Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

R. Timothy Edwards

Johns Hopkins University
Department of Electrical and Computer Engineering

March 26, 1999

Advisor: Dr. Gert Cauwenberghs
Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion
**Research Goals:**

Investigate signal processing systems for which primary constraints are **power consumption** and **physical area**.

- Develop these systems into efficient analog and digital hardware
- Maintain high accuracy with respect to microprocessor and DSP solutions
- Use circuits with ultra-low (µW) power consumption
- Exploit parallel architectures to achieve robustness in the presence of noise and mismatch

**Performance depends critically on the choice of algorithm and hardware.**

**Applications:**

- Handheld or mobile (battery-operated) environments
- Can be used to monitor environment and wake up systems in “standby” mode
- “Smart” A/D conversion

![Diagram of sensor, micropower frontend processor, and computer or DSP on standby]
Microelectronic Systems Laboratory Research: Systems and Algorithms

- Continuous wavelet transform (CWT) processor
- Log-domain filters for audio-frequency applications
- Micropower mixed-mode template correlation acoustic pattern classifier
- Model-free adaptive correction of optical aberrations
Bandpass Filterbank vs. Complex Demodulation

\[ X(j\omega) \]

\[ H_{BP}(j\omega) \]

\[ H_{LP}(j\omega) \]

\[ X(j\omega) H_{BP}(j\omega) \]

\[ (X(j\omega) \ast \delta(j\omega - j\omega_k)) H_{LP}(j\omega) \]
Tilings of the Time-Frequency Plane

Area of uncertainty in time and frequency: $\Delta f \Delta t \geq \frac{1}{2}$

- (a) FFT, short window: Gives good time resolution but poor frequency resolution
- (b) FFT, long window: Gives good frequency resolution but poor time resolution
- (c) Wavelet Basis: Most efficient time/frequency tradeoff for arbitrary inputs
- (d) Wavelet Packets: The most efficient description of a specific signal; difficult to compute on the fly
Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion
Acoustic Event Classification Using Correlation

**Goal:** Try to classify an acoustic event by matching the time-frequency decomposed input against a set of templates. Each template has $M \times N$ values and represents a “typical” example of one class.

Signal ($x$) characteristics (output of frontend filtering system):
- $M$ frequency channels (log-spacing)
- Continuous-time (or Discrete-time)
- Continuous-valued
- Normalized energy envelope

Template ($p_z$) characteristics:
- One template per class ($z$)
- $M$ frequency channels (log-spacing)
- $N$ time-bins (1 ms per bin for transients)

Basic Correlation equation:
\[
c_z[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x[t - n, m] p_z[n, m]
\]
Acoustic Event Recognition: Example

**Inputs: acoustic data**

![Amplitude vs time graph showing Dink, Clap, Tub, and Can with corresponding time and frequency channels.]

**Frontend processing**

![Frequency channel norms across time for Dink, Clap, Tub, and Can.]

**Classification via template correlation**

Dink, Clap, Tub, Can
Acoustic Signal Classification

Three main components:

1. **Filterbank**: Time-frequency decomposition converts the input signal into an efficient representation for subsequent processing.

2. **Template Correlator**: Simple correlation is sufficient to recognize complicated acoustic events.

3. **Digital Post-processing**: Uses correlator outputs as features for tasks such as speech recognition.

This talk will focus on current research (parts 1 and 2) and conclude with comments about (3).
Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion
Filterbank Frontend Systems

- Cochlear-model filters:
  - Mead and Lyon (1989)
  - ... and many others

- Unnikrishnan, Hopfield, and Tank (1991)
- Shihab Shamma (1994)

Our frontend architecture: two versions

![Diagram of two versions of the frontend architecture](attachment:image.png)
Bandpass Filterbank Frontend Processor

System characteristics:

- Parallel bank of 2nd- or higher order bandpass filters with constant $Q$
- Logarithmically-spaced center frequencies across the bank
- Rectification and smoothing of output to get energy envelope of signal
- Automatic gain control through $L$-$1$ normalization across outputs (or other method)
A BiCMOS Current-Mode Filterbank Frontend

System implementation:

- Log-domain filters* based on translinear circuits
- BiCMOS design for greatest dynamic range and linearity

Why a Current-Mode VLSI Implementation?

- Large dynamic range compared to voltage-mode circuits
- Dense circuit layouts
- Filter parameters \((f_c, Q)\) are tunable over multiple decades
- \(\mu\text{W}-\text{range} power consumption: good linearity at low power
- Most convenient format to interface to current-mode backend processors

The Log-domain Bandpass Filter

\[ I_{out} = I_{in} \left( \frac{\tau s}{1 + (1/Q)\tau s + \tau^2 s^2} \right) + I_{DC} \quad \text{where} \quad \tau = \left( \frac{V_tC_1}{I_f} \right) \quad \text{and} \quad Q = e^{V_Q/V_t} \]
The Envelope (Peak) Detector

Simple Op-Amp Peak Detector

![Op-Amp Diagram]

Behavior of a Peak-to-Peak Detector

\[ I_{\text{out}} = I_{\text{max}} - I_{\text{min}} \]
The Envelope Detector

Maximum-peak Finding Circuit

Minimum-peak Finding Circuit

$I_{\text{max}}$

$I_{\text{in}}$

$(I_{\text{out}} \text{ from } \text{bandpass})$

$(V_{\text{out}} \text{ from } \text{bandpass})$

$V_{\text{leak}}$

$V_{\text{ss}}$

$V_{\text{dd}}$

$M_1$

$M_2$

$M_3$

$C_1$

$C_2$

$M_4$

$M_5$

$M_6$

$M_7$

$M_8$

$M_9$

$M_{10}$

$I_{\text{in}}$

$Q_1$

$Q_2$

$Q_3$

$V_{\text{min}}$

$V_{\text{max}}$

$V_{\text{fbp}}$

$V_{\text{fbn}}$

$V_{\text{in}}$

$I_{\text{in}}$

$I_{\text{max}}$

$I_{\text{min}}$

$V_{\text{out}}$
The fifteen-channel bandpass filterbank fabricated in 1.2 µm technology inside a 2.2 mm × 2.2 mm padframe.
Measured results of the Log-domain filters

Measured frequency response of one filter (4th-order bandpass) measured at three different center frequencies and three different $Q$-values.

