WIV-based periodicity analysis

Or: A new way to look at "pitch"

Musical Pitch:

You know it when you hear it, but nobody has figured out how the ear does it.

Why?

Because everybody has been using the wrong technical approach...frequency spectrum analysis

Some background:

Around 160 years ago Ohm and von Helmholtz, using experiments based on pure tones proposed Fourier spectrum analysis to explain how the ear hears different pitches.

In 1842 Seebeck, testing impulsive waveforms, showed valid objections to spectrum theory.

Seebeck was squelched by Ohm and Helmholtz.

Since then, spectrum analysis has remained the standard auditory model...without anyone sitting down trying to understand the totality of the ear's purpose and the deficiencies of the Fourier model.

Doubts about spectrum methods resurfaced with Schouten in 1940* as a result of his psychoacoustic experiments on pitch perception that baffled him...and continue to baffle auditory scientists.

* J. F. Schouten, *The residue, a new component in subjective sound analysis*, Natuurkundig Laboratorium der N.V. Philips Gloelampenfabrieken, Eindhoven, Holland, Feb. 24, 1940s

Here is *my* bafflement:

Why has the auditory community (even Schouten) continued to hold to a theory that is widely known* to be wrong?

*Gabor 1948, based on Heisenberg's uncertainty principle

A syllogism to ponder:

Currently, the ear is thought to be a spectrum analyzer.

Spectrum analyzers can't analyze impulses instantaneously.

Animal's ears hear impulses instantaneously.

If an animal can't hear impulsive sounds it won't survive.

Therefore, ears must not use spectrum-type processing.

And here's another (in)famous unsolved auditory problem: *the cocktail party effect.*

Simply put; it's the ability to hear everything and listen to anything.

It's another evolutionary requirement: All animals have it.

But the 'standard auditory model' can't ever do it.

It should now be obvious:

The ear uses a method that is <u>not</u> based on spectrum analysis.

A Soliloquy

Over years of pondering these paradoxical questions
while I designed radar intelligence intercept systems
it became apparent that,
if acoustic waveforms were treated as impulses
(something like radar pulse intercepts,)
it would be possible
to resolve the auditory paradox.

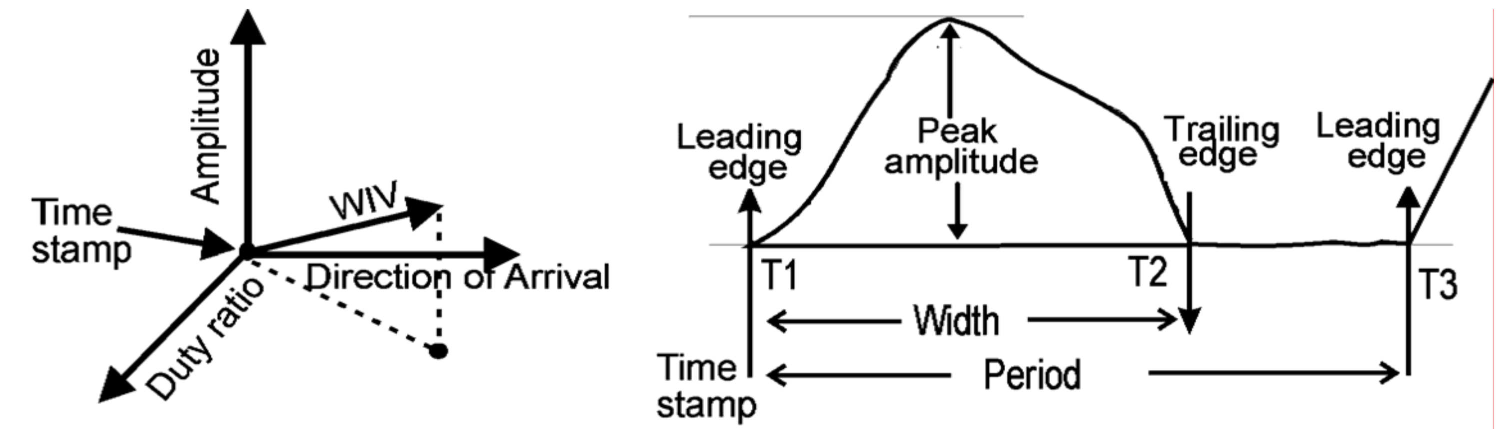
These thoughts led me to develop a particle-like* 'waveform information vector,' the WIV.

The WIV has the time, space, and energy information that makes it possible to recognize to separate, locate, and understand <u>meanings</u> of the sounds in the auditory scene.

*a concept proposed by Epicurus, 300 B.C.

Impulse processing based on WIVs solves crucial practical difficult problems of amplitude dynamic range and direction of arrival within intermixed sound sources.

For theorists, it circumvents the Gabor time frequency ambiguity that denies Fourier the ability to deal with the requirements for synchronous real-time source recognition, location, and separation that are key to explaining the cocktail party effect.



(a) Waveform information vector

(b) Derivation of WIV parameters
duty ratio = width/period

As a fundamental unit of acoustic information the WIV contains elements of time, space, and energy within an interval bounded by the real zeros of the waveform.

The many meanings in the WIV

WIVs encode waveform energy, location, and time.

Waveshapes have characteristic timbres both as individual impulses or in sequential combinations.

Timbres can have various meanings that depend on circumstances in time and space.

Example: narrow, sharp-edged impulses imply irritating or threatening sounds, while symmetrical, smooth shaped impulses are attractive.

Thus, WIVs posess the elementary meanings about the sounds and their locations, and the kinds of sources in the auditory scene.

With these elementary meanings, it is possible to select and respond instantaneously to threatening sounds from within the scene's background.

Elements of a workable system

The WIV deals with the auditory problems that present technology hasn't solved. For example:

- Classifying WIVs according to <u>elementary</u> meanings
- Separating mixed sounds by their meanings
- Selecting sounds for attention
- Handling effects of amplitude dynamic range
- Operating in real time, synchronized with time of arrival

Main features of WIV-based auditory model:

- Patterns detected instantly at a <u>common time reference</u> as each WIV arrives into the system
- Assures real-time synchronization with all sensory modes
- Based entirely on selectionist pattern recognition*****
- Patterns are learned in real time and augmented with experience*
- Patterns are recognized in hierarchies of sequences of WIVs, thus creating efficient and meaningful real-time processing
- Boolean logic and delay-line methods detect patterns
- Consistent with both neural processing and LSI implementation
- Does <u>not</u> require computing mathematical algorithms

***** For example, Gerald M. Edelman, *A universe of consciousness*, Chapter 7 *Selectionism,* Basic Books, New York, 2000

Evolution of the ear

Auditory processing evolved in steps of specific kinds of patterns in WIV sequences.

Sequence steps have time durations that correspond with familiar, meaningful auditory symbols.

A hierarchy of meanings derived from impulse sequences:

Impulse timbre

Tonal timbre (includes pitch)

Atonal timbre (includes phonetic timbre)

prosody

phonetic classes

phonemes

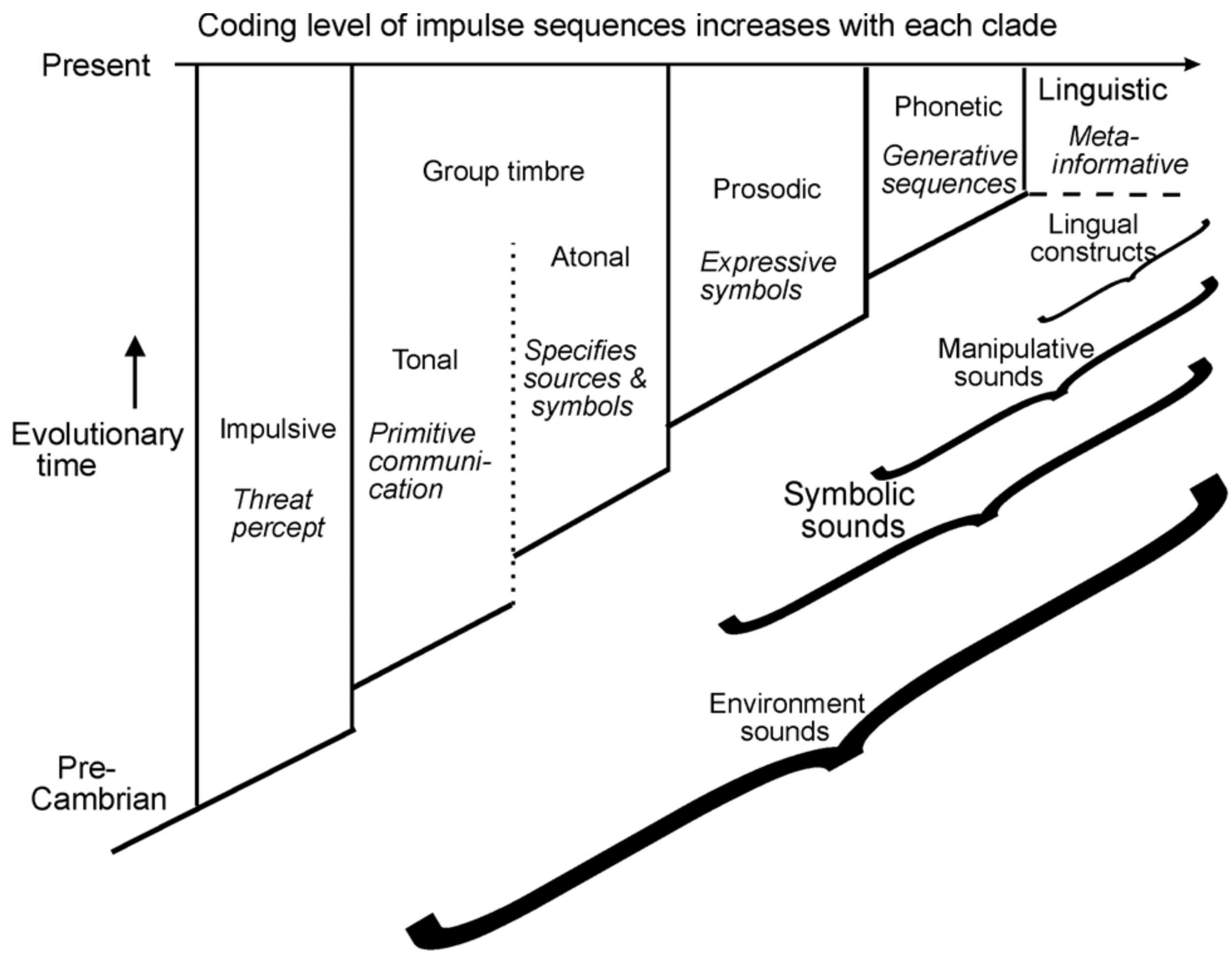
syllables

words

etc.,etc = language

It's interesting that each step in the hierarchy corresponds with epochs in the evolution of animal phyla.

The progression of steps indicates that improvement of auditory capability has enabled corresponding advancements in environmental success.



