Hallucinations in Auditory Perception!!

Malcolm Slaney
Yahoo! Research
Stanford CCRMA
Hadoop
"Bummer of a birthmark, Hal."
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
One Dimensional
(waveform)

Two Dimensional
(not a spectrogram)

Three Dimensional
(neural movie)
Correlogram

Distance down cochlea

Center Frequency

QuickTime™ and a Video decompressor are needed to see this picture.

Time Interval (s)
Autocorrelation Lag

With help from Richard O. Duda
Correlogram

QuickTime™ and a Video decompressor are needed to see this picture.
Success

• Reconstructing from correlogram
  – NIPS Keynote
Problems

• Continuation
  – Tone and Noise
  – Parliament Cough

• Hear two voices?

• What do you hear?
  – Waveforms?
  – Ideas?
Speech Examples

Wedding

Sine

Natural
What Vowel is This?

Peter Ladefoged

Word 1

Word 2

Word 3
QuickTime™ and a
Animation decompressor
are needed to see this picture.
QuickTime™ and a Animation decompressor are needed to see this picture.
QuickTime™ and a Animation decompressor are needed to see this picture.
McGurk

QuickTime™ and a Cinepak decompressor are needed to see this picture.
Speech

Object

Sinewave

Speech

Object

Wedding

Speech

Environment

Vowel?

Vision

Ventraloquism

Audio Locate

Vision

Dots

Audio Locate

Vision

McGurk

Speech
ASR

Language model for the words: “one”, “two”, “three”

Word model showing phonemes for the word one

Acoustic (phoneme) model for the phoneme /Λ /

S1 → S2 → S3
Conventional Scene Analysis

Computer models of grouping

- “Bregman at face value” (e.g. Brown 1992):

- feature maps
- periodicity cue
- common-onset boost
- resynthesis
Barker—ASR
Goto—CASA with MIDI

MIDI Sequence

Music scene descriptions

Local descriptions

Melody line
Bass line
Hierarchical beat structure

Global descriptions

Repeated sections
Chorus sections

Compact disc (CD)
Musical audio signals
Old plus New Principle

The effect of context

- Context can create an ‘expectation’: i.e. a bias towards a particular interpretation

- e.g. Bregman’s “old-plus-new” principle: A change in a signal will be interpreted as an added source whenever possible

- a different division of the same energy depending on what preceded it
Ellis—Prediction Driven
Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map

Saliency
Saliency Example

- Time-frequency display
- Saliency map shows high-interest locations
Saliency Maps

- Longer tones better
- Missing parts salient
- Modulation more salient
- Forward masking works
Sound Examples

- Birds
- Calls
- Cows
- Horse
- Waterfall
Saliency Comparison

- Details of saliency comparison
- Model predictions
Relational Network (Simple)

- Patches of neurons
- Each measure one quantity
- Bidirectional relations for feedback/feedforward

Thanks to Rodney Douglas
Relational Network (example)
ASR Relational Network

Note: We don’t know how to represent delays

A patch of neurons (one of N output)
**Desired Results**

<table>
<thead>
<tr>
<th>Phoneme Patch</th>
<th>Relational Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>/A/ Phoneme Patch</td>
<td>Without</td>
</tr>
<tr>
<td>/I/ Phoneme Patch</td>
<td></td>
</tr>
<tr>
<td>AI Word Patch</td>
<td></td>
</tr>
<tr>
<td>IA Word Patch</td>
<td></td>
</tr>
<tr>
<td>Phoneme Input</td>
<td></td>
</tr>
</tbody>
</table>
Simulation

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.
Simulation 2
Simulation 3
Grossberg—ART

Statistical Means

• ICA
  – Different distributions

• One Microphone
  – GMM models of distribution

Separated signals after 4 steps of FastICA
Conventional
Better?
Thanks

malcolm@ieee.org
Pitch

Integrated Correlogram

Narrowed Autocorrelation

Pitch Peak

Time Lag (ms)
Silicon Frequency Response

• Tone ramps into two cochleas
Cochlear Best Frequency
Cochlear Rate Profiles

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

Spikes per utterance

Left Cochlea    Right Cochlea
Hardware Overview

- Cochlea
- PCI-AER (for remapping)
- Phoneme Learning
- Word Learning
- PCI-AER (for remapping)

Implemented in MATLAB

Shih-Chii Liu

Giacomo Indiveri
LSH Movie
Auditory Map

By Lloyd Watts
Please do more Neurophysiology!

David  Jerry  Prabhatkar
Rushdie, Lakshmi split
Controversial author Salman Rushdie and "Top Chef" host Padma Lakshmi will end their marriage. » Why?
• 'Top Chef' info, videos and more
• Criss Angel's wife wants a divorce
Salman Rushdie, Padma Lakshmi divorcing
iPhone buyers smash and dissect new gadget
Bonds to start All-Star Game at home park
Brain scans reveal why meditation works
More Featured
In the News
As of 11:15 a.m. PDT
• U.K. arrests Iraqi, Jordanian doctors in bomb probe
• Appeals court refuses to delay Libby's prison sentence
• Bush, Putin project united front on Iran
• American embassy says U.S. diplomat in Cyprus found dead
Playing host

Barry Bonds overtook Alfonso Soriano in the fans voting and will start in the All-Star Game at home park.  Tim Brown column  Griffey happy

Manager shakeups: Reds fire Narron  Mariners' Hargrove resigns
U.K. probe nets Iraqi, Jordanian doctors

GLASGOW, Scotland - Doctors from Iraq and Jordan were among the seven suspects arrested in the failed car bombings in London and at Glasgow’s airport, officials said Monday. A witness said police were closing in on the terror network just before attackers rammed the Scottish terminal building.

