2-Microphone Noise Reduction for Cell Phones based on the Human Auditory Pathway

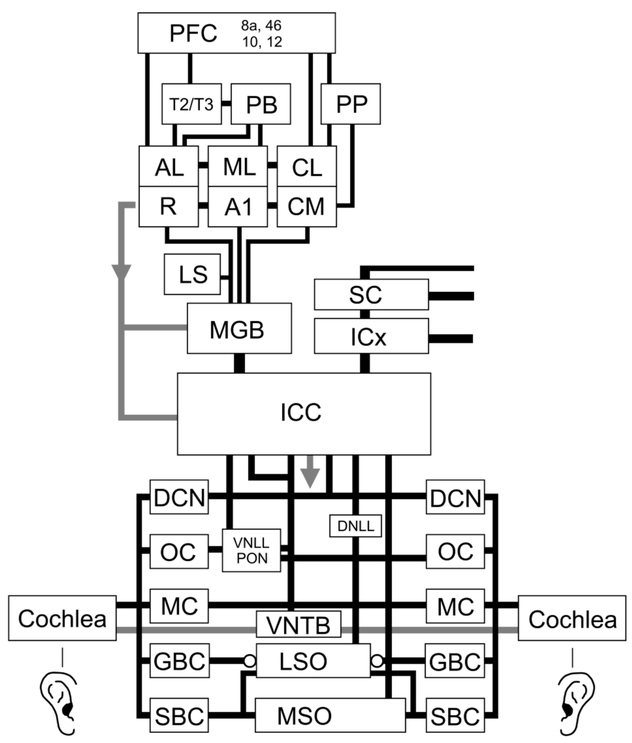
Lloyd Watts
Audience, Inc.
lwatts@audience.com

Abstract

Recent advances in computing and auditory neuroscience have now made it possible to produce high-resolution, real-time simulations of major portions of the human auditory pathway, including high resolution, real-time models of the cochlea, major cell-types of the cochlear nucleus, lateral and medial superior olivary complex, as well as polyphonic pitch perception based on combination-sensitive cells measured in inferior colliculus and auditory cortex. These technologies are the foundation of Audience's Voice Processor chip, which provides advanced two-microphone noise reduction for cell phones. The system achieves an average 0.77 MOS improvement over an SNR range of 0-12dB, with an active current consumption of 32mA, and by 2010 has been widely deployed in a number of popular cell phone models.

Introduction

Recent advances in computing and auditory neuroscience have now made it possible to produce high-resolution, real-time simulations of major portions of the human auditory pathway, including high resolution, real-time models of the cochlea, major cell-types of the cochlear nucleus, lateral and medial superior olivary complex, as well as polyphonic pitch perception based on combination-sensitive cells measured in inferior colliculus and auditory cortex. The human auditory pathway consists of many organs and brain regions, as shown below [1-5]:

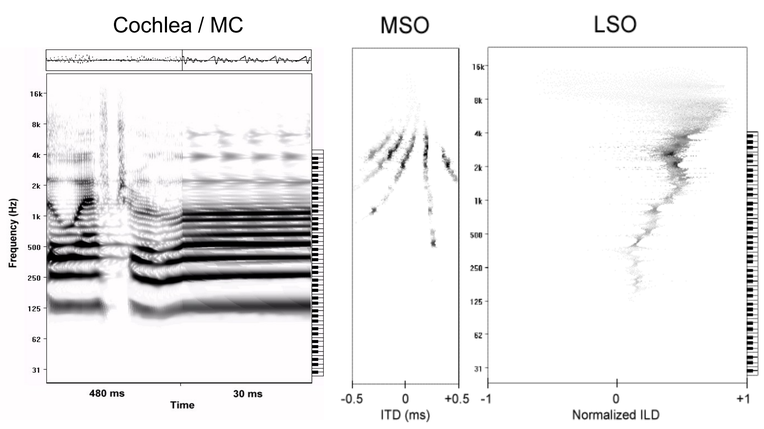


Audience's product is based on the operations of the Cochlea, Multipolar Cells (MC), and the cell regions associated with binaural processing – Globular and Spherical Bushy Cells (GBC, SBC) in the Cochlear Nucleus, the Lateral and Medial Superior Olivary Complexes (LSO and MSO), and the Central Nucleus of the Inferior Colliculus (ICC).