Cladogram illustrates the evolution of auditory perception

All of these factors are embedded in my auditory model

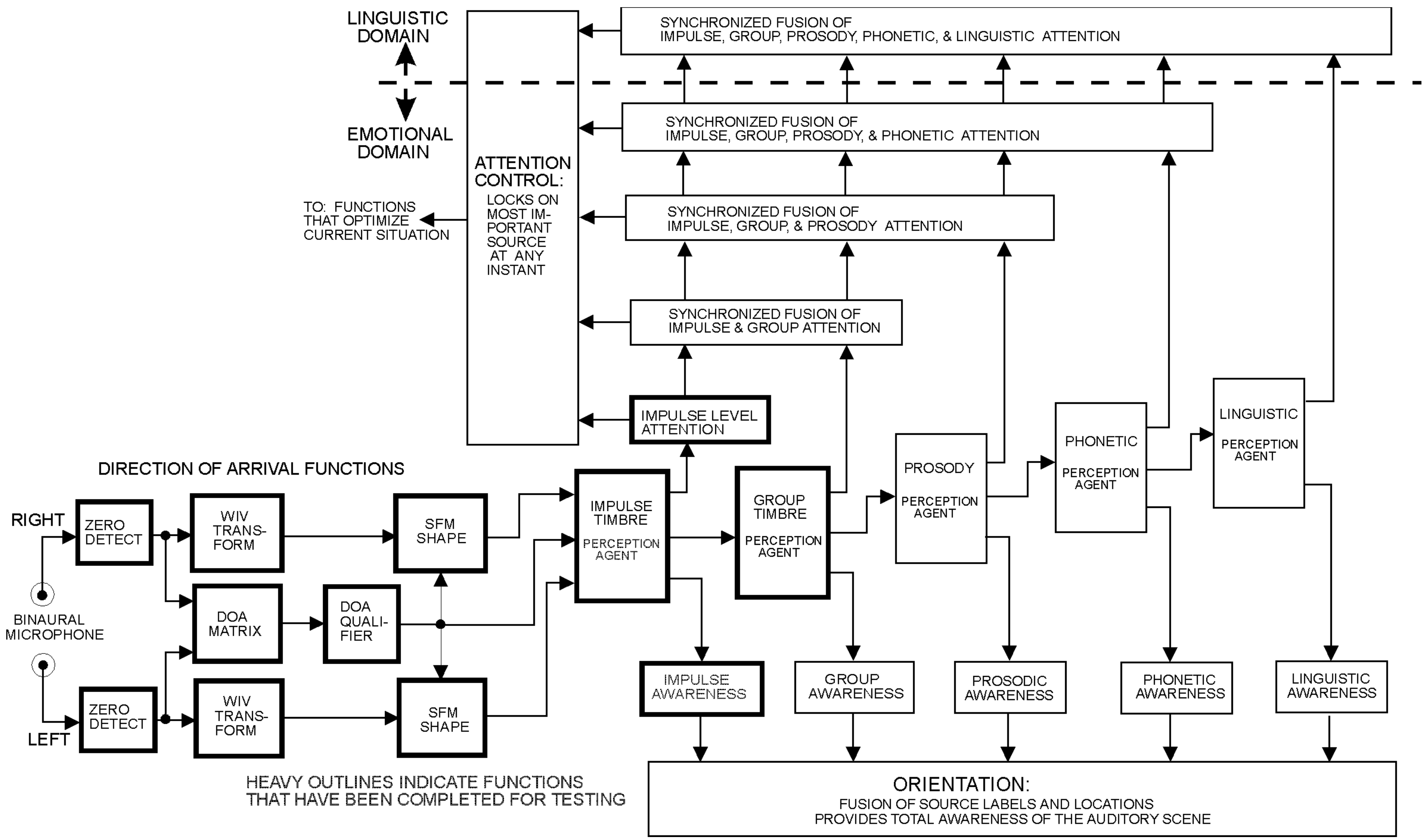
- Its configuration follows the evolutionary cladogram:

- Each clade recognizes the patterns and meanings at its level.

- Pattern recognition runs together with each clade so that their information products can be fused and used instantaneously as soon as they are available.

- Clade information is evaluated in a decision-making stage that focuses attention on the most interesting source.

Eventual (long-term) implementation of the WIV-based computational auditory model

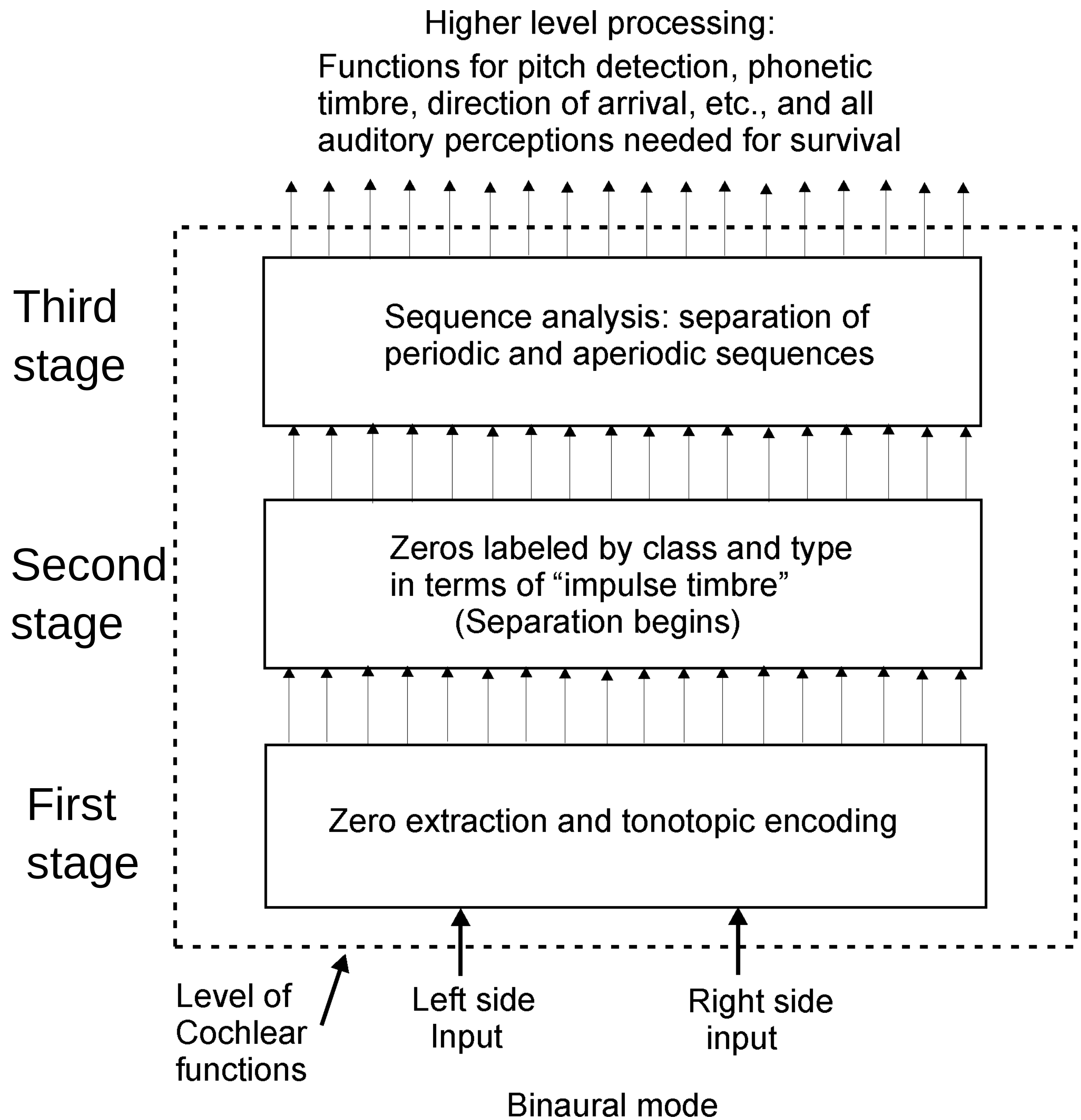


Blocks with heavy outlines denote functions that have been successfully demonstrated
I will call them the "***proto-cochlear functions.***"

Now consider entry-level functions:
the *protocochlea*.

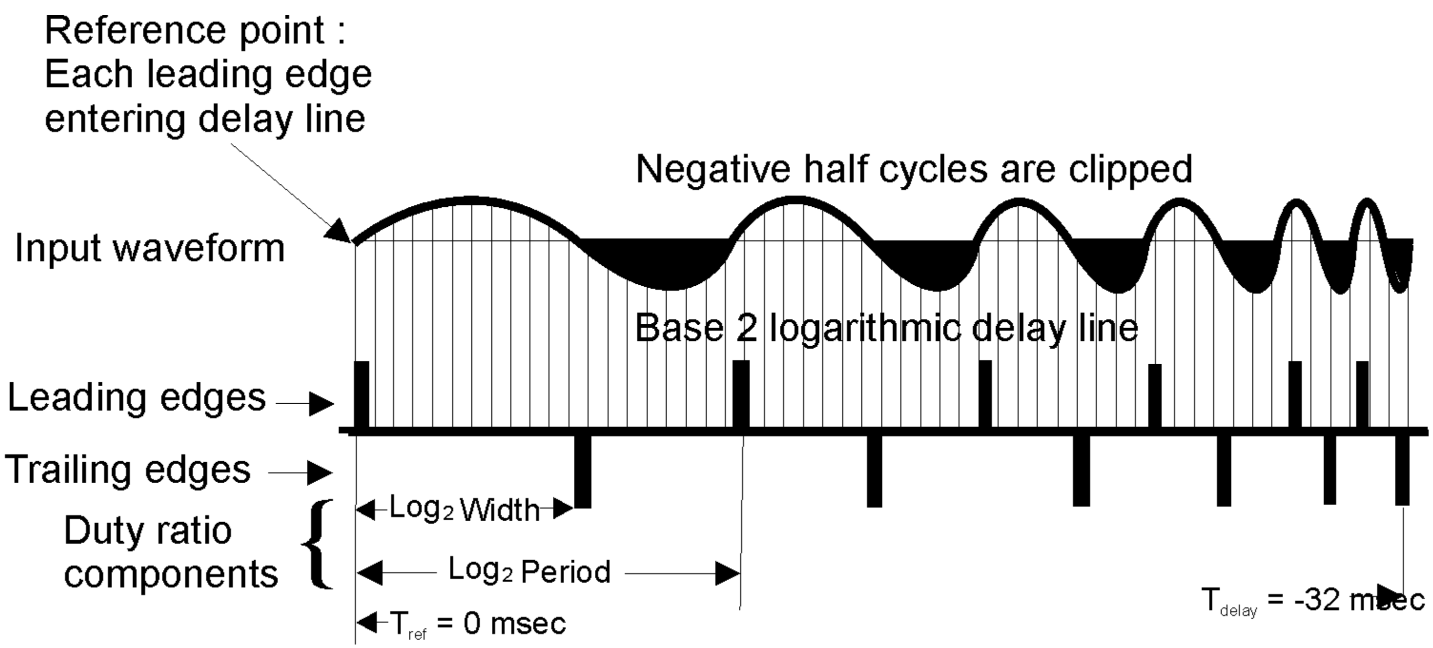
The protocochlea encodes the waveform *analogous* to the ear's cochlea and the cochlear nucleus.

Converts the acoustic waveform into spikes amenable with logical processing.
(Think LSI logic arrays!)



Basic functions of the proto cochlea

First stage of the proto cochlea:
Creates the WIV-to-Zero extraction and tonotopic encoding



0

Second stage of protocochlea functions**:**

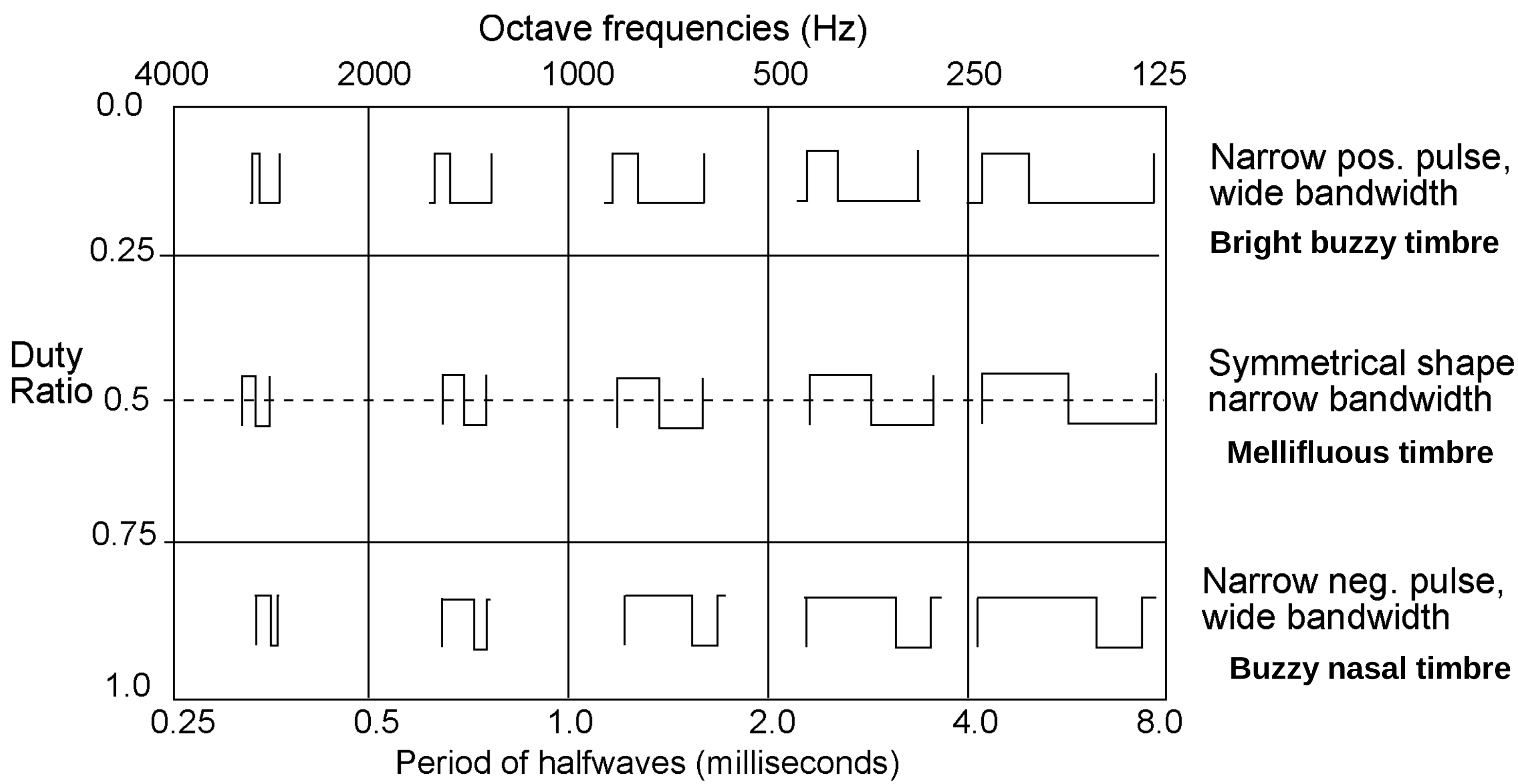
Attaching meaning to each WIV

Impulse timbre is labeled for meaning:

- according to the way a single WIV impulse sounds to listeners, or
- emotional implications
- according to a particular physical characteristic, based on the use of the system

  (for example, in geophysics or medical tomography.)

Illustrating the SFM's auditory waveform classification method



SFM encodes WIVs according to the phonetic class number of its location in the matrix.

Class No.	Phone label
1	/ih/
2	/aa/
3	/rr/
4	/ah/
5	/aw/
6	/uh/
7	/uu/
8	/l l/
9	/oo/
10	/od/
11	/ud/
12	/ad/
13	/md/
14	/nd/
15	/shwa/

These phones (10–14) can have dual perception modes both as individual WIVs and in atonal combinations of mixed front and center WIVs.

1 4 7 10 13 16

1	1 1 2 4	4 4 4	7 7 7	10 10 12	12 12 15
2	1 1 2 3	4 4 5	5 7 7	10 10 11	11 11 11
3	1 2 2 3	3 4 5	6 7 7	10 11 11	11 11 11
4	1 2 3 3	4 4 5	6 8 8	9 9 9	14 14 14
5	1 2 3 3	4 5 5	6 8 8	9 9 9	14 14 14
6	1 2 3 3	4 4 5	6 7 7	10 11 13	13 13 13
7	1 1 2 4	4 4 5	5 6 7	10 11 11	13 13 13
8	1 1 2 2	4 4 5	5 7 7	10 10 10	13 13 13

front | middle | back

The SFM thus encodes each halfwave of the waveform in terms of phonetic timbre.

Phonetic labels are not the only kind of meaningful encoding. Sounds also have emotional effects that depend on pitch and waveshape.

Emotional meaning: It begins in the proto cochlea.

Impulses = primal meaning = fear

Shape	Weight
Narrowest impulses = most threat =	4
Narrow symmetrical = less threat =	3
Wide impulses = less threat =	2
Wide symmetrical = least threat =	1

An experiment: simulating emotion in a "Cambrian" brain ,a brain only at the cochlear level

	1			4			7			10			13			16
1	4	4	4	4	3	3	3	3	2	2	2	2	1	1	1	1
2	4	4	4	3	3	3	3	2	2	2	2	2	1	1	1	1
3	4	4	3	3	3	3	2	2	2	2	1	1	1	1	1	1
4	3	3	3	3	2	2	2	2	2	1	1	1	1	1	1	1
5	3	3	3	3	2	2	2	2	2	1	1	1	1	1	1	1
6	4	4	3	3	3	3	2	2	2	2	1	1	1	1	1	1
7	4	4	4	3	3	3	3	2	2	2	2	2	1	1	1	1
8	4	4	4	4	3	3	3	3	2	2	2	2	1	1	1	1

Table of attention-weighting factors***** for each SFM cell

***** Weighting should be revised by psychoacoustic testing

Third stage: Sequence analysis:

Sequence analysis:

- Recognizes periodic (tonal) sequences
- Separates periodic from aperiodic sequences
- Recognizes atonal timbre patterns that can identify sources or phonetic components.

One of the biggest stumbling blocks in audition is "pitch" and pitch detection.

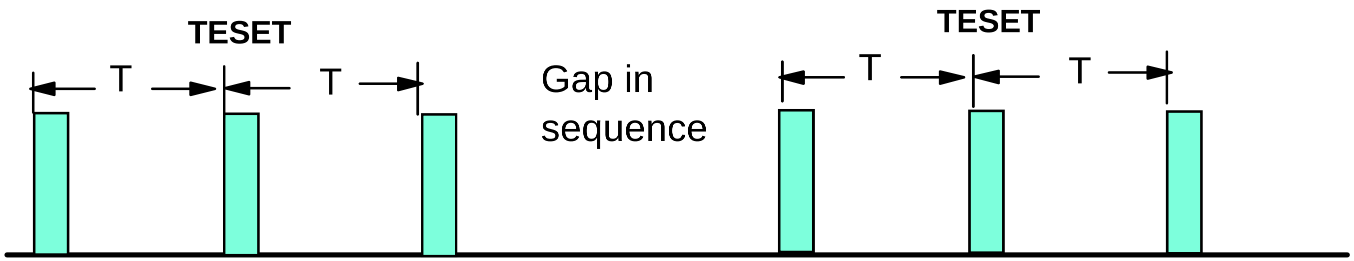
The key to the answer is to realize that pitch detection is:

- recognizing periodic sequences
- separating mixed sequences
- in real time

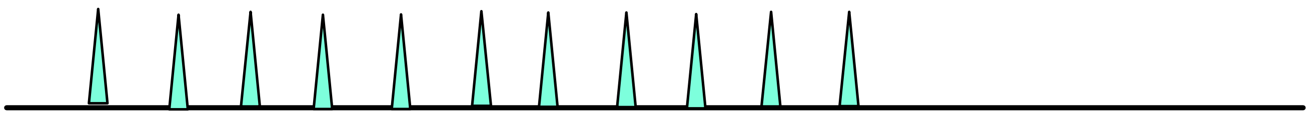
The periodicity sorting matrix (PSM) does these things. Spectral processing does not.

Properties of WIV-based sequences

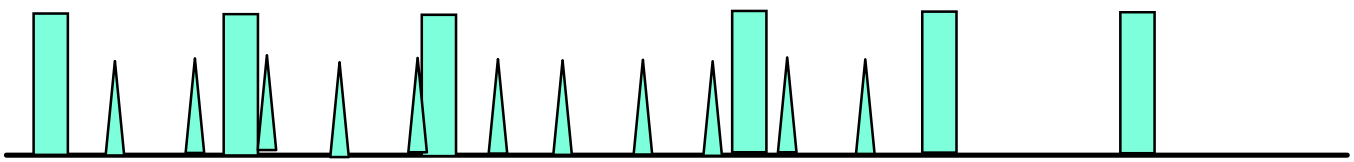
- Two sequential WIVs = three leading edges = one tonal event; a "teset" for **T**hree **E**qually **S**paced **E**vents in **T**ime
- Tesets should be treated as independent events.
- Uninterrupted sequences of similar tesets have a pure tonal quality.
- Interrupted sequences of similar tesets have a rough tonal quality or timbre.
- Sequences of uneven WIV spacing have atonal timbre, often with phonetic attributes, or sounds of sources.



Tesets as independent events...are heard as rough tonal sounds



Sequences of continuous tesets have pure tones



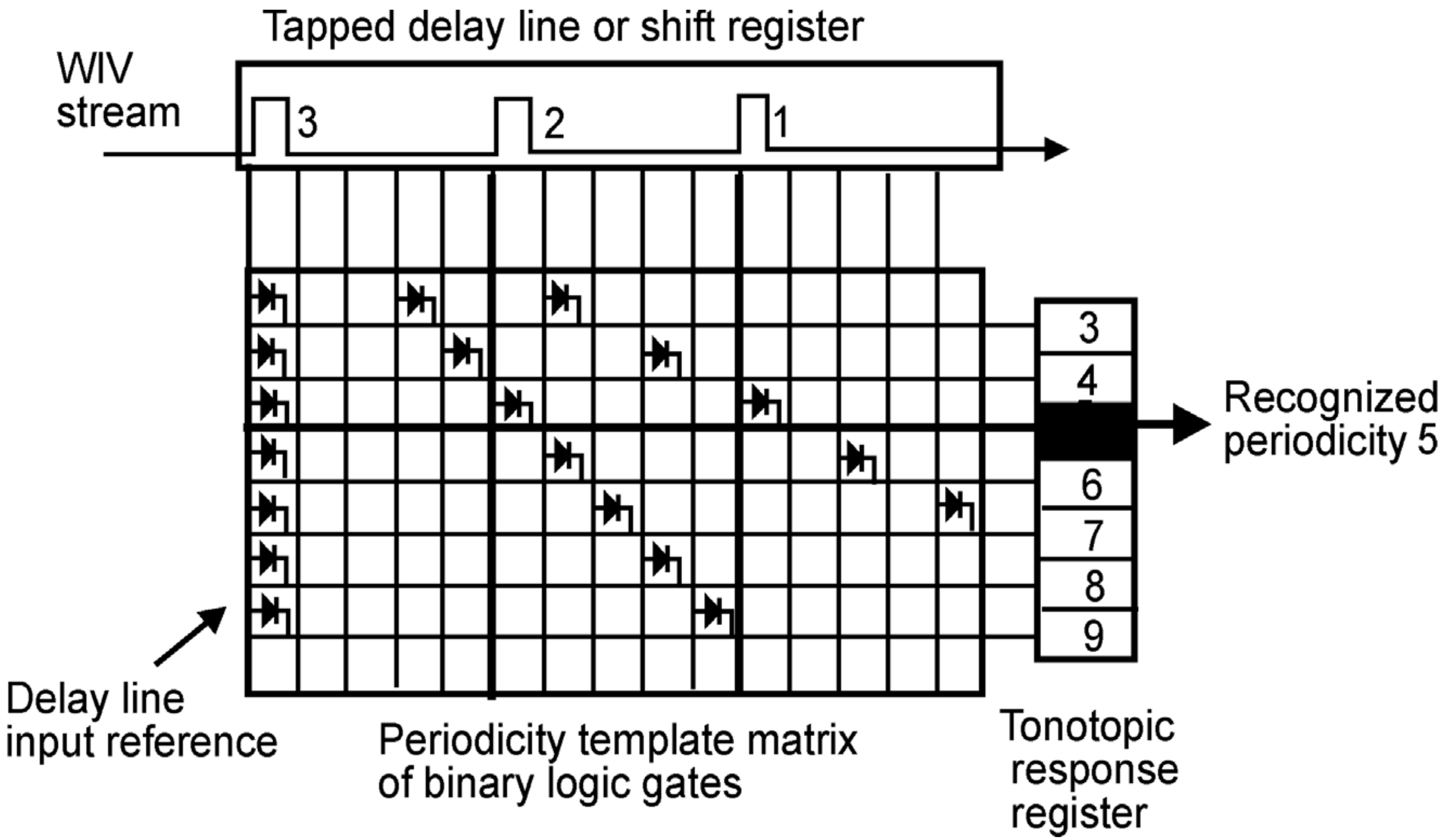
Mixed periodic sequences.

