MEL-FREQUENCY CEPSTRUM ENCODING IN ANALOG FLOATING-GATE CIRCUITRY

Paul D. Smith, Matt Kucic, Richard Ellis, Paul Hasler, David V. Anderson *

Georgia Institute of Technology
Department of Electrical and Computer Engineering
Atlanta, GA 30332

ABSTRACT

This paper presents a continuous-time mel-frequency cepstrum encoding IC using analog circuits and floating-gate computational arrays. We present the dynamics of several floating-gate computational buildings blocks and accompanying experimental measurements. We also present a novel approach to programmable signal spectrum decomposition, analog frequency transforms, and spectrum compaction. Experimental data is presented from circuits fabricated on a 0.5μm nwell CMOS process available through MOSIS. This system can act as the front-end for larger digital or analog speech processing systems.

1. CEPSTRAL ANALYSIS OF SPEECH SIGNALS

From the general model of speech production, speech is the convolution of an excitation sequence, which is a pseudorandom sequence, with an impulse response of the vocal system [1]. Many applications require some knowledge or at least an estimate of either the pseudorandom excitation or the vocal tract response or both. For example, speech recognition is concerned with the vocal tract response independently of the excitation sequence; and speech coders can achieve better compression if the excitation and vocal tract characteristics are coded separately. Extracting the excitation or vocal tract response from a speech signal is nontrivial because they are combined by convolution:

$$s(n) = e(n) * \theta(n), \tag{1}$$

where $s(n)$ is the speech signal, $e(n)$ is the excitation, and $\theta(n)$ is the vocal tract response.

Cepstral analysis is a special case within a general class of methods known as "homomorphic" signal processing [2]. The purpose of the real cepstrum (RC) is to resolve the two convolved pieces of the speech, $e(n)$ and $\theta(n)$, into two additive components, that can then be separated or analyzed

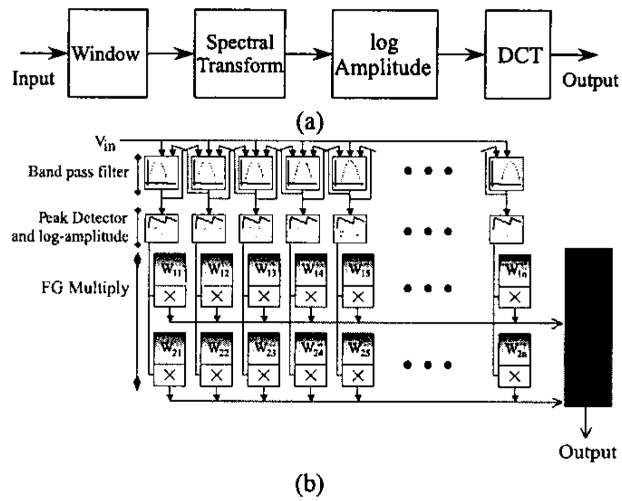

Fig. 1. a.) The traditional cepstrum computation which is performed in digital circuitry. b.) Floating-gate system to perform cepstrum front-end computation for speech processing systems. The system contains 32 frequency taps that can be spaced arbitrarily by programming the corner frequencies for the bandpass filter banks. The peakdetectors provide a power spectrum of the input signal for any given time slice.

using spectral (cepstral) analysis. The real cepstrum is usually defined as

$$\begin{aligned} c_s(n) &= DTFT^{-1}\{\log|DTFT\{s(n)\}|\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(\omega)|e^{j\omega n}d\omega \end{aligned} \tag{2}$$

By substituting $S(\omega) = E(\omega)\Theta(\omega)$ we obtain

$$c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|E(\omega)\Theta(\omega)|e^{j\omega n}d\omega \tag{3}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} (\log|E(\omega)| + \log|\Theta(\omega)|)\, e^{j\omega n}d\omega \tag{4}$$

$$= c_e(n) + c_\theta(n). \tag{5}$$

In general, the spectrum $\Theta(\omega)$ of the vocal tract response is assumed to be smooth with only slow changes as a function

*This work was partially supported by grants National Science Foundation (CISE-1068549, ECS (CAREER): 0093915, ECS-9988905) and by corportate donations to the Georgia Tech Analog Consortium by Texas Instruments and Motorola, Inc.

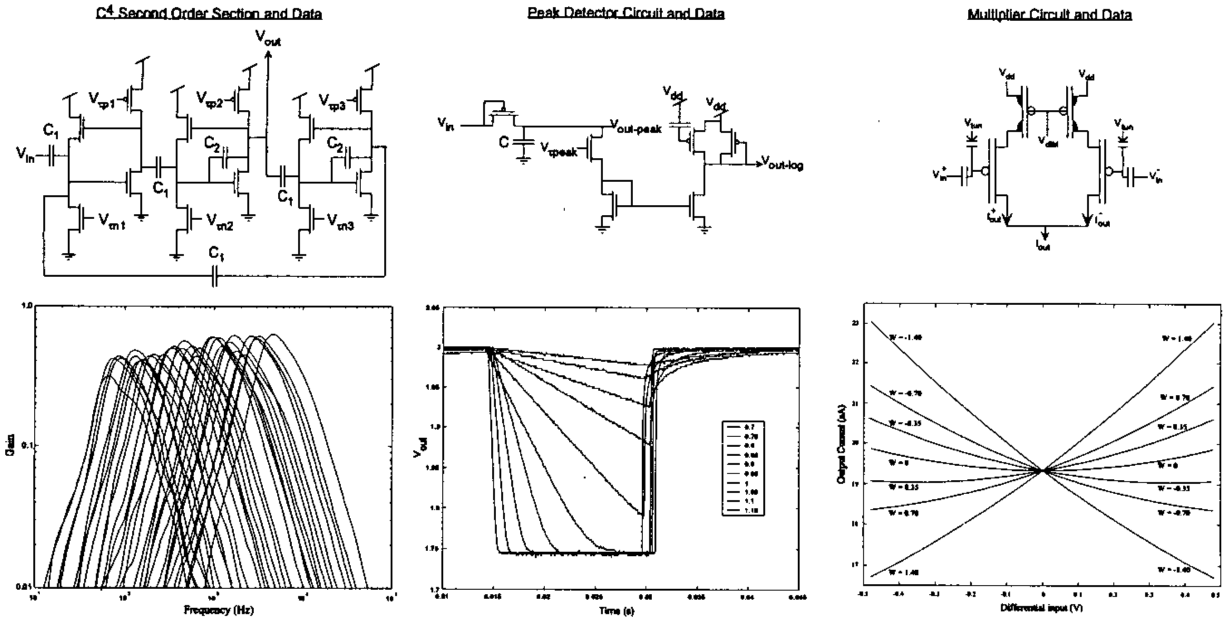

Fig. 2. a.) Floating-gate C^4 second-order-section and its corresponding frequency response. The high and low corner frequencies can be independently tuned for each filter bank. Arbitrarily programmable corner frequencies allow these filters to be spaced linearly, octave, logarithmically or any other values desired by the user. b.) Differential floating-gate multiplier structures multiply two differential signals by constant factors that are stored on the floating gate elements. c.) Floating-gate peak detectors. The frequency response of the peak detector is controlled by a bias voltage which controls the gate of nFET M3. This element sets a constant resistance and the total R,C value shifts the high corner frequency. The frequency response is shown for different values of v_{tau}.

of frequency; the spectrum $E(\omega)$ of the excitation signal is assumed to have rapid variations as a function of frequency. When the inverse discrete-time Fourier transform is taken in Eq. (1) or Eq. (5) the energy from $\log|E(\omega)|$ will be mapped to high values of n and the energy from $\log|\Theta(\omega)|$ will be mapped to low values of n. Therefore, the two signals, $c_e(n)$ and $c_\theta(n)$, will largely occupy different parts of the frequency axis and can be analyzed as separate entities.

1.1. Cepstrum Implementation

Equations 1-5 illustrate the concept of using cepstral analysis to decompose a signal into more basic components. However, speech signals are non-stationary; therefore, the time-frequency spectrum of speech signals is not constant. In practice, short-term Fourier analysis is performed using the discrete Fourier transform (DFT) and inverse DFT respectively.

1.2. Mel-Cepstrum

A common variation on the real cepstrum is the Mel-cepstrum. Implemented in the discrete domain, the Mel-cepstrum is calculated by combining the output of the $\log|S(\omega)|$ into critical band energies and then performing the discrete cosine transform (DCT) on the sequence of critical band energies [1].

The Mel-cepstrum loses the strict interpretation as a deconvolution tool but it has several qualities that make it useful for speech recognition. First, the critical band filtering stage produces an output that is very similar to that observed at early stages of human auditory perception. Second, the final DCT serves to decorrelate the critical band energies for improved automatic pattern recognition performance. The DCT also has the advantage over the Fourier transform that it yields real values.

2. ANALOG CEPSTRUM

Large arrays of floating gate computational blocks have been developed that perform similar processing as mel-cepstrum algorithm. The processing is all done in continuous analog circuitry. By keeping the signals as purely analog signals we are able to neglect noise that is introduced during conversion and also process higher order properties of the signals. One example of such higher order properties would be the high frequency information that is present in speech signals during fricatives and or dipthongs. The spectral content for these transitive sounds appear as a spike in the time domain which translates into a broad spectrum sig-

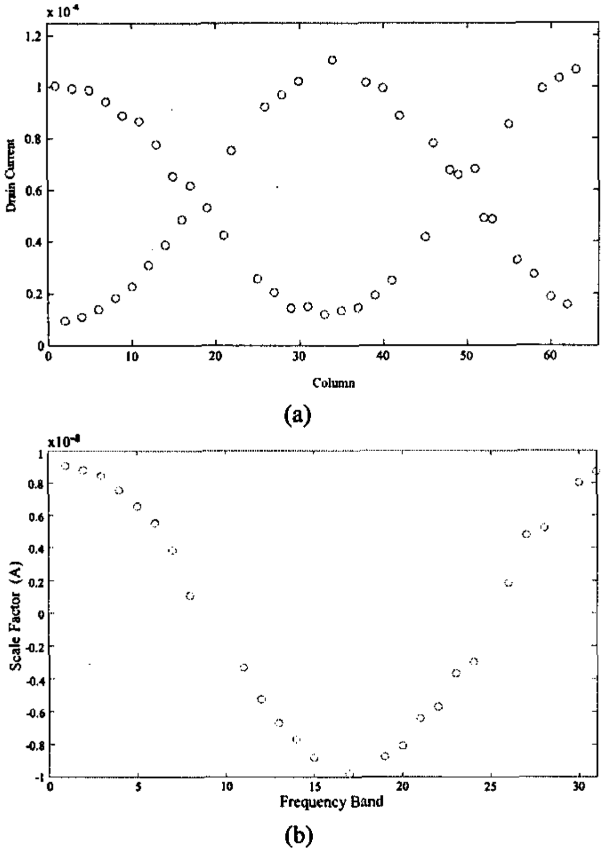

Fig. 3. a.) Floating-gate multiplier with programmed weights to a single cosine period. b.) The resulting scale factor as applied to a single-ended input vector.

nal in the frequency domain. The peak detectors are not all programmed at the same cut off frequency, and are therefore able to capture the high frequency transitions occurring in the time domain. This additional frequency information will result in a "richer" frequency spectrum and possibly better resolution of speech components containing sharp transitions such as *b, c, d,* and *t*.

2.1. Analog Frequency Decomposition

The basic building block of the cepstrum begins with a continuous spectrum decomposition similar to a Discrete-Fourier Transform (DFT). The spectrum decomposition is done using differential C^4 second-order-section bandpass filters [3]. For simplicity only one half of the differential structure is shown in Fig. 2. The spacing of the bandpass filters is arbitrary because each can be programmed to have a desired high-frequency corner and low-frequency corner [4]. Programming the C^4s is handled as if each filter were two floating gate elements. The entire row is viewed as a single row and the floating-gate elements for the high and low corner frequencies are accessed by column. Control circuitry guarantees injection isolation by latching the gate and drain voltages to the power supply for elements within C^4s not selected [5].

The magnitude of each spectrum passes through a peak detector stage to produce a constant magnitude output. This magnitude is similar to taking the power spectrum density or real spectrum of an input signal. At this point, phase information is unchanged, however the frequency response of the peak detectors must be programmed to it's respective frequency band. The circuit is shown in Fig. 2. The input diode has the following current relationship

$$I_D = I_0 e^{\frac{\kappa V in - V_2}{U_T}} \qquad (6)$$

This current $I = V_2/R$, where R is the total resistance of the bias transistor and diode. The floating-gate transistor on the output provides an offset current to set the DC output voltage. The output diode has the a logarithmic relationship to current.

$$\Delta V_{out} = -\frac{U_T}{\kappa} ln(\frac{I - I_{offset}}{I_0}) + v_{dd}. \qquad (7)$$

Each peak detector has an individually programmable corner frequency. Because the output magnitude is continuous, this allows us to capture additional high frequency content within each band. The peak detector programming blocks are isolated similarly to the C^4s. The entire bank is treated as a single row and within that row the individual elements are accessed by column. Control circuitry on the rows and columns ensures isolation.

2.2. Analog Floating-Gate Multiplier

The analog differential multiplier is shown in Fig. 2. The equation governing its operation is as follows:

$$\begin{aligned} I_{Out} &= I_{SO}(W^+ + W^-)cosh(\Delta V_{in}/V_y) \\ &\quad - I_{SO}(W^+ - W^-)sinh(\Delta V_{in}/V_y) \end{aligned} \qquad (8)$$

The linearized case including only the first order terms simplifies to:

$$I_{Out} = I_{SO}(W^+ + W^-) + I_{SO}(W^+ - W^-)(\Delta V_{in}/V_y)$$

Where the two weights W^+ and W^- are programmable floating gate voltage values. These values can be programmed to any arbitrary value, but for operations involving spectrum decomposition and transforms, these values are programmed to cosine scale factors. Their differential operation requires each pair to have a DC bias voltage and from there, the cosine values are scaled around this bias current.

Fig. 3 shows programmed cosine values for a single row of multipliers. Each row acts as a 32-tap DCT basis vector. Similar vectors are programmed into each row and typically they will be set to produce a wavelet type of decomposition by scaling each basis vector frequency by a constant factor.

2.3. Relation of the Analog Mel-Cepstrum to the Mel-Cepstrum

The mel-cepstrum, as used in digital signal processing (DSP) is based on a signal sampled in time and in frequency. The analog cepstrum is an approximation to the mel-cepstrum or cepstrum (depending on how the filters are defined) in which frequency is sampled but time is not. The output of each filter contains information similar to the short-time Fourier transform and can likewise be assumed to represent the product of the excitation and vocal-tract within that filter band. The primary difference here is that the DSP mel-cepstrum approximates the critical band log frequency analysis of the human ear by combining DFT bands while the analog system actually performs a critical band-like analysis on the input signal. Thus higher frequency critical band energies are effectively computed using shorter basis functions than the lower frequency bands. This is more in agreement with analysis in the human auditory system and is better suited to identifying transients.

One other difference between the analog cepstrum described herein and the real cepstrum described in Eq. (1) is that the magnitude function (inside the log) is estimated using a peak detector rather than using the true magnitude of the complex spectrum.

3. CONCLUSION

Chips performing an analog cepstrum have been fabricated through MOSIS. Initial results show that the analog cepstrum will be useful components within speech recognition systems. The full system is currently being tested with preliminary results are shown in Fig. 4 The system output from the analog peak detector was processed using simulated multiplier cells in matlab.

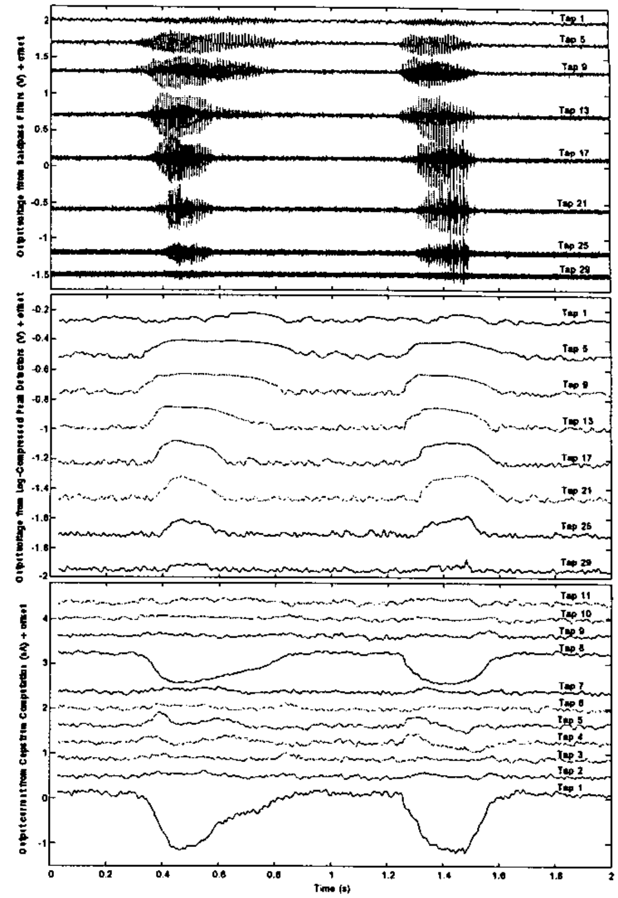

Fig. 4. Cepstrum system output. The system input is a sequence of speech using a standard speech database. There are 12 continuous cepstrum coefficients calculated for this section of speech and more coefficients is only a matter of chip area since the calculation is performed in parallel analog circuits. From the graph one can see the two distinct periods of speech.

4. REFERENCES

[1] John R. Deller, John G. Proakis, and John H. L. Hansen, *Discrete-time Processing of Speech Signals*, Institute of Electrical and Electronics Engineers Press, New York, 2000.

[2] Alan V. Oppenheim and Ronald W. Schafer, *Digital Signal Processing*, Prentice-Hall, New Jersey, 1975.

[3] David W. Graham and Paul Hasler, "Capacitively-coupled current conveyer second-order section for continuous-time bandpass filtering and cochlea modeling," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2002.

[4] M. Kucic, A. Low, J. Neff, and P. Hasler, "A programmable continuous-time floating-gate fourier processor," *IEEE Journal of Circuits and Systems II*, January 2001.

[5] M. Kucic, J. Dugger, P. Hasler, and D. Anderson, "Programmable and adaptive analog filters using arrays of floating-gate circuits," in *Proceedings of the 21st Conference on Advanced Research in VLSI*, Atlanta, GA, March 2001, pp. 148 –162.