The advent of GHz PCs in 2000 made it possible to run biologically accurate models in real-time on inexpensive personal computers. And with Audience's dedicated low-power chip, it became possible in 2007 to run these algorithms in real-time in a cell-phone, consuming about 30mW of power.

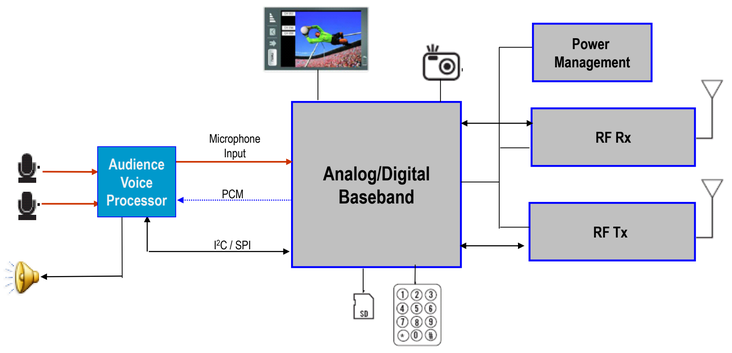
Biological Accuracy

The output of the Cochlea / Multipolar Cell model [6-7] and MSO / LSO model [8] is shown below, and is available at the meeting as a running real-time demonstration.

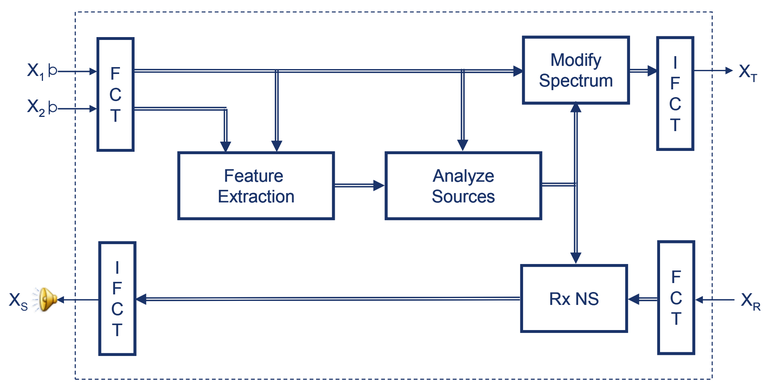


System Architecture

Audience's Voice Processor chip fits into a standard cell phone architecture as shown below:



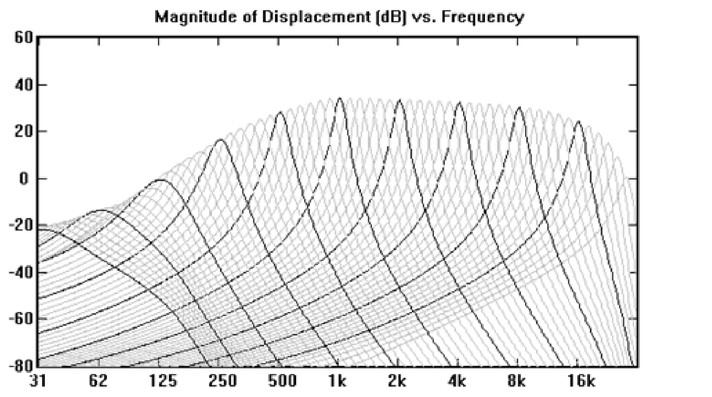
The chip receives analog inputs from two microphones, analogous to the two ears in the human auditory pathway. The chip then digitizes those signals, and transforms them into the cochlea domain using Audience's proprietary Fast Cochlea Transform (FCT). These cochlea-domain signals are then analyzed for binaural differences in the Feature Extraction block, analogously to the MSO / LSO in the human auditory pathway. These computations allow the Analyze Sources block to identify the primary voice based on physical location, and then modify the spectrum from the primary microphone X_1, and then the Inverse FCT (IFCT) transforms the modified spectrum back into the time domain for encoding for transmission by the Baseband Processor.



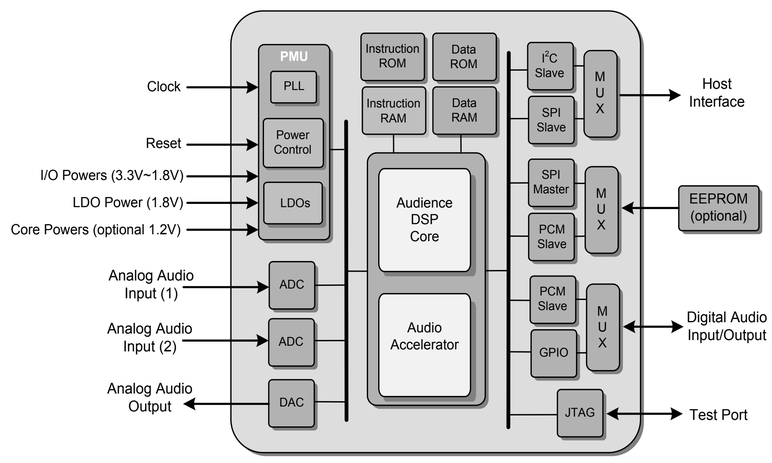
Similar operations are performed in the receive direction to provide stationary noise reduction using only one channel.

Hardware Acceleration

The Audience Voice Processor has been optimized to efficiently run the most common and computationally intensive algorithms, including the FCT, which is a proprietary implementation of the Pole-Zero Filter Cascade [6-7]. The core operations of parallel multiplies and adds in the FCT have been implemented in dedicated hardware to provide accelerated performance at reduced power consumption for this critical operation. In addition, there are dedicated instructions for other common nonlinear operations such as *sqrt*, *log*, *exp*, etc.