Measured output of the envelope detector of all filter outputs, showing mismatch between channels.
Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion
The Acoustic Transient Processor Correlation Algorithm

\[ c_z[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x[t - n, m] p_z[n, m] \]
Original Software Trials of the Transient Classifier
Confusion Matrix for Cross-Validation Test
Baseline Algorithm Results

<table>
<thead>
<tr>
<th>Event</th>
<th>Bar</th>
<th>Book</th>
<th>Can</th>
<th>Dink</th>
<th>Door</th>
<th>Finger</th>
<th>Hand</th>
<th>Mallet</th>
<th>Shelf</th>
<th>Tub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Book</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Can</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dink</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Door</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hand</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mallet</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Shelf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tub</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>0</td>
</tr>
</tbody>
</table>

Statistics:

Total instances presented: 222
Correct: 214
Incorrect: 8
Accuracy: 96.4%
Template Correlation: Computational Requirements

- $M = 32$ frequency channels
- timestep = 1 ms
- $N = 100$ bins (timespan of 1/10 second)
- one correlation per timestep

>3 million multiply-accumulate operations per second per template

Implementation goals:

- minimize operations per timestep (modify algorithm)
- minimize power per operation (analog design)
### Different Correlation Architectures

<table>
<thead>
<tr>
<th>Method</th>
<th>Both Cont.</th>
<th>Binary Input</th>
<th>Both Binary</th>
<th>Binary (1, 1) Template</th>
<th>Binary (1, 0) Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-to-One</td>
<td>96.40%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time Difference</td>
<td>85.59%</td>
<td>65.32%</td>
<td>59.46%</td>
<td>82.43%</td>
<td>81.98%</td>
</tr>
<tr>
<td>Channel Difference</td>
<td>90.54%</td>
<td>53.60%</td>
<td>95.05%</td>
<td>94.59%</td>
<td>94.14%</td>
</tr>
<tr>
<td>Center-Surround</td>
<td>92.79%</td>
<td>53.60%</td>
<td>95.05%</td>
<td>92.34%</td>
<td>92.34%</td>
</tr>
</tbody>
</table>

**Conclusion:**
Best architectures use channel difference computation and binary template values and either binary or continuous-valued input.
Software Trials of the Transient Classifier
Confusion Matrix for Cross-Validation Test
Channel-differenced input and binary-valued template

<table>
<thead>
<tr>
<th>Event</th>
<th>Bar</th>
<th>Book</th>
<th>Can</th>
<th>Dink</th>
<th>Door</th>
<th>Finger</th>
<th>Hand</th>
<th>Mallet</th>
<th>Shelf</th>
<th>Tub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Book</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Can</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dink</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Door</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hand</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mallet</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Shelf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tub</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
</tr>
</tbody>
</table>

Statistics:

Total instances presented: 221
Correct: 205
Incorrect: 16
Accuracy: 92.7%
An Efficient Architecture for the Acoustic Transient Classifier

\[
c_z[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} (x[t - n, m] - x[t - n, m - 1]) p'_z[n, m]
\]

\[
p'_z[n, m] = \begin{cases} 
1 & (p_z[n, m] - p_z[n, m - 1]) > 0 \\
0 & (p_z[n, m] - p_z[n, m - 1]) \leq 0 
\end{cases}
\]
ATP Chip Architecture—Static Memory Core Cell

\[ \sum_{m} x(m,t) p'(n,m) \]
\[ q_1(n,t) = \sum_{m} x(m,t) p'(n,m) \]
\[ q_2(n,t) = \sum_{m} x(m,t) p'(n,m-1) \]

Delay line 1
\[ + \]
\[ \Sigma \]
\[ - \]
\[ output \ c_2(t) \]

\[ V_{dd} \]
\[ \Phi \]
\[ M_1 \]
\[ M_2 \]
\[ M_3 \]
\[ M_4 \]
\[ M_5 \]
\[ \bar{\Phi} \]

\[ bit(n) \]
\[ p'(n,m-1) \]
\[ \bar{bit}(n) \]

input \( y(0,t) \)
\[ \ldots \]
input \( y(M,t) \)
ATP—Complete Chip Architecture

![Diagram of ATP architecture](image)

- **Inputs**: V, I
- **Row Select**: A0, A1, A2, A3
- **Column Select**: bit0, bit1, bit2
- **64 x 16 memory array**: 1-bit template cell
- **Bucket Drivers**: V_{casc}, V_{in}, V_{out}
- **Top Bucket Brigade**: S1, C1, V_{ref} = 1/2 Vdd
- **Bottom Bucket Brigade**: S2, C2

**Accum. Pulse** (φ3b)

**Write**
The acoustic transient correlator, a 2.2 mm × 2.2 mm die fabricated in a 1.2µm CMOS technology.
Simple template (left) detects the transient “can” (bottom left) but rejects the transient “snap” (bottom right).

Blue: numerical simulation.
Green: measured chip response.
Black: Residual error.

Chip measured power consumption: 50 µW
Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion