 4-III summarizes the measured performance of the 4.5mm \times 4.5mm chip, details of which are provided in the following sub-sections.

![Figure 4.26: A die photo of the 5-channel OZGF-with-LI system chip.](image)
Table 4-III: Measured performance of the OZGF-with-LI system

<table>
<thead>
<tr>
<th></th>
<th>50dB with &lt;1% THD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peak Gain</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Bandwidths (CFs)</strong></td>
<td>$Q = 1$ 610, 740, 880, 1050, 1280 Hz</td>
</tr>
<tr>
<td></td>
<td>$Q = 5$ 760, 920, 1100, 1320, 1590 Hz</td>
</tr>
<tr>
<td><strong>Input DR @CF</strong></td>
<td>92dB with &lt;5% THD</td>
</tr>
<tr>
<td><strong>Noise Floor</strong></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>9.6pA @CF = 250Hz</td>
</tr>
<tr>
<td>Max.</td>
<td>35.3pA @CF = 4kHz</td>
</tr>
<tr>
<td><strong>Total on-chip Capacitance</strong></td>
<td>1650pF</td>
</tr>
<tr>
<td><strong>Chip Area</strong></td>
<td>4.5mmx4.5mm (20.25mm$^2$)</td>
</tr>
<tr>
<td><strong>Power Consumption</strong></td>
<td>28µW</td>
</tr>
</tbody>
</table>

4.5.1 Optimized Device Sizing and Layout

Table 4-II shows the optimized device sizing for each building block of the system. As presented earlier, for blocks whose functionality strongly relies on the exponential $V/I$ characteristic of their constituent transistors, their sizes were carefully chosen to ensure that those transistors always operate in weak-inversion (WI) for different current levels. For instance, a large aspect ratio of $W/L = 300/1.5$ was employed to maintain the log-conformity of the transistors in the biquads for a µA-range of their drain currents. Despite the limit to the WI regime (i.e., the available subthreshold range) imposed by a particular CMOS process, our sizing efforts attempted to make full use of this limited voltage range in such a way that it is able to accommodate a high DR of currents from several pica amperes (pAs) up to a µA-range for the biquads.

Besides higher DR (one of our design goals), linearity, or more specifically the THD at the OZGF output, is also intended to be optimized via device sizing. To maintain an external linearity in log-domain filtering, the device sizing should guarantee that each transistor in the biquads acts as a true translinear element — a device whose transconductance is linearly related to the current it carries [15]. Therefore, we adopted a small transistor length $L$ together with a large width $W$ to obtain the foregoing large aspect ratio, and to maximize the linear range of the $\ln(I_{DS})-V_{GS}$ transfer characteristic since the slope of the characteristic corresponds to the transconductance (i.e., $d(I_{DS})/d(V_{GS})$) normalized with respect to the current (i.e., $I_{DS}$). Furthermore, too small $L$ was avoided, considering the likely channel-length modulation effects, and a desirable trade-off was found when $L$ was set equal to approximately four times the feature size (i.e., $L = 1.5\mu m$). It is worth noting that for the basic current mirrors and cascode ones,
their constituent transistors (mostly NMOS) are not constrained to operate in WI and favor larger $L$ for higher output impedance.

From the perspective of current matching, large device areas (i.e., $WL$) have been used to optimize the standard deviation of the drain current mismatch, which is inversely proportional to $\sqrt{WL}$ [26, 27]. For the matching purpose, special efforts were also made during the layout phase, exploiting inter-digitization and common-centroid techniques. One such example is the layout for the Class-AB topology: the purpose therein is to minimize potential mismatches between the two Class-A branches, which likely cause distortions at the differential output; the efforts therein (see Figure 4.27 and 4.28) include interdigitating each ‘upper’ transistor with its corresponding ‘lower’ counterpart to form the matched transistor pairs within every Class-AB biquad, and arranging each of these pair in a two-dimensional common-centroid array with both horizontal and vertical axes of symmetry. Other building blocks employing inter-digitization/common-centroid techniques for their layout include the GMSs (balanced and single-ended) with their matched building TL components, the smoothing LPF with its matched input and output transistors, the OTA with its matched differential input-pair and various current mirrors (simple and cascode) used for distributing signals or biases, as well as the ‘upper’ and ‘lower’ capacitors matched in the Class-AB biquads.

Besides matching pairs of transistors of equal size, the foregoing layout techniques were also employed when relative device sizing with high accuracy was required to implement certain ratios of currents, e.g., the scaling of various biasing currents (i.e., $I_b$, $I_{Z}$ and $I_{LPF}$) and the AGC output current $I_Q$ as well as the weighting (in $S_1$-$S_3$) of coupling current signals. Specifically, to match two transistors of different sizes, they were respectively divided into different numbers of fingers (segments) of equal size while the ratio between the counts of their fingers corresponds to the one desired; these fingers were then interdigitated and arranged in a common centroid array. For instance, the dimension 15µm/4µm shown in Table 4-II was used as the finger size for all the transistors in the coupling circuit and the three different sets of the weighting factors (i.e., $S_1$-$S_3$) were implemented via the ratios between the finger counts (i.e., the aspect ratios indicated in Figure 4.18).
Figure 4.27: The ten pairs of the matched transistors (‘upper’ and ‘lower’) which constitute the Class-AB BP-biquad, corresponding to the schematic shown in Figure 4.7 (the unnecessary $M_{11}$ was removed when implementing the lossy BP-biquad). The I/O terminals, the biasing current sources $I_0$ and $I_Z$, and the AGC output $I_Q$ as well as the two capacitors (denoted by $C$), are respectively connected to the source or the drain terminal of some of these transistors.
Figure 4.28: The common-centroid layout of a PMOS transistor pair. Each transistor (A or B) was divided into eight segments (i.e., fingers). The shaded portions on the illustrative pattern indicate the shared source terminals.
4.5.2 Measurement Setup

The measurement setup is illustrated in Figure 4.29, where the device under test (DUT) is our chip. The major instrument is a Standford Research 1mHz-to-100kHz spectrum analyzer (SR785), the internal source of which provides a variety of test signals (will be presented later). The voltage-mode nature of the SR785 means a $V$-to-$I$ conversion required prior to the GMS function, as well as an $I$-to-$V$ conversion required to follow the OZGF output. In this setup, the $V$-to-$I$ function was implemented through:

1. An audio-frequency transformer generating two balanced differential voltages at its output for a single-ended one from the internal source of the SRT785.

2. Two resistors connected to a virtual DC bias ($V_{bias,in}$), which was set equal to the one of the transformer so that no DC current flows through the resistors in principle.
(3) Two unity-gain current amplifiers respectively via two Op-amps matched on-chip for buffering the output currents \( I_{IN+} \) and \( I_{IN-} \) to the GMS different input \( I_{IN} = I_{IN+} - I_{IN-} \).

In practice, differences among the DC bias of the transformer and the ones of the two on-chip Op-amps may be deliberately introduced to partly overcome the mismatch between the two signal paths. The mismatch could be further minimized using precision foil resistors (±%0.01 tolerance). At the OZGF channel output, the filtered current was measured through standard Op-amp \( I\text{-to-V} \) converters followed by the SR785 (with 1MΩ input resistance).

In particular, our setup provides flexibility to utilize external current sources that are commercially available, e.g., Keithley 6221 precision AC/DC current sources. By this way, more precision current sourcing is possible since the \( V\text{-to-I} \) conversion at the input, which may degrade the precision to some extent, is not required. The penalty, however, is that two external current sources are needed at the same time for the differential input and they must operate in a synchronous manner (via externally triggering) so that the two generated AC current waveforms have an exact 180° phase shift relative to each other.

4.5.3 Frequency Response

As presented earlier, the fabricated five channels were calibrated in such a way that their nominal centre frequencies (CFs) correspond to those of the 7th~11th channels within the simulated sixteen-channel filterbank (see Figure 4.8–4.10), i.e., the CF values shown in Table 4-III. For convenience, we apply the indexes \( i = 7 \sim 11 \) to denote the 5 channels respectively in this section and the ones that follow.

Figure 4.30 shows the tunability of the stage \( Q \) and hence peak gain of the 4th-order OZGF channel with \( i = 9 \) for different input strengths (via adjusting the internal source of the spectrum analyzer). This tunability arises from the action of the AGC that is directly adaptive to the channel output and thus indirectly adaptive to the input. Observe the noisy response far away from the CF (e.g., the low frequency tail); the SNR there is much worse than that within the passband (near the CF) due to severe out-of-passband suppression of the signal (with considerably low gain) relative to the noise (e.g., mains noise). Also note that the filter gain changes nonlinearly with the input due to the following facts: a) the peak gain is inherently nonlinear with \( Q \), which stems from the OZGF transfer function [16]; b) the AGC input-output characteristic is quasi-
logarithmically compressive as shown earlier; c) $Q$ is determined by the ratio between the biasing currents, i.e., $Q = I_0/I_Z$. More details on the gain adaption with the input will be given as input-gain transfer curves in Section 4.5.4.

![OZGF Channel Measured Frequency Response](image)

Figure 4.30: The gain-tunability of the 4th-order OZGF channel ($i = 9$) with the maximum peak gain at around 50dB ($Q = 5$). Each trace corresponds to a different input (and hence output) strength.

Figure 4.31 shows that the low-frequency tail is tunable with the biasing current $I_Z$ as this bias determines the location of the zero relative to the pole via $I_0/I_Z$. For clarity, the responses were measured with large signals and hence low $Q$ values so that the SNR at the out-of-passband tail is not too low. Observe that the CF and gain is almost not affected by the variation of the low-frequency tail despite the foregoing design deviation (introduced via (9)) from an ideal OZGF transfer function; this is consistent with our previous explanation in Section 4.3.2.
Figure 4.31: Low-frequency-tail tunability of the OZGF channel (via varying $I_Z$).

Figure 4.32 compares the adaptive OZGF responses for the open-loop and closed-loop cases: the former case was realized on-chip by adding switches to each channel-AGC output so that the AGC can be tuned off as required, and in this case the different $Q$-values were set manually via $I_Q$; the latter case correspond to the weighting scheme $S_0$, through which the AGC coupling become inactive and multi-channel compression is performed alone. It can be seen that compression broadens the frequency response except when $Q = 5$, compared to the open-loop case. This is due to the asymmetric amplification over frequencies performed by compression: the frequency region with weak spectral content (far away from the CF) gets enhanced relative to the intense one (close to the CF). Note that, however, such an asymmetry disappears when the signal strength sensed by the AGC is as low as the threshold of the AGC characteristic, giving $Q = 5$ and hence the maximum peak gain of $\sim 50$dB via the AGC settings presented in Table 4-II.
It is interesting to see in Figure 4.33 that the compressed responses get sharpened and thus have better frequency selectivity in the AGC coupling-ON case (with $S_1$); this indeed simulates the LI/2TS mechanism, which serves to sharpen the frequency tuning in the auditory system [28, 29], whereas the impairment of these mechanisms, which often accompanies hearing loss, can lead to a broadening of tuning curves [30]. Again, note that all of these nonlinear mechanisms become inactive as the AGC input strength approaches the threshold, giving approximately the same frequency responses (with the maximum peak gain of $\approx 50$ dB) as the one in the open-loop case. Our single-tone (Section 4.5.6) and multi-tone experiments (Section 4.5.7) later will further demonstrate the simulated LI effects on across-channel contrasts in a quantitative manner.
Figure 4.33: A comparison of frequency responses with AGC coupling ON ($S_1$) and OFF ($S_0$).

Figure 4.34 show the measured frequency responses of the fabricated and CF-calibrated five OZGF channels with $S_1$ (i.e., AGC coupling ON) for both $Q = 5$ and $Q = 1$ cases. Despite these five channels with $i = 7 \sim 11$ being on-chip, Figure 4.35 demonstrates that other channels within the sixteen channels, e.g., the two boundary channels shown therein with $i = 1$ and $i = 16$, are obtainable by correspondingly scaling their biasing currents $I_0$ with that for $i = 9$ (i.e., the reference channel) – an extended bias distribution network can help. Observe that the frequency response exhibits quite stable characteristics in all these cases.

Figure 4.36 and Figure 4.37 together show the offsets in response measured across a total of fifteen chips for each fabricated channel, which does not exceed 5dB. The offsets could be minimized electronically in our next-generation OZGF-with-LI system (see the next chapter), where the system parameters are programmable, offset calibration bits being particularly introduced there.
Figure 4.34: Measured frequency responses of the fabricated and CF-calibrated five channels with $S_1$ for $Q = 5$ (upper) and $Q = 1$ (lower).
Figure 4.35: Tunability and gain adaption of the response over the frequency range used for the simulated sixteen channels ($i = 1\sim 16$).

Figure 4.36: Across-chip offsets (15 chips) for $i = 9$ with the two extreme $Q$ values (max. and min.).
Figure 4.37: Across-chip offsets (15 chips) for the other four channels with the two extreme $Q$ values.
4.5.4 Linearity Performance

The maximum allowable total harmonic distortion (THD) for the OZGF-with-LI system was specified at 5% over the whole input-level range; as suggested in [5], this figure was chosen based on the THD values reported in performance specifications (regarding linearity) of various commercial and academic CIs/hearing-aids. It is worth noting that for audio/auditory processors, their linearity performance affects listeners’ perception of sound. Based on this fact, a lower THD value does not necessarily “win” as such sound signals might be less “aesthetically pleasing” to human ears. Indeed, this could be regard as a significant difference between traditional amplifiers (in their classical uses) and sound processors; for the latter, especially those in CI systems, where patients hear the artificial sound created by electrical stimulation, moderate THD values are commonly tolerable.

Since in [5] various open-loop tests have been done for the linearity of the OZGF channel, including measuring the intermodulation distortion (IMD), our efforts focused on how the measured output THD varies when the filter’s peak gain changes adaptively with the input level because of the AGC, and varies for the four different sets of weighting ($S_0$–$S_3$). Attention was also paid to different channels – whether the THD is also affected by different biasing conditions (for different CFs) was of interest. The measured results are shown in Figure 4.38 and 4.39. The peak gain decreases with an increase in the input level for different channels ($i = 1$ and $i = 9$, see Figure 4.38) and for different weighting schemes ($S_0$–$S_3$, see Figure 4.39); meanwhile, the corresponding THD at the CF gradually increases over the whole input-level range.

Observe that the peak gain versus input in the AGC uncoupled case ($S_0$) is the same for $i = 1$ and $i = 9$ (via tuning $I_0$) because of the scaling of $I_Q$ with $I_0$, as presented in Section 4.4. On the other hand, the difference between the THD values in the two cases starts to become noticeable when the increased input level goes into the µA-range where the $i = 1$ case gives a output THD of 4.8% for a 3µA input tone while the THD for $i = 9$ is lower at 3.5%. For $i = 16$, the measured THD values (not plotted here) are even lower for the µA-range input where the THD is at 1.7%, 2.5% and 2.8% for an input level at 1µA, 2µA and 3µA respectively. While the upper limit of the input level is around 3µA for $i = 1$ as the resulting THD approaches 5%, the limit can be higher till 5µA for $i = 16$, which gives a THD of 4.75%.

The THD difference across channels could be due to different biasing conditions: with small biasing currents for the low CFs, some transistors in the log-domain biquads may
deviate downward (due to small $V_{GS}$) from the optimum linear range of their $\ln(I_{DS})-V_{GS}$ transfer characteristic, resulting in higher THD at the biquad output; furthermore, this case was likely to be worsened by the use of large $W/L$ to extend the upper limit of the WI range, which could give a further decreased $V_{GS}$ for a given small $I_{DS}$. For a certain biasing current range for the desired CFs, it seems difficult to optimize the THD at both higher- and lower- end of the WI range by adjusting the transistor sizes merely. However, if the capacitors in the biquads are also allowed to be adjusted, which is at the expense of some area, we could have flexibility to avoid that the biasing currents (and hence $V_{GS}$) for the low CFs are too small relative to the ones for the high CFs. This idea will be developed in our future work.

Also observe that for all the weighting sets ($S_0$–$S_3$), the OZGF channel ($i = 9$) provides approximately the same peak gain of $\sim 50$dB at 80pA and the same one of $\sim -3$dB at 3µA. This was achieved by the careful manipulation of the AGC parameters (especially the combination of $I_{0\_control}$ and $I_{th}$), which yields the same $Q_{max}$ and $Q_{min}$ respectively for $S_0$–$S_3$ as presented in Section 4.4.5. The resulting THD values for $S_0$–$S_3$ are close to each other; the difference does not exceed 0.5% over the whole input-level range.

![Figure 4.38: Plots of measured adaptive peak gain (at CF) and the corresponding THD vs. channel input for different channel indexes ($i = 1$ and $i = 9$).](image-url)
4.5.5 Input Dynamic Range (DR)

In this work, the definition of input DR was associated with 1) a given maximum allowable output THD, which corresponds to the upper limit of the input level, and 2) the input noise floor, which sets the lower limit – in practice, the minimum input level should be higher than this figure to some extent for good signal integrity. The measured DR takes into consideration the AGC’s contribution: it was calculated as the ratio of the maximum input signal yielding ~5% THD with $Q_{\text{min}} \approx 0.85$ over the minimum input with $Q_{\text{max}} = 5$.

As presented in the last section, the worst and best THD case were found to happen at the lowest and highest CF channels (i.e., $i = 1$ and $i = 16$) respectively where the corresponding maximum input signals for <5% THD were at 3µA and 5µA. The measured noise floors (RMS values) of the two ‘boundary’ channels together with other channels are shown in Table 4-IV and 4-V. Note that these figures were measured for an
input tone of 80pA, or the minimum allowable input signal for the OZGF-with-LI system, which gives $Q_{\text{max}} = 5$ via the calibrated AGC (see Section 4.4.5). In other words, the minimum input signal level was specified, as mentioned earlier, to exceed the noise floor even in the worst case where $i = 16$ and $80/\sqrt{2}$ (RMS) $> 35.3$ in pA. Also note the different noise figures across channels; these differences can be accounted for by different biasing conditions (which also account for the THD difference) – the small DC biasing currents for the low-CF channels give better noise performance to these channels (but impair their linearity performance) compared with high-CF channels. On the other hand, the measured noise floor is approximately the same for the four different weighting schemes as well as the open-loop case where the $Q$ value was set manually equal to the ones via the AGC (i.e., $Q = 5$). This finding suggests that the noise contribution of the AGC is negligible for the overall system noise; this is likely due to the fact that the AGC gives a quasi-DC output, the frequency of which is far away from the OZGF passband.

In summary, the input DR of the overall OZGF-with-LI system for $S_0$–$S_3$ is 92dB with <5% THD, although the DR difference among channels exists (< 6dB). This DR figure has taken into consideration the worst THD and the worst noise floor cases which were found to happen at the lowest ($i = 1$) and highest ($i = 16$) CF channels respectively. Specifically, the DR reported herein was derived by means of dividing the upper limit of the input level for $i = 1$ (3µA, the same for $S_0$–$S_3$) by the specified lower limit of 80pA.

<table>
<thead>
<tr>
<th>Channel Index $i$</th>
<th>CF (Hz) when $Q = 5$</th>
<th>Noise Floor (pA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
<td>9.6</td>
</tr>
<tr>
<td>7</td>
<td>760</td>
<td>14.3</td>
</tr>
<tr>
<td>8</td>
<td>920</td>
<td>16.4</td>
</tr>
<tr>
<td>9</td>
<td>1100</td>
<td>17.8</td>
</tr>
<tr>
<td>10</td>
<td>1320</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>1590</td>
<td>21</td>
</tr>
<tr>
<td>16</td>
<td>4000</td>
<td>35.3</td>
</tr>
</tbody>
</table>

Table 4-IV: Measured noise floor for the fabricated five channels.

<table>
<thead>
<tr>
<th>AGC Scheme</th>
<th>Noise Floor (pA) for $i = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_0$</td>
<td>17.8</td>
</tr>
<tr>
<td>$S_1$</td>
<td>17.62</td>
</tr>
<tr>
<td>$S_2$</td>
<td>17.7</td>
</tr>
<tr>
<td>$S_3$</td>
<td>17.9</td>
</tr>
<tr>
<td>Open-loop</td>
<td>17.58</td>
</tr>
</tbody>
</table>

Table 4-V: Measured noise floor for the four weighting schemes and the open-loop case.
The tests presented in Section 4.5.3 have shown the ability of the simulated LI effect to sharpen the OZGF frequency response. We now look into how this effect is distributed across channels through a single-tone test. This test was performed concurrently with the previously presented THD measurement for $i = 9$ (shown in Figure 4.39); that is, we measured at the neighbouring channels’ outputs (i.e., $i = 7, 8, 10$ and $11$) the amplitudes of the copies of the test tone used in the THD measurement. The results are shown in Figure 4.40 (a)-(d) where the measured output amplitudes are normalized with respect to that of $i = 9$ and plotted versus input amplitudes. Observations on these figures are as follows:

1. The output amplitudes of the neighbouring channels are suppressed relative to that of $i = 9$ (being negative in dB) because of the frequency selectivity of each channel.

2. The suppression happens at $i = 8, 10$ is less than at $i = 7, 11$ as the passbands of the former are closer to that of $i = 9$.

3. The suppression is stronger via the AGC coupling ON ($S_1 - S_3$) than the coupling OFF ($S_0$).

4. The suppression becomes successively stronger from $S_1$ to $S_3$, i.e., when more and more ‘weights’ are given to the neighbouring channel signals in the coupling function (23).

5. The extent of suppression decreases as the input amplitude is increased since the channels’ frequency selectivity degrades with decreasing $Q$ via the compressive action of the AGC. As a consequence, the four schemes ($S_0 - S_3$) provide very similar extents of suppression for a large input signal (e.g., $1\mu\text{A}$). This is also partly due to the fact that these channels’ AGCs operate almost in saturation when the input signal level is high.

The above observations (3) and (4) reveal that the LI effect simulated via the AGC coupling can enhance across-channel contrasts and thus provide a more channel-specific spectral response pattern for a certain input signal. This is further supported by the results of our complex-tone test presented in the next section.
Figure 4.40: Single-tone output amplitudes vs. input ones of neighbouring channels ($i = 7, 8, 10, 11$) normalized with respect to Channel $i = 9$.
4.5.7 Complex-tone Test

In the complex-tone test, the system input changed to be connected to two Keithley 6221 precision AC/DC current sources. The two devices were connected to MATLAB via their GPIB interfaces, programmed and externally triggered to generate two AC current waveforms (complex tone) which have an exact 180° phase shift relative to each other for the differential input of the global GMS.

The out-of-phase complex tones synthesized in MATLAB are depicted in Figure 4.41 where the two tones are represented by the continuous and dotted lines respectively; as illustrated by its amplitude spectrum, the complex tone was synthesized from five pure tones. For convenience, we use Tone 1~5 to denote the five tones respectively and Table 4-VI gives their frequencies and normalized amplitudes (within the -1~1 scale). Note that the complex tone’s amplitude has been normalized after a superposition of the five pure tones (see Figure 4.42), which accounts for the difference (dB) between the respective FFT amplitudes of its constituent tones and the amplitudes (before the superposition) given by Table 4-VI. The five tone frequencies were set in such a way that they were close to the CF of $i = 7$~$11$ ($Q = 5$) respectively and give a fundamental frequency of 60Hz (and hence a period of 16.7ms) of the complex tone. When programming the current sources, practical amplitudes of current signals were assigned equally to the complex tone and its out-of-phase version.
Figure 4.43 plots the measured output RMS levels across channels (normalized with respect to the maximum one among the five) and for different input RMS levels. Observe that the AGC coupling schemes ($S_1$–$S_3$) enhance the difference among the channel outputs (i.e., across-channel contrast) relative to the uncoupled case ($S_0$), which simulates the LI effect. More specifically, the two stronger constituent tones (Tone 2 and 4, corresponding to $i = 8$ and 10) have the effect of suppressing the other weaker tones (Tone 1, 3 and 5, corresponding to $i = 7, 9$ and 11, i.e., the neighbouring channels of $i = 8$ and 10). In addition, the enhancement becomes successively stronger from $S_3$ to $S_1$ and weaker as the input level increases. All these observations are consistent with the findings from the sing-tone test. Figure 4.44 provides a clearer description of how the extent of the enhancement, or suppression, varies with the input RMS level for different channels, and Figure 4.45 summarizes the increase in the across-channel contrast ($i = 7, 9$ and 11 relative to $i = 8$ and 10) averaged over the input RMS levels. It can be seen that the maximum average increase (20dB) happens with $S_1$ at $i = 7$ while the worst case (8dB) correspond to $S_3$ at $i = 11$; the increase is slightly smaller at the higher CF channel ($i = 11$) due to the passband asymmetry of the OZGF – the frequency selectivity is lower at low frequency side of the CF, and thus this channel gets less suppression from that of $i = 8$ and 10.

Figure 4.41: The complex-tone input and its FFT spectrum.

Table 4-VI: The five pure tones for synthesis of a complex tone.

<table>
<thead>
<tr>
<th>Tone Index</th>
<th>Normalized Amplitude</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.316 (-10 dB)</td>
<td>720</td>
</tr>
<tr>
<td>2</td>
<td>1 (0 dB)</td>
<td>900</td>
</tr>
<tr>
<td>3</td>
<td>0.316 (-10 dB)</td>
<td>1080</td>
</tr>
<tr>
<td>4</td>
<td>1 (0 dB)</td>
<td>1320</td>
</tr>
<tr>
<td>5</td>
<td>0.316 (-10 dB)</td>
<td>1560</td>
</tr>
</tbody>
</table>
Figure 4.42: Formation of a complex tone (Tone A), represented by the continuous line, and its complementary counterpart with 180° phase difference (Tone B), represented by the dotted line, from the five sinusoids whose frequencies and amplitudes are given in Table 4-VI. The amplitudes of Tone A and Tone B have been normalized to be within -1~1.
Figure 4.43: Output RMS levels (window length = 500ms) vs. channel normalized with respect to the maximum.
Figure 4.44: Normalized output RMS levels vs. input for different channels.

Figure 4.45: Increase in the across-channel contrast ($S_i - S_0$ relative to $S_0$) averaged over the input RMS levels.
4.6 Summary and Conclusion

This work presents an ultra-low-power analog VLSI implementation for the proposed OZGF-with-LI system. Owning to the use of CMOS transistors operating in weak inversion and current-mode techniques, its five-channel silicon IC prototype consumes merely 28µW, and most likely this figure will be 3~4 times\(^*\) higher for a sixteen-channel\(^†\) version of our design and thus comparable to those reported in some pioneering works [31, 32] towards the next generation CIs that are fully implantable inside the human body—our system should therefore have potential for use in such applications, where power consumption requirements are very stringent for signal processing.

Probably the most attractive aspect of this work lies in the VLSI realization of lateral inhibition (LI) via low power and compact coupled AGC circuits. The results measured from the system prototype show that the design has the ability to reproduce this biological link between the responses at different frequencies, which enhances spectral features of sound and may improve hearing performance in noisy listening environments but is missing in current CI systems. The reproduced LI effect is represented by the sharpened OZGF frequency response and increased differences among channel outputs (i.e., across-channel contrast). Besides the LI, our VLSI implementation also ‘inherits’ well other biomimetic features from the computational model presented in Chapter 2, e.g., the passband asymmetry and level-dependent filter gain and shape. In summary, the OZGF-with-LI system can provide a closer bio-mimicking of signal processing that occurs in the normal cochlea, which has been suggest as one of future directions of CIs [33].

Our future work can include incorporating more OZGF channels into the present IC prototype (e.g., a total of sixteen channels as simulated previously) and eventually developing the prototype to form a complete OZGF-with-LI based CI processor by adding its other essential components, e.g., a microphone pre-amplifier preceding the global GMS, electrical stimulation circuits following the OZGF filterbank, and current-mode digital-to-analog converters for programmability. An overview of this processor will be given in the next chapter. In addition, more complex and ‘realistic’ tests using natural speech embedded in noise become possible via this processor. The choice of single-tone and complex-tone signals in the present tests was based on the very limited

\* The simulated power consumption was 106µW.
\† This channel number has been widely used in current CIs.
count of channels and hence poor spectral resolution of the prototype filterbank; despite this, the possibility of VLSI realization of the LI effect has been verified through these preliminary tests. In conclusion, we believe the way of simulating LI, which is presented in this work, provides a good compromise between bio-fidelity, silicon area, and power consumption, since it only needs one special AGC building block to implement the coupling function and thus does not significantly increase the overall system’s area and power consumption. In addition, as discussed in Section 2.5, parameter optimization and customization for each channel could help if even higher bio-fidelity is required.
References


Chapter 5

OZGF-with-LI: Some Likely Next Steps

5.1 Introduction

This chapter describes some likely possibilities for improvements in the proposed OZGF-with-LI system, which can be included in our future work. These possibilities may push the current design closer to a practical patient-specific CI system, where programmability is emphasized to deal with a potential high variation in performance across patients, one of the problems remains to be solved for current CIs. In another aspect, through the programmability, further increased bio-fidelity of the signal processing is obtainable, which may support better CI performance. The programmability is realized by digital circuitry while the signal processing is performed via ultra-low-power analog solutions – the possibilities presented in this chapter attempt to combine the best of both analog and digital worlds and eventually yield a hybrid system which is programmable, robust, power-efficient, and of high bio-fidelity. This chapter is organized as follows. In Section 5.2, a programmable OZGF-with-LI based CI processor is presented. Section 5.3 and 5.4 then describe possible improvements to the OZGF channel and the AGC coupling scheme respectively. Section 5.5 gives concluding remarks on these possibilities.

5.2 Programmable OZGF-with-LI based Processor

Figure 5.1 shows a possible OZGF-with-LI based CI processor which is compatible with the present envelope-based stimulation strategies (see Section 1.4). Compared with the analog filterbank described in Figure 4.1, a significant difference lies in the added current-mode digital-to-analog (DAC) converters for altering the various parameters, which make the system programmable. For instance, the biasing currents for the OZGF and the AGC biasing currents can be programmable via several bits for each channel. This facilitates the channel-by-channel adjustment required in current CIs for patient fitting or that required for high bio-fidelity of the system. In addition, calibration bits can be specifically provided for each channel to compensate potential offsets and mismatches.
across channels. The DACs may be simply several current mirrors whose width is digitally controllable (variable) as illustrated by Figure 5.2.

In this processor, sound is picked up by a commercial electret microphone and transduced into an electrical signal ($V_{IN}$), which is then processed by a preamplifier, or analog front end (AFE). The functions of the AFE include: 1) voltage-to-current conversion, since all the subsequent signal processing is current-mode; 2) single-ended-to-differential conversion for fitting the AFE output to the differential input of the balanced global GMS ($I_{IN}^{u}$ and $I_{IN}^{l}$). The GMS functions as an input conditioner for the Class-AB operation of the OZGF channel as described in Chapter 4 and its output copies are distributed for all the channels. The OZGF-with-LI filterbank performs biomimetic filtering via the OZGF, in conjunction with both multichannel compression and spectral enhancement via the coupled AGC scheme.

Note that the output signal of each channel is taken from the AGC’s GMS output $I_{SUM}$ (instead of $I_{OUT}$ in Figure 4.1) and hence a full-wave rectified signal. For enveloped-based stimulation, the filtered signals require to be rectified and subsequently smoothed. With the reuse of the GMS, the rectification is implemented without needing to resort to additional circuitry. On the other hand, the smoothing after rectification is performed by a separate LPF added to each channel: the LPF in the AGC is not reused just like the GMS. This is based on the fact that for the AGC, the cut-off frequency of the LPF requires to be sufficiently low (i.e., long time constant) as presented in Section 4.4.2, while that for cochlear electrical stimulation is typically set at 200Hz or higher to cover the fundamental frequency ($F_0$) range of speech sounds as noted in Section 1.4. By using separate LPFs, appropriate cut-off frequencies can be set respectively for the two different applications. The extracted envelope information at the LPF output of each channel is then used to modulate trains of biphasic pulses, after which the modulated pulses are directed to a single electrode for stimulation. The pulses are interleaved across channels to ensure non-simultaneous stimulation for minimizing the interaction between channels. The pulse generator can be tailored for various stimulation forms in current use (Section 1.4 have provided a review of these forms).
Figure 5.1: Programmable OZGF-with-LI based CI processor.
Due to the ultra-low-power analog processing, the processor has potential for fully implantable use. It operates without using transistors as switches and thus avoids the high energy dissipated during switching on or off, which is commonly seen in traditional analog-to-digital (A/D)-then-DSP systems, as noted in Section 1.5. If the channel count is sixteen (i.e., $N = 16$ in Figure 5.1), its power consumption (excluding the part for pulse stimulus) should be similar to that predicted in Section 4.6 (i.e., <120 µW) because the increase in power consumption due to the added reference circuits, current-mode DACs, AFE and LPF is negligible compared to the total power consumption of the sixteen-channel system. In the next two sections, we will look into possible improvements at lower-level of the processor – the OZGF channel and its coupled AGC scheme.

5.3 Reconfigurable and Hybrid Channel Architecture

This section presents two possible efforts to make the OZGF channel reconfigurable with respect to the filter cascade and more robust to the noise and offset accumulated along the cascade of the biquads.

5.3.1 Programmable filter order and biomimetic phase response

Figure 5.3 shows that the OZGF slope $S_2$ defined in Figure 3.4 (at the low-frequency side of $\omega_{\text{CF}}$) varies with both the filter order $N$ and quality factor $Q$. On the other hand, the high-frequency side slope $S_3$ is almost constant for different $Q$ values while largely depends on $N$ as illustrated by Figure 3.8. Therefore, increased bio-fidelity of modelling cochlea transfer functions (represented by BM responses) is obtainable for the present OZGF-with-LI system by making the filter order $N$ programmable while $Q$ is level-
dependent via the AGC. Furthermore, fitting of the CIs based on this system can benefit from the programmability of the $N$, by which *patient-specific* frequency selectivity (i.e., that optimized for the individual patient) of the OZGF is implementable.

![Figure 5.3: OZGF slope $S_2$ versus stage $Q$ for different filter order $N$.](image)

Besides the filter shape, the parameter $N$ affects the phase response as described in (27) and (28), where $H_{OZGF}(j\omega)$ is the OZGF transfer function (for its detailed description, see (1) in Section 2.2.1), and $\omega_0$ and $\omega_z$ are the pole- and zero- frequency respectively. The expression (28) provides a *group-delay* view of the phase by negatively differentiating the phase response with respect to frequency. Note that the $\beta$-terms arise from the presence of a non-DC zero (at $\omega_z$) in the numerator of the OZGF transfer function. Based on (28), a series of simple group-delay expressions normalized with respect to $\omega_0$ can be derived at certain frequencies, e.g., $\omega = \omega_{\text{peak}}$ (corresponding to the maximum group delay), $\omega$ at DC and the CF; all the derived delay parameters are functions of both $Q$ and $N$. In [1] (see Figure 18 and Table 3 there), Katsiamis *et al.* have shown that it is feasible to fit such delay parameters to physiological measurements of real cochleae for the Differentiated All-Pole Gammatone Filter (DAPGF), the transfer function of which is similar to that of the OZGF except that the zero is placed at DC (i.e., $\omega_z = 0$). For the OZGF, such
feasibility also exists; in fact, the OZGF response approaches the DAPGF when its zero’s location on the real axis is sufficiently close to DC.

\[ \angle H_{OZGF}(j\omega) = \arctan \left( \frac{\omega}{\omega_0} \right) - N \cdot \arctan \left( \frac{\omega_0 \omega}{Q (\omega_0^2 - \omega^2)} \right) \]  

(27)

\[ T(\omega) = -\frac{d\angle H_{OZGF}(j\omega)}{d\omega} = -\frac{1}{\beta \omega_0} + \frac{N (1 + x)}{\beta \omega_0 [x^2 - 2(1 - \frac{1}{Q^2}x) + 1]} \]  

(28)

(where \( x = (\omega/\omega_0)^2 \) and \( \beta = \omega_z/\omega_0 \))

The foregoing parametric dependence of the OZGF phase response on the filter order \( N \) can be exploited to improve the accuracy of modelling the level-dependent differences in phase response at different places along the BM. In Section 2.2.2, we have presented the potential benefit of this modelling and its possible implementation – the level-dependent regulation of \( Q \) (via the AGC) customized for each channel modelling a different BM segment (with a different phase lag along the BM – see Section 1.2). A programmable \( N \) should result in more flexible parametric modifications towards higher bio-fidelity of the simulated phase cues.

The programmability can be realized by the cascade-architecture described in Figure 5.4, where each lowpass (LP) biquad stage contains an output tap and these output taps are multiplexed into a single output of the channel. Thus, the OZGF channel is \( N^{th} \)-order if the \( N^{th} \) tap is selected out of the total \( M \) taps. The channel architecture is similar to a classical travelling-wave cochlea model except that merely one tap is selected out for use and this tap must locate after the first stage of the cascade, the bandpass (BP) biquad, to guarantee the overall OZGF response. The architecture can be regarded as reconfigurable because it allows flexible changes to the signal path starting from the BP stage and then ending at any LP stage we wish.
Figure 5.4: Improved OZGF cascade architecture with programmable order $N$. 

$N_{th}$-order $(N \leq M)$ 
Output
5.3.2 Hybrid OZGF Channel

A well-known issue with analog processing is that noise and offset accumulate along a cascade of processing stages. For a short cascade (e.g., \(N = 4\), as previously used for the system prototype) this issue is tolerable, but it gets worse and worse as the cascade length increases. For the foregoing \(N\)-programmable channel architecture, the filter order \(N\) is allowed to be high for a large \(M\) potentially required for a high bio-fidelity of the filtering; thus, attention needs to be paid to its offset- and noise-robustness. A potential solution to this issue, which may also optimizes the overall system’s area and power consumption, is illustrated in Figure 5.5(a). The architecture performs distributed and hybrid computation, the idea of which was proposed by Sarpeshkar [2] and is as follows:

- It employs an analog encoder to distribute the information present on a single wire over many wires (channels) for the subsequent processing occurs on each wire. The analog encoder is an A/D/A system [3] – an analog-to-digital (A/D) converter immediately followed by a digital-to-analog (D/A) one. For instance, given that the information is first represented by 8 bits (via 8-bit A/D), then it can be encoded as four analog signals onto four wires each carrying 2 bits of the information (each via 2-bit D/A). Note that the analog processors (denoted “A”) on different wires does not operate independently; their interactions (indicated by the double arrows) are necessary for preserving the distributed representation. Sarpeshkar demonstrated that the total area or power consumption cost could be efficiently reduced by using many distributed low-precision processors instead of a single high-precision processor.

- It offers the offset- and noise-immunity for a long cascade of analog processing stages by dividing the cascade into a sequence of hybrid links each containing an A/D/A block inserted after a short cascade. The A/D/A restores the analog signal to one of \(M\) predefined attractor states: it is rounded off to the closest state level by comparing with \(M-1\) threshold levels. The robustness arises if the difference between an attractor-state level and a threshold level is sufficiently large compared to the variance of the accumulated offset or noise (including the input referred noise of the A/D/A from its internal circuits) that arrives at the comparison stages in the A/D/A. The offset/noise accumulation is ‘reset’ to a restored signal level after every hybrid link, and the precision is maintained to be \(N\)-bit \((N = \log_2(M))\) at the output of the whole cascade—given that all the
A/D/As each are \(M\)-state—even if the cascade consists of a considerably large number of processing stages. In the example of Figure 5.5(a), the A/D/A in the first half of each wire (channel) is 2-bit, giving four attractor states (i.e., \(M = 4\)) to decide for signal restoration (denoted ‘00’, ‘01’, ‘10’, and ‘11’); in the next hybrid link, the A/D/A reduces the output precision to only 1 bit so that the analog processing circuits there becomes even less sensitive to noise and offset.

Figure 5.5(b) shows the OZGF channels applying A/D/As, where the foregoing analog processing is now performed by bandpass (BP) or lowpass (LP) biquad. Each channel is reconfigurable (as described in Figure 5.4); for clarity, its multiplexing operation is not shown in the figure. The insertion of an A/D/A after every short cascade of LPs, which results in a hybrid link, should increase each channel’s robustness, but the precision achievable at the channel output is not fully determined by the resolution of a certain A/D/A (say \(N\)-bit). In practice, the precision is related to the location of the selected output tap: the closer to the preceding A/D/A it locates, the closer to \(N\)-bit its output precision is; nevertheless, the actual output precision will not be significantly worse than \(N\)-bit since each hybrid link is a short cascade of LP biquads.

The A/D/A may be implemented as depicted in Figure 5.6. Note that the D/A circuit is very similar to that shown in 5.2; the difference lies in the control of the switches via A/D: the copies of an analog input are compared with the three reference currents to generate the three controlling signals respectively. We choose to implement the current-mode A/D/A, rather than the voltage-mode (e.g., that used in [2]) mainly because it requires no additional I-to-V circuitry to fit its preceding current-mode filters. Also, for a given power-supply voltage, it may provide more available attractor states (and hence more bits of precision) while maintaining a sufficient bit distance in current space (i.e., \(I_{\text{bit}}\)) instead of voltage space (limited by the power supply voltage) to minimize the probability of a bit-error.

It is interesting to see that the OZGF-with-LI filterbank performed signal processing naturally in a distributed manner. The IDN and BP biquad together serve as the foregoing analog encoder but without the A/D/A operation: the former makes a global analog-input signal available for all the channels, while the latter tailors this signal to the passband of each channel for the subsequent channel-specific processing via the LP-cascade. In addition, the foregoing interaction between the analog processors is now

---

* It is only an impression of the A/D/A circuitry; there are much more efficient alternatives.
realized by the coupled AGC scheme, which simulates the LI – a competitive interaction between neurons.

Figure 5.5: (a) Distributed and hybrid computation [2]; (b) Improved OZGF with LI filterbank using hybrid channel architecture.
Figure 5.6: A 2-bit A/D/A: (a) the three threshold currents ($I_{T1}$, $I_{T2}$ and $I_{T3}$) and their resulting four attractor states ($I_{L1}$, $I_{L2}$, $I_{L3}$ and $I_{L4}$); (b) a possible circuit implementation.
Note that for practical use of the above architecture, a few parameters (besides those inherent in the OZGF and its AGC) still need to be decided: the optimal number of bits allocated to each channel and of analog processing stages prior to each A/D/A, etc., in terms of the total area or power consumption, or both. Although in [2] Sarpeshkar has demonstrated the way to optimize these parameters, his work only takes into account the simplest cases for providing insight, e.g., each channel processes the same number of bits of information, and all the analog processing stages are unity-gain. In practice, the precision (represented by the number of bits) does not have to be the same for each channel. For instance, we may specify higher precision (more bits) for low CF channels, and vice versa, as illustrated in Figure 5.5(b) (one bit less from low CF to high), since most useful spectral information for speech perception is known to be at low frequencies. An extension of Sarpeshkar’s work regarding the parameter optimization can be involved in our future work and therein specific to the OZGF-with-LI filterbank.

### 5.4 Enhanced Lateral Inhibition

In Section 4.4.3, we have presented a circuit implementation of the LI (see Figure 4.18), which uses pairs of current-scaling current mirrors to realize different sets of weighting factors used in the AGC coupling function (see (23)). While such an implementation has proven efficient, it may cost much silicon area if a large number of sets of weighting factors are required for more options of the amount of the simulated LI: $M$ sets of weighting factors requires $M$ pairs of current-scaling current mirrors; given that they are programmable via $N = \log_2(M)$ bits, then every time adding one bit more requires two more pairs of current mirrors correspondingly. Thus, we propose here an alternative implementation for saving area. The circuit implementing a scaling factor of $k_1/k_2$ is shown in Figure 5.7 where all the transistors are assumed to be matched. It operates in accordance with the translinear (TL) principle as follows (the transistors $M_1 \sim M_4$ forms a TL loop):

$$k_1 I_s \cdot I_{IN} = k_2 I_s \cdot I_{OUT} \quad \Rightarrow \quad I_{OUT} = (k_1/k_2) \cdot I_{IN}$$

(29)

where $k_1$ and $k_2$ are two programmable factors and $I_s$ is a reference current. In its practical use for the AGC coupling scheme, $I_{IN}$ here corresponds to the signal $I_{ED}$, or the sum of the four signals respectively from the four neighboring channels in (23) (i.e., $I_{ED, Ness} = I_{ED, i+1} + I_{ED, i+2} + I_{ED, i-1} + I_{ED, i-2}$), and correspondingly $W_0 = k_1/k_2$ or $W_H = W_L = k_1/k_2$. Given that $k_1$ is programmable via $N_1$ bits and $k_2$ via $N_2$ bits, and given that we
use the DAC circuit shown in Figure 5.2 where $I_{LSB}$ corresponds to $I_w$ in (29) for the programmability, then adding one bit more to $N_1$ or $N_2$ only requires two more transistors in the circuit – one for $W_0$ and another for $W_1$ ($= W_I$), instead of the foregoing two more pairs of current mirrors.

Another possible improvement lies in the AGC-coupling manner itself. Previously, we chose to implement the AGC coupling as a simple summation operation rather than a Root-Mean-Square (RMS) operation of signals, because of our finding that the former is able to achieve a good trade-off between performance and complexity through a careful manipulation of the weighting factors (for discussions, see Section 2.3.2 and Figure 2.13 therein). The flexibility of such manipulation, however, is limited by the finite number of bits in a digitally programmable OZGF-with-LI system (e.g. that described in Figure 5.1). For a given number of bits, the amount of the simulated LI may be enhanced by use of the RMS coupling because it adds the ‘weights’ to the strong signals relative to the weak signals by squaring these signals along with their respective weighting factors in the coupling function.

Figure 5.8 shows a possible TL circuit for the RMS coupling (the transistors $M_1 \sim M_4$ forms a TL loop) whose operation is governed by
\[
\frac{I_{ED,i}^*}{2}, \frac{I_{ED,i}^*}{2} = I_s \cdot (I_s \pm I_{ED,Neigh}^*)
\]  
(30)

where \(I_{ED,i}^*\) is an weighted (denoted ‘\(^*\)’) envelope signal detected from the outputs of the channel \(i\) while \(I_{ED,Neigh}^*\) is the sum of those detected from the neighbouring channels, and \(I_s\) is the drain current of the transistor \(M_t\) given by

\[
I_s = \frac{1}{2} (I_{CP} \mp I_{ED,Neigh}^*)
\]  
(31)

where \(I_{CP}\) is the coupled output current. By substituting (31) into (30), we have the following input-output relation:

\[
I_{CP} = \sqrt{(I_{ED,i}^*)^2 + (I_{ED,Neigh}^*)^2}
\]  
(32)

Note that the above function is not strictly RMS as it omits the averaging operation to get the arithmetic mean of the squares; more accurately, it may be called root-sum-of-squares; nevertheless, we use the term ‘RMS’ here to refer to (32) and its corresponding circuit in Figure 5.8 since the omitted averaging can be thought to be already lumped into the weighting operation.

Figure 5.8: TL-based implementation of the RMS coupling (a CMOS version of the counterpart in [4, 5]. The double arrow indicates that the sign of \(I_{ED,Neigh}^*\) in (30) and (31) can be either positive or negative.

Adding a TL-based RMS circuit like that in Figure 5.8 should not significantly increase the total area of the OZGF-with-LI system since it is very simple and consists of only five transistors. On the other hand, if larger area consumption is allowable for achieving even higher bio-fidelity of the simulated LI, then we may make the simulated LI specific
to the individual constituent signals in the coupling function by using the weighting factors customized to each of these signals and squaring them each instead of the combined signal $I_{ED,Neigh}$. The resulting coupling function is

$$I_{CP} = \sqrt{(w_i \cdot I_{ED,i})^2 + (w_{i+1} \cdot I_{ED,i+1})^2 + (w_{i+2} \cdot I_{ED,i+2})^2 + (w_{i+3} \cdot I_{ED,i+3})^2 + (w_{i+4} \cdot I_{ED,i+4})^2}$$

(33)

where are $w_i$, $w_{i+1}$, and $w_{i+2}$ are the respective weighting factors for the coupling signals. For example, $w_{i+1}$ can be set larger than $w_{i+2}$ to more accurately mimic one of the 2TS effects: the tone-to-tone suppression becomes stronger when the two tones are closer to each other, as mentioned in Section 2.1.1 and simulated in Section 2.3.3. Equation (33) can be implemented using the TL circuit shown in Figure 5.9, by which the RMS-coupling manner is generalized to apply to $k$ (an integer) coupling signals as given by

$$I_{CP} = \sum_{j=1}^{k} I_{CP,j} = \sum_{j=1}^{k} (I_{ED,j})^n \rightarrow I_{CP} = \sqrt[2n]{\sum_{j=1}^{k} (I_{ED,j})^n}$$

(34)

where the superscript **"n"** denotes the weighting, and $I_{CP}$ is the sum of $k$ coupling signals raised to the $n$-th power; $k = 5$ and $n = 2$ in (33). Stronger LI is obtainable by use of higher $n$ due to more ‘weights’ given to strong constituent signals in (34).

Figure 5.9: The root-sum-of-powers circuit implementing (34) (a CMOS version of the counterpart in [5, 6]).

The above circuit should be very helpful for an OZGF-with-LI filterbank with more channels (e.g., 32 channels) where more constituent signals (i.e., larger $k$) need to be present in the coupling function to maintain the quality of spectral enhancement, as suggested in Section 2.5. Certainly, $k$ and $n$ must not be too large since tolerable area consumption is also attempted: it can be seen from Figure 5.9 that adding one to $k$
requires three more transistors while adding one to $n$ requires $k$ more transistors! Having this in mind, our future work will look into the possibility of making both $k$ and $n$ programmable, e.g., by means of multiplexing (also applicable for a programmable filter order of the OZGF as presented previously).

5.5 Conclusion

In this chapter, we looked into some likely next steps in developing the present OZGF-with-LI filterbank towards a practical CI system. These possibilities place an emphasis on three aspects of the system: programmability, robustness and bio-fidelity. We suggested that the improvement to the three could be achieved by:

- Adding DACs to alter system parameters, through which offset compensation is implementable and higher bio-fidelity is obtainable.
- Using a reconfigurable and hybrid channel architecture where the filter order and its associated phase information are programmable via multiplexing while the noise/offset robustness could be maintained due to signal restorations.
- Adopting the TL design strategy to implement the AGC coupling as it provides an economical (in silicon area) way to further enhance the simulated LI effect.

The above possible efforts may increase the complexity, area and power consumption of the system to some extent, but we believe the cost will be quite tolerable and should not affect the system’s potential for use in fully implantable CIs, since:

- All the signal processing is still maintained in the analog domain where ultra-low power design techniques are applicable.
- The current-mode DACs and multiplexers can be very simple (e.g., via current-mirror architecture), and thus their contribution to the cost is small.
- The TL circuitry is simple and has high function density [5] (e.g., a single RMS circuit can implement the function involving squaring, summation and square-rooting), thereby having potential for saving area while implementing the AGC coupling. Note that its complexity will increase when more constituent signals are required to be present in the coupling function; however, since in this case the total number of channels in the system also increases, its contribution is still low relative to that from the increase channels.
Now a major contribution to the cost may be from the signal restorations each performed after a short cascade of analog processing stages, because it involves the switching of several discrete states. Fortunately, the precision required there is low (e.g., for high-CF channels) or moderate (e.g., for low-CF channels) and the resulting cost may be tolerable. Furthermore, we might deliberately simplify the circuit topology of the analog processors and/or scale down the sizes of the transistors therein since they can afford to be noisier and/or yield higher offsets due to the restorations.

In conclusion, the possibilities presented in this chapter could be regarded as a good trade-off between performance improvement (on programmability, robustness and bio-fidelity) and its area and power costs. Certainly, there are still a large number of possible efforts we may make to improve the present design. The eventual goal of our future work will be to yield an OZGF-with-LI based CI system which is fully-implantable and meantime offers high programmability, robustness and bio-fidelity.
References


Chapter 6

Conclusion

My PhD research went through four phases: concept design, computational modeling, speech-recognition evaluation and hardware implementation. Aspects (contributions) of each phase are summarized as follows:

- **Concept design.** We proposed a novel filterbank architecture termed “OZGF-with LI”, which mimics the lateral inhibition (LI)—a biological spectral enhancement mechanism that may account for high noise robustness of the human auditory system—by means of feedback AGCs that are cross-coupled across channels so that each channel-AGC ‘senses’ not only the signal intensity of the corresponding channel output but also those of the neighboring channel outputs. The simulated LI effect manifests itself as a soft local “winner-take-all” link between different frequency regions in the spectrum; that is, spectral prominences (e.g., formants of speech) suppress their neighboring weak frequency content, resulting in enhanced local spectral contrast. The architecture attempts to restore the LI mechanism that is ‘missing’ in current CI processing which only provides a very crude approximation to the human auditory processing.

Within the filterbank, each channel filter is called the One-Zero-Gammatone-Filter (OZGF) [1, 2], which provides an asymmetric bandpass-like function with desirable tunability and hence a robust foundation for modeling cochlea transfer functions and a variety of the associated auditory data. The OZGF was adopted in our work because of its high bio-fidelity and hardware-compatibility. Its quality factor, which controls both the filter gain and shape, changes adaptively via the coupled AGC in accordance with a highly programmable, compressive control law. As a consequence, the OZGF-with-LI architecture produces a close intertwining of both filtering and compression as observed in the auditory
system rather than the artificial separation of the two functions in traditional CI processors.

Owing to the LI-based spectral enhancement, the OZGF-with-LI system performs multi-channel compression while preserving well the spectral contrast of the system input. Therefore, the “spectral-contrast-friendly” compression scheme provided by this system is advantageous over that used in current CI processors, which degrades spectral contrast and thus impairs CI users’ ability to pick out the spectral peaks as the identity of various speech sounds embedded in noise. From the perspective of noise reduction, due to the simulated LI mechanism, spectral prominences in the input spectra can have the effects of suppressing the background noise around them. In conclusion, the proposed system, and its associated new re-processing strategy, which is highly compatible to current stimulation approaches, provides a potential solution to CI users’ susceptibility to noise, which has been considered as a major limitation of current cochlear prosthesis.

- **Computational modeling.** We developed a computational model of the OZGF-with-LI system in MATLAB, through which we investigated thoroughly how the parameters and some alternative topologies affect the effectiveness of the spectral enhancement provided by this system. The model is VLSI-compatible, and more specifically, it is very amenable to low-power analog VLSI implementation due to its simple structure and parameterization – our design aimed at achieving a good trade-off between performance and complexity of hardware implementation. This phase of research realized the proposed concept of the OZGF-with-LI processing, and demonstrated its working and benefits with a variety of simulation experiments. The findings and draw conclusions during this phase can serve as “guidelines” for how to optimize the parameters and the topology effectively 1) to achieve even higher level of spectral enhancement and improved bio-fidelity, and 2) to maintain the effectiveness of enhancement in the filterbank of increased channels, thereby extending the use of this architecture in future CIs of improved spectral resolution, 3) to fit the OZGF-with-LI based CI processor of the future to patients’ individual hearing situations.
Speech-recognition evaluation. In order to complement the simulation evaluation, we evaluated the proposed OZGF-with-LI system via a psychophysical study of its effects on the speech intelligibility of listeners, which was measured during speech-recognition tasks. This phase of research had two major attempts: the first was to investigate whether the OZGF, a filter type of increased bio-realism including sharpness and asymmetry in passband, could produce improvements in speech intelligibility in noise; the second was to investigate whether the side effect of multi-channel compression schemes, i.e., reduced spectral contrast, could be effectively compensated by reproducing the LI mechanism.

Correspondingly, two perceptual experiments were conducted based on the acoustic simulation of CIs via a noise-excited envelope vocoder incorporating the OZGF-with-LI filterbank, and based on comparisons in various processing conditions. The results showed that 1) the bio-realistic asymmetric sharpness of tuning of the filter, provided by the OZGF, cannot produce an improvement in the speech intelligibility in a steady-state noise background, at least with our present parameter settings and with only 16 channels, which is likely accounted by the fact that the potential benefits have been severely counteracted by the loss of useful spectral content due to the very limited channel number combined with the sharp tuning of channel filters – however, it remains to see whether the potential benefits would emerge when more channels are applied; 2) the simulated LI mechanism can provide a large improvement in speech recognition performance in noise, compared with the case in which multi-channel compression was performed alone, thereby degrading the spectral contrast that is present in the input speech stimuli.

Hardware implementation. Based on the computation model, we realized an ultra-low-power analog OZGF-with-LI system by employing MOS transistors operating in their weak-inversion regime (i.e., subthreshold circuits) and low-power wide-DR design techniques including Log-domain, Class-AB and Dynamic Biasing via AGC (referring to syllabic companding). The system itself does not impose a limit on any technological medium – it can be implemented in analog or digital. We opted for analog approaches because they are known to offer desirable power- and area-saving solutions; our efforts aimed at an IC prototype that can meet the very stringent power consumption requirements of
fully implantable CIs of the future and thus has potential for such applications. A five-channel silicon IC prototype of this system, which was fabricated in the commercially available 0.35µm AMS CMOS process, has demonstrated such potential; measurements from this prototype shows that it consumes merely 28µW of static power while providing a wide input DR of over 90dB (at <5%THD) at channel CFs. It is most likely that the prototype’s sixteen-channel version will have a 3~4 times higher power consumption compared to the prototype; this power consumption figure is still very low though and comparable to those reported in pioneering works towards fully-implantable CIs. Therefore, the analog OZGF-with-LI system should have potential for use in fully-implantable CIs of the future. Probably the most attractive aspect involved this phase of research is the VLSI realization of the LI via low power and compact coupled AGC circuits; we believe our way of realization of the LI provides a desirable compromise between bio-fidelity, silicon area, and power consumption, since it only needs one special AGC building block to implement the coupling function and thus does not significantly increase the overall system’s area and power consumption.

In this thesis, we have also described some possibilities for further improvements in our current design with a good trade-off between improved system performance and its area/power costs. These possibilities, which can be taken into account in our future work, include digital control of the parameters via current-mode DACs, reconfigurable and hybrid channel architecture with programmable filter order (via multiplexing) and increased noise/offset robustness (via A/D/A), and TL strategy based implementation of the AGC coupling. The idea behind these possible efforts is to push the present OZGF-with-LI system closer to a practical cochlear prosthesis system with desirable programmability, robustness and bio-fidelity—which can be regarded as the eventual goal of our future work—by combining the best of both analogue and digital world.

Other possibilities in our future work, which have been presented in the conclusion sections of previous chapters (Chapter 2, 3 and 4) and correspond to the above four research phases, can be summarized as follows:

---

* This channel number has been widely used in current CIs.
† The simulated power consumption was 106µW.
- **Psychophysical/physiological-based parameter optimization of the computational modeling.**

Our present parameter settings are primarily based on the resulting quality of spectral enhancement and also chosen to facilitate various simulation experiments (e.g., having the formants with approximately the same magnitude in different processing cases so that comparison in spectral contrast can be easily done by checking the valleys between the formants). Since our psychophysical evaluation has shown the benefits of the coupled AGC using the present parameter settings, the next step for obtaining even larger improvements in speech intelligibility may be the parameter optimization based on listeners’ score and feedback during speech-recognition tasks. In addition, thanks to high programmability of the OZGF response and the simulated LI, we may choose parameter values to have further improved bio-fidelity, which may in turn improve the intelligibility; specifically, reported data from physiological measurements can be used as the reference for choosing parameter values – we attempt to fit the OZGF-with-LI processing to physiological data. This may require customizing the AGC parameters for each channel (e.g., to better simulate the level-dependent differences in phase response along the BM while each channel is mapped to a different BM section) as well as the weighting factors for different frequency sides (i.e., \( W_{HL} \neq W_{IL} \), whose respect values can be chosen to fit to the differential growth of suppression for low- and high-frequency side suppressors).

- **“Preprocessing-then-vocoding” based speech-recognition evaluation.** Our present perceptual tests employed a “pared-down” version of the traditional vocoding-based experimental set-up; that is, we integrated the OZGF-with-LI processing into the vocoding, rather than a traditional separation of the two processing stages, since our evaluation was intended for a typical number of channels (i.e., 16 channels) used in current CIs and thus the OZGF-with-LI processing under evaluation is compatible to the present processing strategies. In future work, we can adopt a preprocessing-then-vocoding implementation (i.e., using the OZGF or BP-cascade pre-processed stimuli as inputs to the vocoder) so that the preprocessing stage can have much more channels than the latter vocoding stage while the vocoding stage can capture those important aspects of current CIs with a typical number of channels and typical filter types (e.g., the Butterworth filter), as illustrated in Figure 3.10. Then we can investigate whether the sharp
frequency tuning caused by the OZGF pre-filtering can provide improvements in intelligibility when the spectral resolution is largely improved due to much more preprocessing channels. In addition, this experimental set-up can facilitate our further study on some other aspects of the CI processing since the parameters of the two stages and hence their performance can be adjusted separately. For instance, we can adjust the vocoding parameters to introduce a mismatch between the slope of the analysis- and the synthesis- filters so that the spectral-smearing effects caused by channel interactions can be simulated, while all the parameters of the pre-processing stage are retained. Then it will be interesting to see whether the foregoing sharp frequency tuning of the pre-filtering can render input spectral features more robust to the simulated effects of channel interactions at the vocoding stage. Other possibilities based on this experimental set-up include tests in some of the most challenging listening situations for CI users, e.g., speech with spectrally and/or temporally fluctuating maskers, as opposed to the steady-state noise used presently. Following these vocoder-based evaluations with NH people, direct perceptual tests with real CI users can be conducted.

- **A complete low-power, noise-robust, programmable OZGF-with-LI based CI processor.** Since measurements from the fabricated five-channel prototype have shown the working and benefits of the coupled AGC scheme, the next step in VLSI realization can be to add more channels (e.g., resulting in a total of 16 channels) together with other essential components like a microphone pre-amplifier preceding the global GMS, electrical stimulation circuits following the OZGF filterbank, and D/A converters for altering the parameters digitally. An overview of the resulting CI processor has been provided in Figure 5.1. Despite more channels and components, the potential power consumption can still meet the stringent requirement of fully implantable CIs, and thus this CI processor is suitable for such implant systems. As noted previously, improvements in programmability, bio-fidelity and noise/offset robustness may be provided by reconfigurable filter cascade (via a multiplexer) organized in a hybrid A/D/A channel architecture and TL-based AGC coupling circuits. Future tests on this CI processor can be more complex and ‘realistic’ through using natural speech embedded in noise, which will complement our present simplified tests using single-tone and complex-tone input signals.
Three publications on my PhD research are in progress, and involve 1) the design, simulation and evaluation of the OZGF-with-LI system, which corresponds to the work presented in Chapter 2, 2) the psychophysical evaluation of the OZGF-with-LI processing strategy, which corresponds to the work presented in Chapter 3, and 3) the ultra-low-power analog VLSI implementation of the proposed system, which corresponds to the work presented in Chapter 4. Some ideas described in Chapter 5 are also included in the third publication.

References


Appendix A:

Acoustic Experiment – Summary/Protocol

**PI:** E.M. Drakakis, Reader, Dept of Bioengineering  
**Co-I:** Mr Guang Yang, PhD applicant, Dept. of Bioengineering

**Study:** Assessing the effect of spectral enhancement upon intelligibility when speech is processed by a bank of filters.

**Background & Project Summary:**
As part of our research in electrical stimulators for the inner ear we wish to investigate experimentally whether: a) the shape of the filters of our audio pre-processor affects speech intelligibility and b) whether the coupling of channels also affects intelligibility. To this end we have developed a novel piece of a computer program which allows for acoustic experimentation along the lines explained above. The purpose of this application is to seek for permission to elicit interest for at least 16 native English speakers (ideally 8 men and 8 women) from our current UG and MSc students who would take part in one experiment. The experiment itself will involve listening (by means of a headset attached to the PC where our program runs) to specific sentences processed in different ways. The sound intensity level will be kept constant and at comfortable level. Each volunteer will be given the opportunity to familiarize with the platform for ~20mins. Subsequently to familiarization the ~40min experiment will take place. This particular piece of research is original and the results will contribute to the improvement of electrical stimulators for the inner ear.
**Protocol:**

**Test Material:** Speech stimuli (sentences) of the form “Ready Arrow, go to <colour> <number> now.” <Colour> is one of the set [white, red, blue, and green]. <Number> is one of the digits between one and eight. The stimuli will be presented via earphones and at a comfortable loudness level.

**Procedure:** During the test, listeners need identify the keywords “<colour> <number>” in the noisy sentences by moving the mouse cursor and clicking the corresponding button on a graphical user interface. These sentences are processed by various speech processing strategies that stimulate the effects of cochlear implants.

The experiment consists of two phases:

- **Phase 1 (Practice)** – practice (around 20 minutes) will be allowed for familiarization with the computer’s graphical interface and the processed speech materials. Feedback will be provided during this practice phase but not in the following tests, showing the correct <colour> <number> if the identification is wrong.

- **Phase 2 (Test)** – The test takes about 40 minutes. By using the mouse, listeners identify the key words in the sentences processed by various strategies including spectral enhancement that potentially improve speech performance in noise. These strategies will be compared and evaluated according to the resulting speech intelligibility (scored in terms of the correctness).

**Risk:** The participants’ risk associated with the particular study is equivalent to the risk of using a PC and a pair of headphones. No risk can be identified.

**Location:** Room B503, Level 5, Bessemer Building

**Persons responsible for the conduction of the experiment:** Mr Guang Yang and Dr. E.M. Drakakis