Two atonal sequences

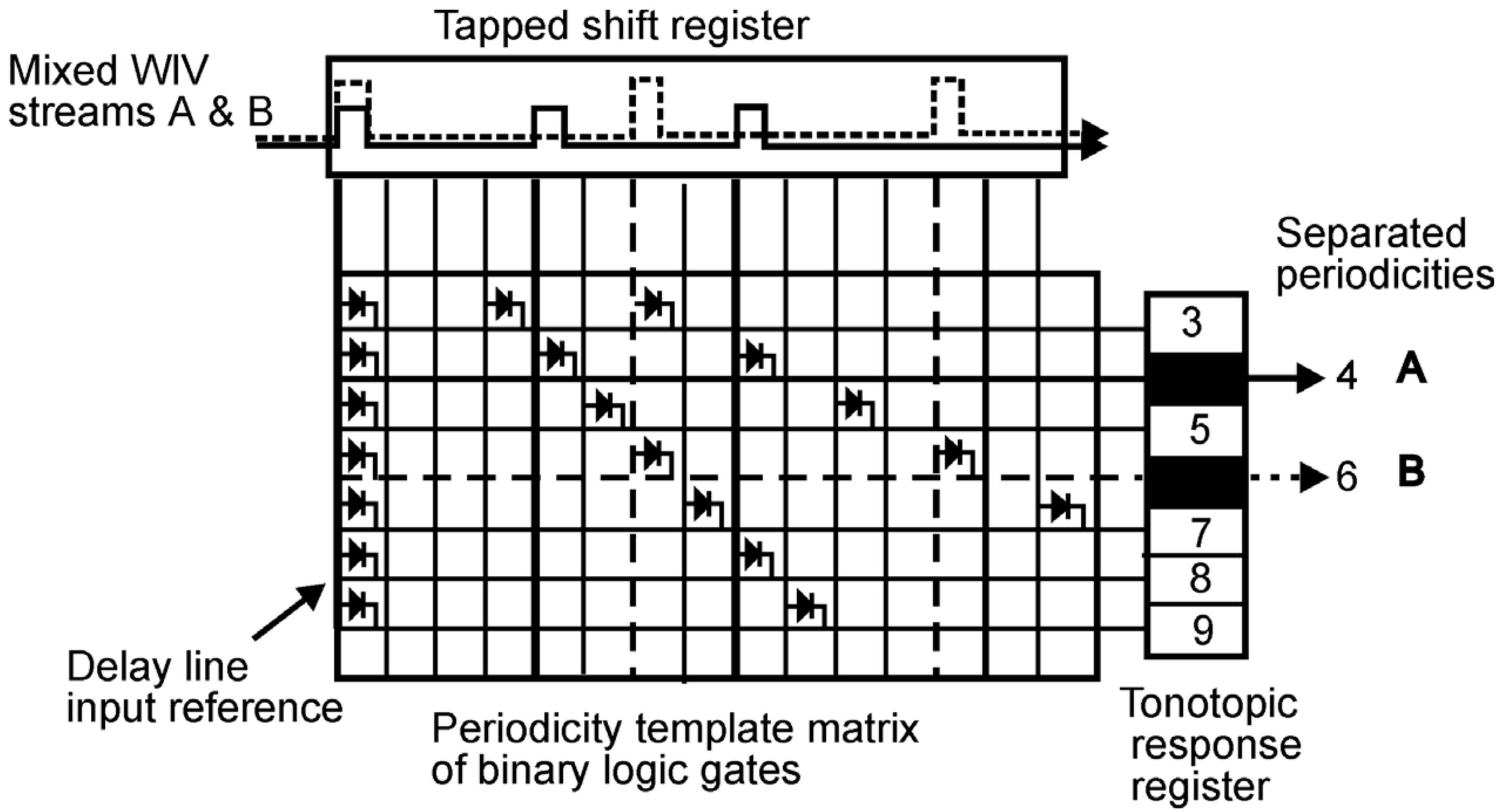
Illustrations of types of sequences

Recognizing tesets requires these functions:

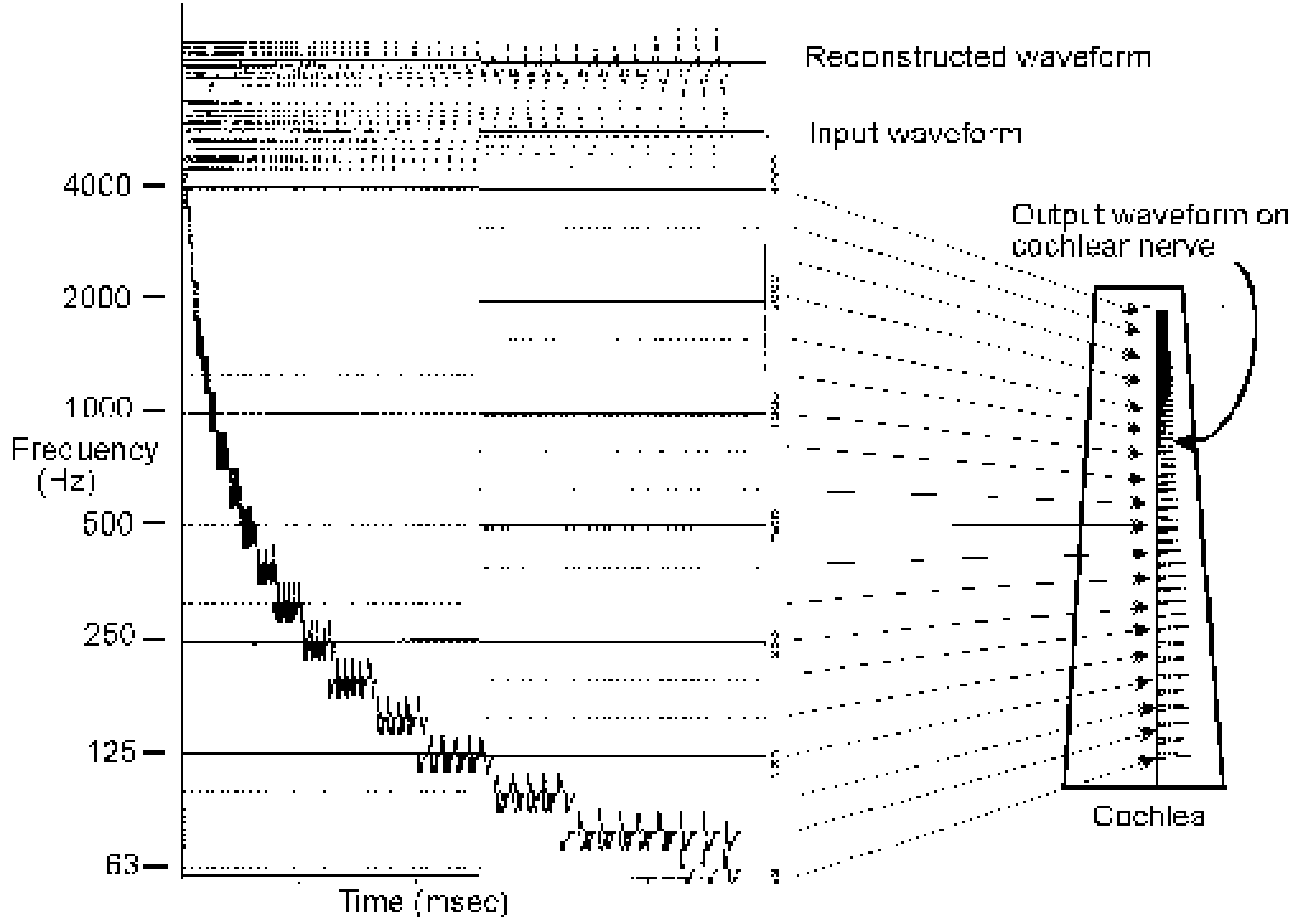
- An input stream of labeled WIVs
- A tapped delay line to store WIVs <u>in their time sequence </u>for a duration equal to the longest teset to be recognized
- Connection of the delay line taps to a set of contiguous templates specifying *all possible tesets* in the selected range
- Logic that identifies the point in time when any teset in the input stream is recognized by a template
- A means to label each recognized labeled WIV teset with the periodicity that matches its template
- A means to prevent erroneous recognition of submultiples in continuous tonal streams



Illustrating recognition of a single teset



Illustrating recognition and separation of two tesets



Example of PSM display showing its 6-octave range and
its inherent tonotopicity matching that of the human cochlea. "

A new signal processing paradigm

My algorithm for separating periodic sequences (pitch detection) holds the answer to the entire problem of separating and understanding sound sources, called computational auditory scene analysis (CASA.)

It can become a new paradigm for all higher-level methods of extracting meaning from sound.