LECTURE 05: PERCEPTION AND MASKING

- Objectives:
  - Frequency and temporal masking
  - Introduce other impairments such as echo and appreciate how they impact speech processing systems
  - Appreciate how we can exploit properties of masking in speech analysis
  - Summarize our digital models/approximations

Note that this lecture is primarily based on material from the course textbook:

> X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

In addition, information from:

> D. O'Shaughnessy, *Speech Communications: Human and Machine*, IEEE Press, ISBN: 0-7803-3449-3, 2000.

has been used.

Return to Main

Introduction:

01: Organization
(html, pdf)

Speech Signals:

02: Production
(html, pdf)

03: Digital Models
(html, pdf)

04: Perception
(html, pdf)

05: Masking
(html, pdf)

06: Phonetics and Phonology
(html, pdf)

07: Syntax and Semantics
(html, pdf)

Signal Processing:

08: Sampling
(html, pdf)

09: Resampling
(html, pdf)

10: Acoustic Transducers
(html, pdf)

11: Temporal Analysis
(html, pdf)

12: Frequency Domain Analysis
(html, pdf)

13: Cepstral Analysis
(html, pdf)

14: **Exam No. 1**
(html, pdf)

15: Linear Prediction
(html, pdf)

16: LP-Based Representations
(html, pdf)

Parameterization:

17: Differentiation
(html, pdf)

18: Principal Components
(html, pdf)

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

19: Linear Discriminant Analysis (html, pdf)

LECTURE 05: PERCEPTION AND MASKING

- Objectives:
  - Frequency and temporal masking
  - Introduce other impairments such as echo and appreciate how they impact speech processing systems
  - Appreciate how we can exploit properties of masking in speech analysis
  - Summarize our digital models/approximations

Note that this lecture is primarily based on material from the course textbook:

> X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

In addition, information from:

> D. O'Shaughnessy, *Speech Communications: Human and Machine*, IEEE Press, ISBN: 0-7803-3449-3, 2000.

has been used.

TONE-MASKING NOISE

- **Frequency masking**: one sound cannot be perceived if another sound close in frequency has a high enough level. The first sound *masks* the second.
- **Tone-masking noise**: noise with energy E_N (dB) at Bark frequency g masks a tone at Bark frequency b if the tone's energy is below the threshold:

  $$T_T(b) = E_N - 6.025 - 0.275g + S_m(b-g) \quad \text{(dB SPL)}$$

  where the *spread-of-masking* function $S_m(b)$ is given by:

  $$S_m(b) = 15.81 + 7.5(b+0.474) - 17.5*\text{sqrt}(1 + (b+0.474)^2) \quad \text{(dB)}$$

- **Temporal Masking**: onsets of sounds are masked in the time domain through a similar masking process.

Key points:

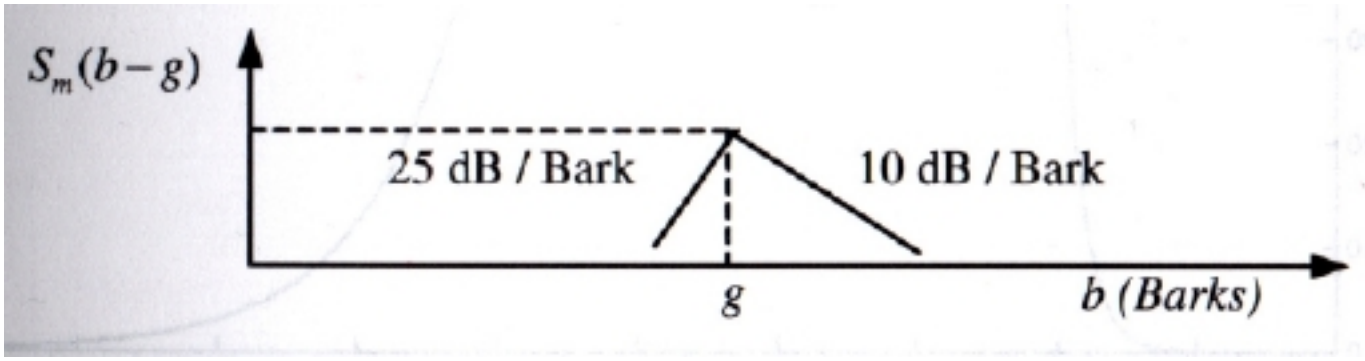
- Thresholds are frequency and energy dependent.
- Thresholds depend on the nature of the sound as well.

NOISE-MASKING TONE

- **Noise-masking tone**: a tone at Bark frequency *g* energy E_T (dB) masks noise at Bark frequency *b* if the noise energy is below the threshold:

$$T_N(b) = E_T - 2.025 - 0.17g + S_m(b-g) \quad \text{(dB SPL)}$$

- Masking thresholds are commonly referred to as Bark scale functions of *just noticeable differences* (JND).