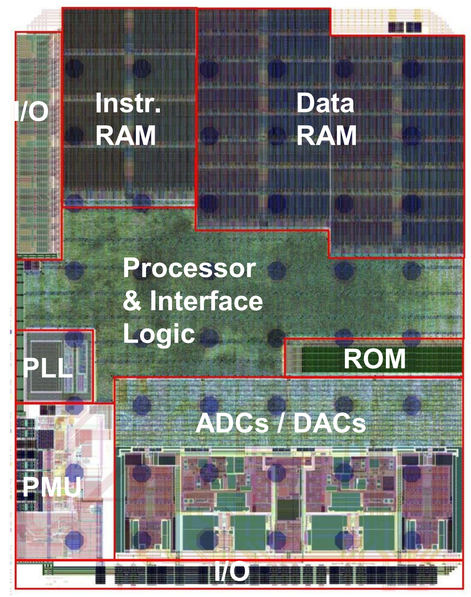


Hardware Architecture



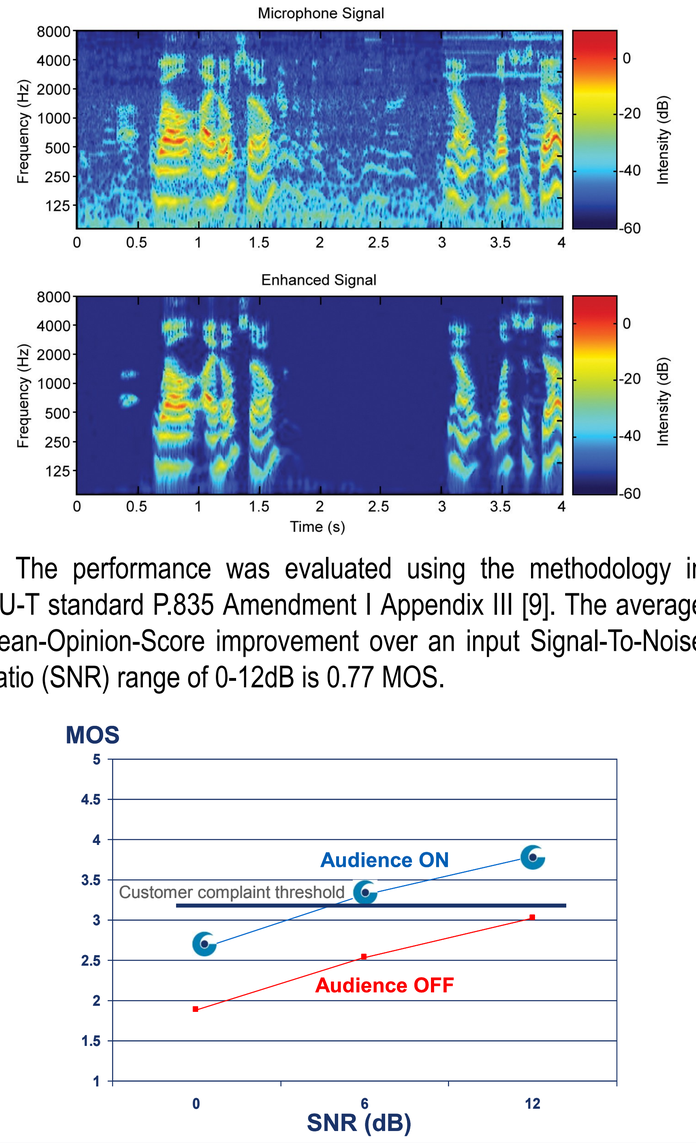
Hardware Design

The Audience Voice Processor runs on 1.8-3.3V power supplies, with optional 1.2V core power input. It is packaged in a 48-pin Chip-Scale Package (CSP), with 0.4mm ball spacing. Active current consumption is 15-32mA, depending upon which features are enabled. It consumes 30µA current when in sleep mode. The chip was fabricated on the TSMC 130nm process with a die size of 2.7 x 3.5 mm.



Performance

The figure below shows the effect of the Audience Voice Processor noise reduction on a real recording made on a busy street.



The performance was evaluated using the methodology in ITU-T standard P.835 Amendment I Appendix III [9]. The average Mean-Opinion-Score improvement over an input Signal-To-Noise Ratio (SNR) range of 0-12dB is 0.77 MOS.

References

1. L. Watts, "Visualizing Complexity in the Brain," in *Computational Intelligence: The Experts Speak,* D. Fogel and C. Robinson, Eds., IEEE Press/Wiley, pp. 45-56, 2003.
2. G. Shepherd, *The Synaptic Organization of the Brain,* Fourth Edition, Oxford University Press, p. vi.
3. E. Young, "The Cochlear Nucleus", in *The Synaptic Organization of the Brain,* Fourth Edition, G. Shepherd, Ed., Oxford University Press, pp. 121-158.
4. Yin, T. (2002), in Oertel, D., Fay, R., and Popper, A., ed., *Integrative Functions in the Mammalian Auditory Pathway*, Springer-Verlag, New York, pp. 99-159.
5. Casseday, J., Fremouw, T., Covey, E. (2002), in Oertel, D., Fay, R., and Popper, A., ed., *Integrative Functions in the Mammalian Auditory Pathway*, Springer-Verlag, New York, pp. 238-318.
6. R. F. Lyon, "A Signal-Processing Model of Hearing", Xerox PARC technical report, 1978. Available at www.dicklyon.com.
7. M. Slaney, "Lyon's Cochlear Model," Apple Technical Report No. 13, Advanced Technology Group, Apple Computer, Inc., Cupertino, CA, 1988.
8. L. Watts, "Computation of Multi-Sensor Time Delays", United States Patent Number 6,792,118, September 14, 2004.
9. ITU-T P.835 Amendment I Appendix III, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm* – Appendix III: *Additional provisions for non-stationary noise suppressors.* Available at http://www.itu.int/rec/T-REC-P.835-200710-I!Amd1/en.

Speech in Noise Workshop
January 6-7, 2011, Lyon, France

Audience