Court won’t delay Libby prison sentence

WASHINGTON - Former White House aide I. Lewis “Scooter” Libby cannot delay his 2 1/2-year prison term in the CIA leak case, a federal appeals court unanimously ruled Monday.

Yemeni temple bombing kills 7 Spaniards

SAN’A, Yemen - An al-Qaida suicide bomber drove into a convoy of Spanish tourists visiting an ancient Yemeni temple, officials said, killing seven Spaniards and two Yemenis less than two weeks after a U.S. terror warning about the area.
Settle your **beef** at **BurgerCon**

**ADVERTISEMENT**

**CELEBRATE 100% BEEF**

Click to
TASTE THE EXCITEMENT

Quarter Pounder®
with Cheese

Big Mac ®

*Weight before cooking 4 oz. (113.4g) ©2007 McDonald's*

**MUSIC VIDEOS**

Yahoo! Music Video Premiere

Sean Kingston - "Beautiful Girls"
Rock around the clock with Sean!

Browse Top-100 Videos >

**RADIO**

Listen to My Station

Select a station below

- 4th of July
- Classic Rock
- Pop & Rock New Releases

**GET MORE**

Looking for the words to your favorite songs?
Find them with the new Yahoo! Music Lyrics

**ARTIST SPOTLIGHT**

Get Your Freak Avril’s Fans-Only

Audition For Avril
Last chance to enter!

More Artist Spotlight

**CRITICAL EYE**

**The Clientele**
Shoegazer rock is back! So stare down at your tiptoes and get lost in the lush, chiming
TOP CHEF' MARRIAGE ON THE CHOPPING BLOCK

PHOTOS
HAPPY 21ST BIRTHDAY LINDSAY!

ADVERTISEMENT

Lipton
Sometimes the lightest touch brings the biggest sensation.

LIPTON TEA
Timbre definition

• Sound color
  – Instruments
  – Vowels

Timbre

Static

Dynamic

Pitch

Loudness

All sound
Multi-Dimensional Scaling of Timbre

- Measure
  - Distances
- Estimate
  - Positions
- Art
  - Label axis

Spectral flux
Spectral centroid
Decay

McAdams et al. (1995)
Desired perception model

- Compact (parsimonious)

- Three Properties
  - Predictive
    - Explain distance perception
  - Simple model
    - Orthogonal axis
  - Linear model
    - Interpolate sounds

Test
Euclidean distance

Assumption
Experimental Contrast

Old Way

Sound → Perception → Model

New Way

Parameter space → Sound → Perception

Guess a model that fits the data
Spectral shape using MFCC

- A huge tapestry hung in her hallway.
MFCC and LFC

\[ C = [C_0, C_1, C_2, \ldots, C_{12}] \]
Kernel function of DCT

- Spectrum
  - superposition of DCT kernels
- Cepstrum coefficients
  - Coefficients for superposition
Parameter space: MFCC

$C = [1, 0, 0, C_3, 0, 0, C_6, 0, 0]$
Parameter space: LFC

\[ C = [1, 0, 0, C_3, 0, 0, C_6, 0, 0, 0] \]
Synthesize stimuli

- Harmonics: pitch and vibrato
  - Amplitude weighted by the spectral shape

Desired spectral shape
Vertical - frequency, Horizontal - amplitude
Experiment procedures

- Paired stimuli (AB, AG, AD, …)
- Rate dissimilarities using 0-9 scale
- 10 subjects
  - Quiet office
  - Individual sessions (headphone)
Euclidean Fitting

- 2D linear regression
  \[ d^2 = ax^2 + by^2 \]
  - Known values: x, y, d - estimate a and b

- Residual from Euclidean model
Results summary

![Graph showing comparison between Tristimulus model, LFC, and MFCC.](image)
Experiment results

• MFCC: most successful timbre model

• Less linearity for high coeffs
DIVERSION
Remix Examples

- Abba *Gimme Gimme*
- Madonna *Hung Up*
- Tracy Young Remix of *Hung Up*
- Tracy Young Remix 2 of *Hung Up*
Specificity Spectrum

Fingerprinting  Remixes  Cover songs  Genre

Look for specific exact matches

Our work (nearest neighbor)

Bag of Features model
Cross-Correlation

- 2M songs
  - 3 minutes
  - 10 frames/second

72 Billion
Curse of Dimensionality

- Histogram of distances between Gaussian data
  - Normalized to the mean

- Nearest Neighbor

Ill-posed?
Distractors
Correlogram

Distance down cochlea

Center Frequency

Time Interval (s)

Autocorrelation Lag