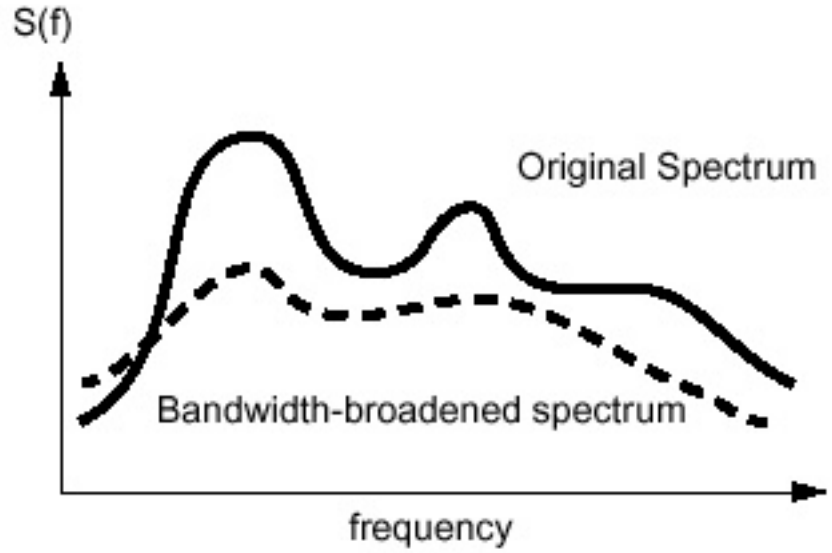


Key points:

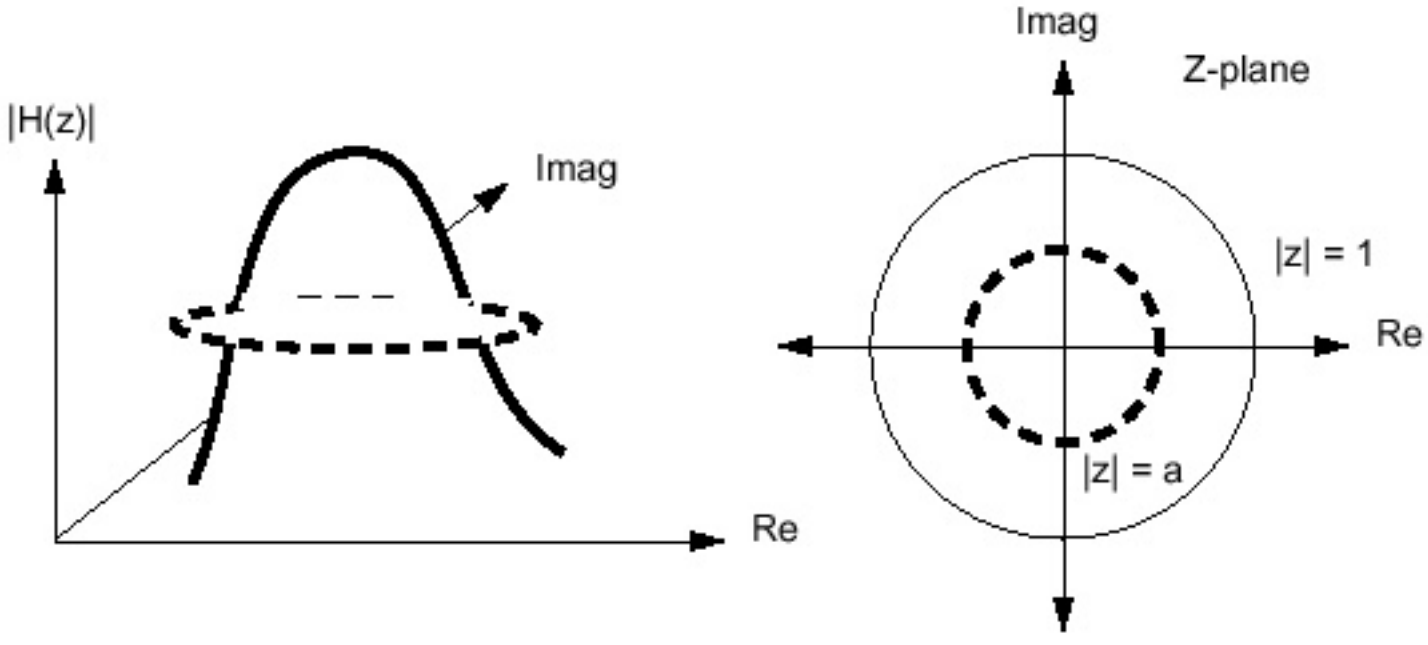
- Thresholds are not symmetric.
- Thresholds depend on the nature of the noise and the sound.

PERCEPTUAL NOISE WEIGHTING

- **Noise-weighting**: shaping the spectrum to hide noise introduced by imperfect analysis and modeling techniques (essential in speech coding).
- Humans are sensitive to noise introduced in low-energy areas of the spectrum.
- Humans tolerate more additive noise when it falls under high energy areas the spectrum. The amount of noise tolerated is greater if it is spectrally shaped to match perception.
- We can simulate this phenomena using "bandwidth-broadening":



- Simple Z-Transform interpretation:



which can be implemented by evaluating the Z-Transform around a contour closer to the origin in the z-plane:

$H_{nw}(z) = H(az)$.

Key points:

- Used in many speech compression systems (Code Excited Linear Prediction).
- Analysis performed on bandwidth-broadened speech; synthesis performed using normal speech. Effectively shapes noise to fall under the formants.

ECHO, THE LOMBARD EFFECT, AND TIME DELAY

- Humans are used to hearing their voice while they speak - real-time feedback (side tone).
- When we place headphones over our ears, which dampens this feedback, we tend to speak louder.
- **Lombard Effect**: Humans speak louder in the presence of ambient noise.
- When this side-tone is delayed, it interrupts our cognitive processes, and degrades our speech.
- This effect begins at delays of approximately 250 ms.
- Modern telephony systems have been designed to maintain delays lower than this value (long distance phone calls routed over satellites).
- Digital speech processing systems can introduce large amounts of delay due to non-real-time processing.

ADAPTATION

- **Adaptation** refers to changing sensitivity in response to a continued stimulus, and is likely a feature of the mechanoelectrical transformation in the cochlea.
- Neurons tuned to a frequency where energy is present do not change their firing rate drastically for the next sound.
- Additive broadband noise does not significantly change the firing rate for a neuron in the region of a formant.
- The McGurk Effect is an auditory illusion which results from combining a face pronouncing a certain syllable with the sound of a different syllable. The illusion is stronger for some combinations than for others. For example, an auditory 'ba' combined with a visual 'ga' is perceived by some percentage of people as 'da'. A larger proportion will perceive an auditory 'ma' with a visual 'ka' as 'na'. Some researchers have measured evoked electrical signals matching the "perceived" sound.

TIMING

- Temporal resolution of the ear is crucial.
- Two clicks are perceived monoaurally as one unless they are separated by at lest 2 ms.
- 17 ms of separation is required before we can reliably determine the order of the clicks.
- Sounds with onsets faster than 20 ms are perceived as "plucks" rather than "bows".
- Short sounds near the threshold of hearing must exceed a certain intensity-time product to be perceived.
- Humans do not perceive individual "phonemes" in fluent speech - they are simply too short. We somehow integrate the effect over intervals of approximately 100 ms.
- Humans are very sensitive to long-term periodicity (ultra low frequency) - has implications for random noise generation.

DIGITAL MODELS FOR PERCEPTION

- Logarithmic processing of energy.
- Energy normalization.
- Nonlinear warping of the frequency scale.
- Filter bank analysis (wavelets).
- Cochlear models have not been extremely effective.

W.A.V.S. Compression

Background on the Psychoacoustic Model

The psychoacoustic model is based on many studies of human perception. These studies have shown that the average human does not hear all frequencies the same. Effects due to different sounds in the environment and limitations of the human sensory system lead to facts that can be used to cut out unnecessary data in an audio signal.

The two main properties of the human auditory system that make up the psychoacoustic model are:

- absolute threshold of hearing
- auditory masking.

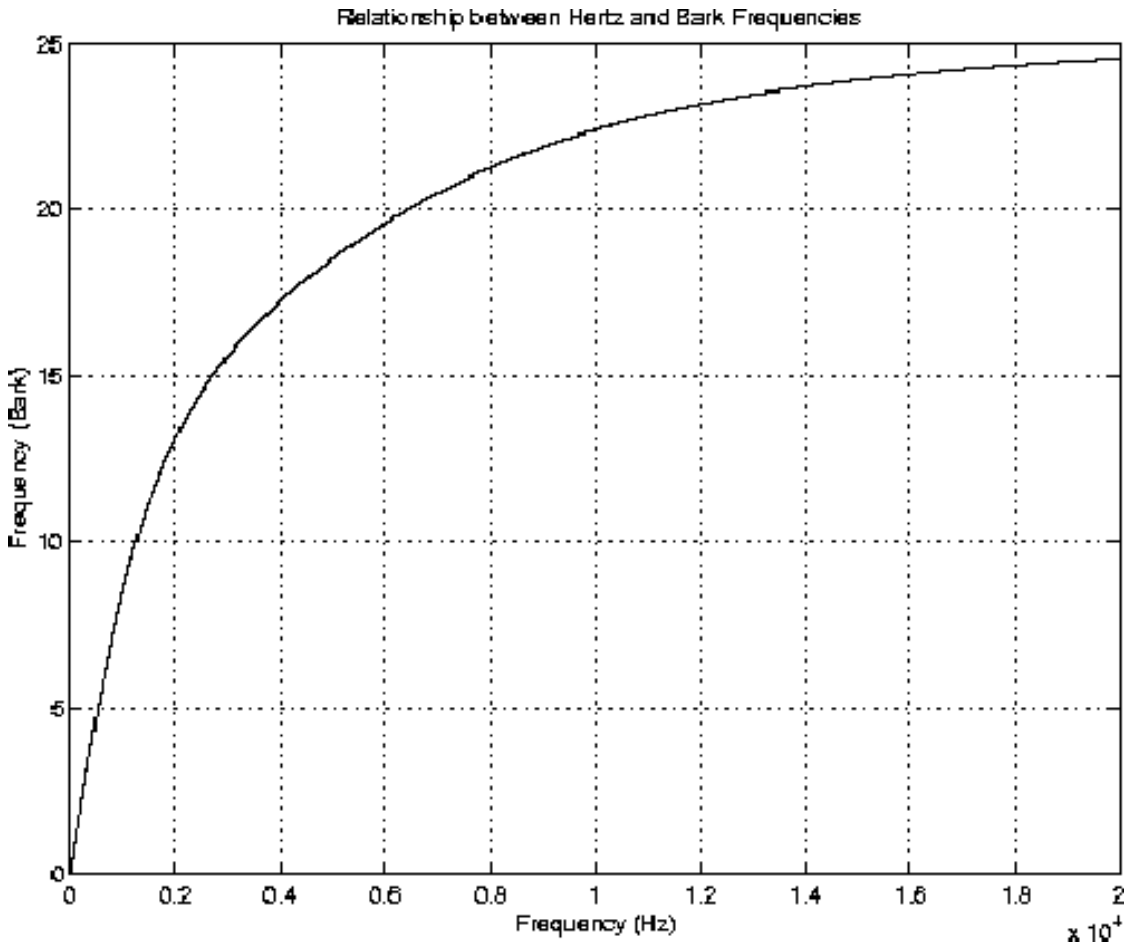
Each provides a way of determining which portions of a signal are inaudible and indiscernible to the average human, and can thus be removed from a signal.

Absolute Threshold of Hearing

Humans can hear frequencies in the range from 20 Hz to 20,000 Hz. However, this does not mean that all frequencies are heard in the same way. One could make the assumption that a human would hear frequencies that make up speech better than others; this is a good guess. Furthermore, one could also hypothesize that hearing a tone becomes more difficult as its frequency nears either of the extremes. Again, this is true.

One other observation forms the basis for modeling. Because humans hear lower frequencies, like those making up speech, more than others, like high frequencies around 20 kHz, the ear probably has better capability in detecting differences in pitch at lower frequencies than at high ones. This, too, is true. For example, a human has an easier time telling the difference between 500 Hz and 600 Hz than he does determining whether something is 17,000 Hz or 18,000 Hz. After many studies, scientists found that the frequency range from 20 Hz to 20,000 Hz can be broken up into critical bandwidths, which are non-uniform, non-linear, and dependent on the heard sound. Signals within one critical bandwidth are hard to separate for a human observer.

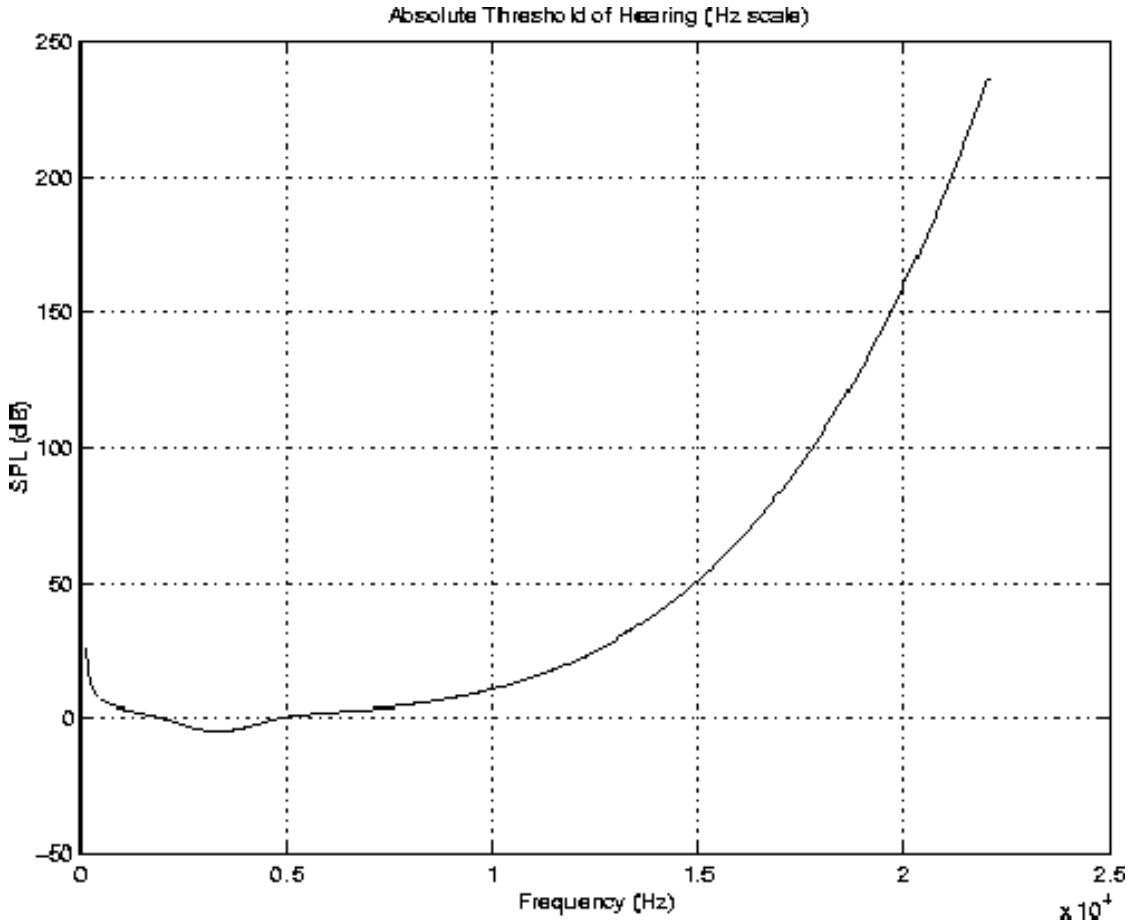
A more uniform measure of frequency based on critical bandwidths is the Bark. From the earlier discussed observations, one would expect a Bark bandwidth to be smaller at low frequencies (in Hz) and larger at high ones. Indeed, this is the case.

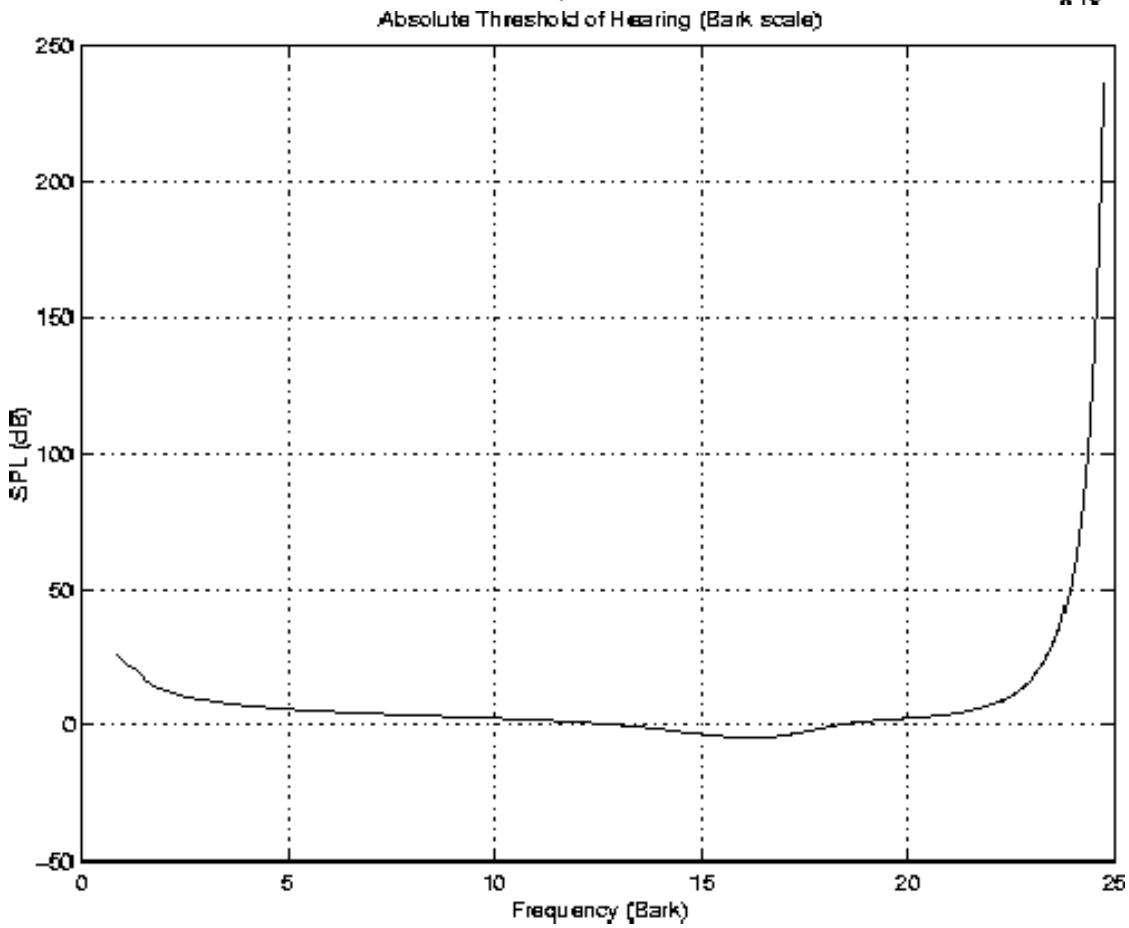


The Bark frequency scale can be approximated by the following equation:

barks = 13*arctan(0.00076*Hz) + 3.5*arctan((f/7500)^2)

To determine the effect of frequency on hearing ability, scientists played a sinusoidal tone at a very low power. The power was slowly raised until the subject could hear the tone. This level was the threshold at which the tone could be heard. The process was repeated for many frequencies in the human auditory range and with many subjects. As a result, the following plot was obtained.





This experimental data can be modeled by the following equation, where f is frequency in Hertz:

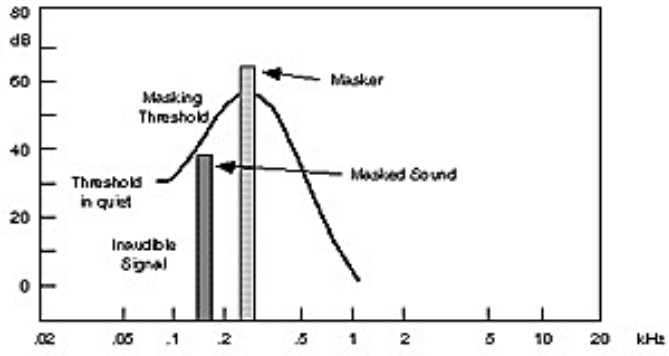
ATH(f) = 3.64 * (f/1000)^-0.8 - 6.5e^(-0.6*((f/1000) - 3.3)^2) + 10^-3*(f/1000)^4 (dB SPL)

Thus, we can make the following jump for the purposes of compression. If a signal has any frequency components with power levels that fall below the absolute threshold of hearing, then these components can be discarded, as the average listener will be unable to hear those frequencies of the signal anyway.

Auditory Masking

Humans do not have the ability to hear minute differences in frequency. For example, it is very difficult to discern a 1,000 Hz signal from one that is 1,001 Hz. This becomes even more difficult if the two signals are playing at the same time. Furthermore, the 1,000 Hz signal would also affect a human's ability to hear a signal that is 1,010 Hz, or 1,100 Hz, or 990 Hz.

This concept is known as masking. If the 1,000 Hz signal is strong, it will mask signals at nearby frequencies, making them inaudible to the listener. For a masked signal to be heard, its power will need to be increased to a level greater than that of a threshold that is determined by the frequency of the masker tone and its strength.



It turns out that noise can be a masker as well. If noise is strong enough, it can mask a tone that would be clear otherwise. For example, a jet engine, which is very noisy, can drown out music easily.

In a compression algorithm, therefore, one must determine:

- tone maskers
- noise maskers
- masking effect due to these maskers.

If any frequency components around these maskers fall below the masking threshold, they can be discarded.

Tone Maskers

Determining whether a frequency component is a tone requires knowing whether it has been held constant for a period of time, as well as whether it is a sharp peak in the frequency spectrum, which indicates that it is above the ambient noise of the signal.

For the purposes of this project, only the second criterion is considered. Determining whether a certain frequency is a tone (masker) can be done with the following definition:

A frequency f (with FFT index k) is a tone if its power P[k] is:

1. greater than P[k-1] and P[k+1], i.e., it is a local maxima
2. 7 dB greater than the other frequencies in its neighborhood, where the neighborhood is dependent on f:
    - If 0.17 Hz < f < 5.5 kHz, the neighborhood is [k-2…k+2].
    - If 5.5 kHz =< f < 11 kHz, the neighborhood is [k-3…k+3].

- If 11 kHz =< f < 20 kHz, the neighborhood is [k-6…k+6].

Noise Maskers

If a signal is not a tone, it must be noise. Thus, one can take all frequency components that are not part of a tone's neighborhood and treat them like noise. Combining such components into maskers, though, takes a little more thought.

Since humans have difficulty discerning signals within a critical band, the noise found within each of the bands can be combined to form one mask. Thus, the idea is to take all frequency components within a critical band that do not fit within tone neighborhoods, add them together, and place them at the geometric mean location within the critical band. Repeat this for all critical bands.

Masking Effect

The maskers which have been determined affect not only the frequencies within a critical band, but also in surrounding bands. Studies show that the spreading of this masking has an approximate slope of +25 dB/Bark before and -10 dB/Bark after the masker.

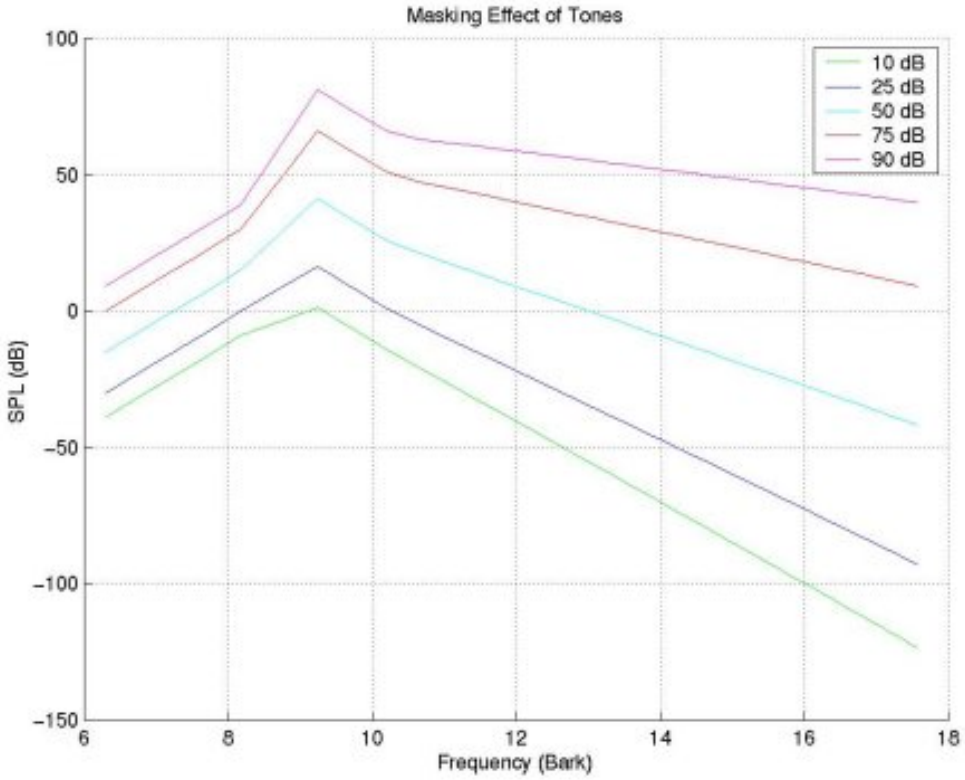
The spreading can be described as a function that depends on the maskee location i, the masker location j, the power spectrum Ptm at j, and the difference between the masker and maskee locations in Barks (deltaz=z(i)-z(j)):

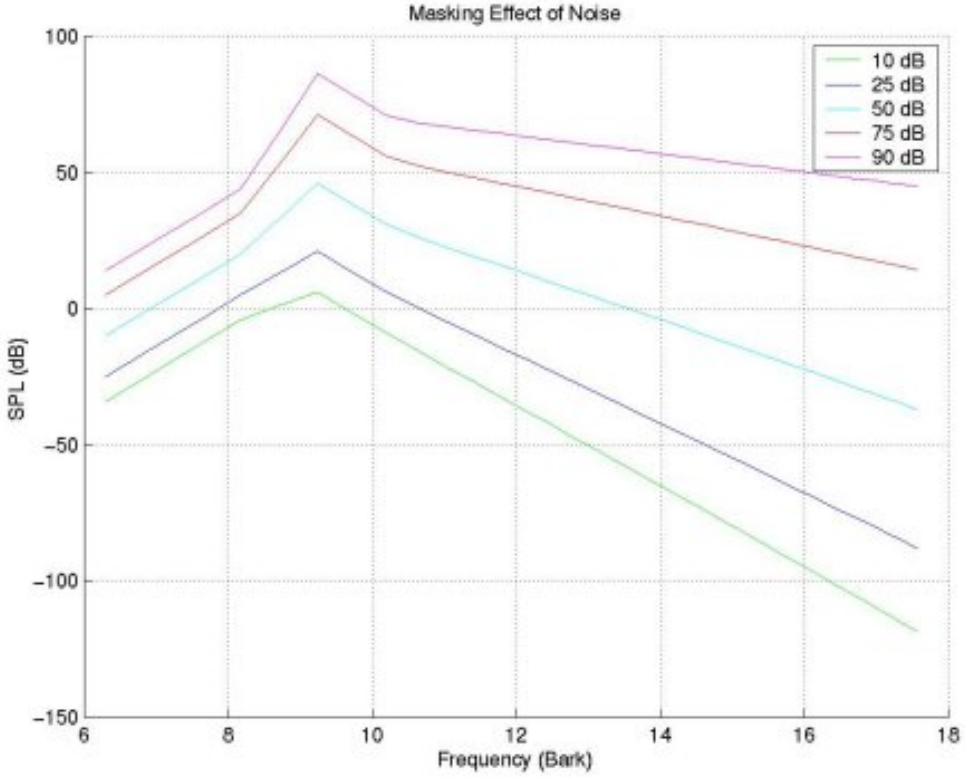
$$SF(i,j) = \begin{cases} 17\text{deltaz} - 0.4\text{Ptm}(j)+11 & -3 <= \text{deltaz} < -1 \\ (0.4\text{Ptm}(j)+6)\text{deltaz} & -1 <= \text{deltaz} < 0 \\ -17\text{deltaz} & 0 <= \text{deltaz} < 1 \\ (0.15\text{Ptm}(j)-17)\text{deltaz} - 0.15\text{Ptm}(j) & 1 <= \text{deltaz} < 8 \end{cases}$$

There is a slight difference in the resulting mask that depends on whether the mask is a tone or noise. As a result, the masks can be modeled by the following equations, with the same variables as described above:

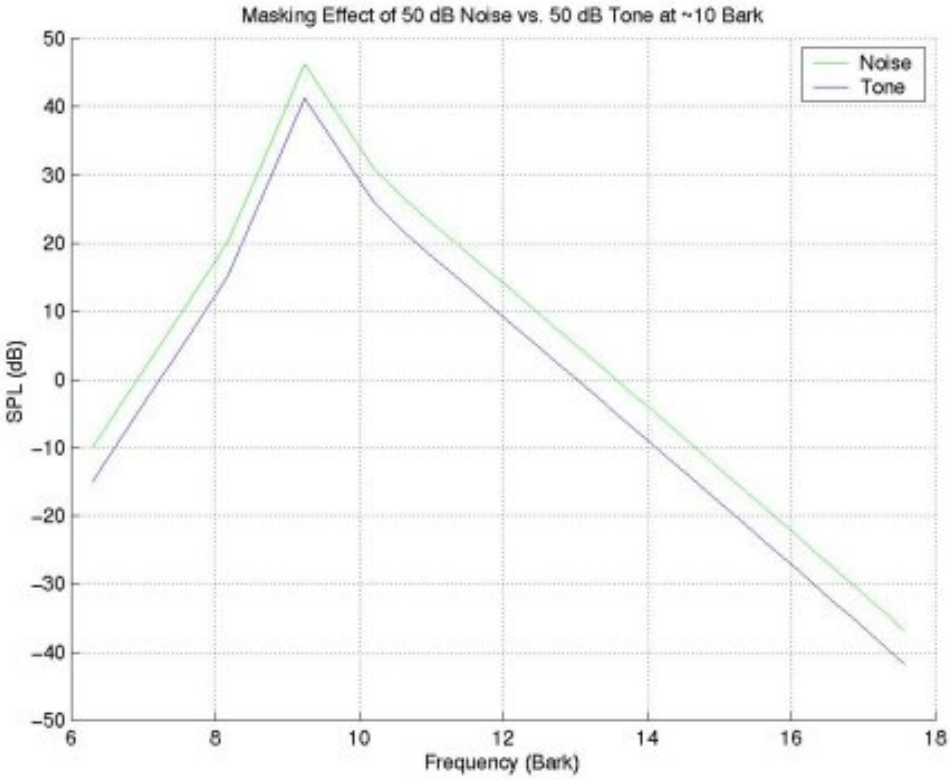
For tones: Ttm(i,j) = Ptm(j) - 0.275z(j) + SF(i,j) - 6.025 (dB SPL)
For noise: Tnm(i,j) = Pnm(j) - 0.175z(j) + SF(i,j) - 2.025 (dB SPL)

The following are plots of various levels of tone and noise maskers.





The final plot compares a tone and noise masker at the same frequency and of the same power.



Naturally, if there are multiple noise and tone maskers, the overall effect is a little harder to determine. In this project, the assumption is made that the effects are power additive. This is a reasonable assumption to make, but note that there is a definitely an interplay that can occur between maskers that would lower or increase thresholds.

[Alex Chen] [Nader Shehad] [Aamir Virani] [Erik Welsh]

Malcolm Slaney

Publications and Pointers

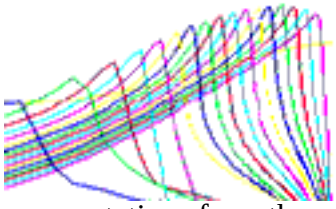
I now work at IBM's Almaden Research Center. I used to work for Interval Research, Apple Computer's Advanced Technology Group, and Schlumberger Palo Alto Research.

Several of my technical reports and papers are available on the net for downloading. The following is a brief list. I have a personal web page for the fun stuff.

This page shows my auditory modeling work, my signal processing work, some of my software tools, and pointers to other work.

Note! My tomography book is now online. Get more information here. The book is back in print and you can order it now from SIAM!!!

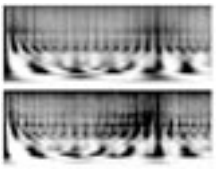
Auditory Modeling



There is now a new version of the Auditory Toolbox. It contains Matlab functions to implement many different kinds of auditory models. The toolbox includes code for Lyon's passive longwave model, Patterson's gammatone filterbank, Meddis' hair cell model, Seneff's auditory model, correlograms and several common representations from the speech-recognition world (including MFCC, LPC and spectrograms). This code has been tested on Macintosh, Windows, and Unix machines using Matlab 5.2.

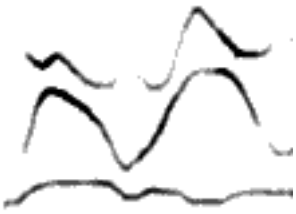
Auditory Toolbox (Version 2.0)

Note: This toolbox was originally published as Apple Computer Technical Report #45. The old technical report (PDF PDF and Postscript) and old code (Unix TAR and Macintosh BinHex) are available for historical reasons.



My primary scientific goal is to understand how our brains perceive sound. My role in this research area is a modeler, I build models that explain the neurophysiological and psychoacoustic data. Hopefully these models will help other researchers understand the mechanisms involved and result in better experiments. My latest work in this area is titled "Connecting Correlograms to Neurophysiology and Psychoacoustics" and was presented at the XIth International Symposium on Hearing in Grantham England from 1-6 August, 1997. Two correlograms, one computed using autocorrelation and other other computed using AIM, are shown on the left.

Abstract



The information in most auditory models flows exclusively bottom-up, yet there is increasing evidence that a great deluge of information is flowing down from the cortex. A paper I wrote for the 1995 Computational Auditory Scene Analysis workshop is called "A Critique of Pure Audition". This paper has been greatly refined and is published in the book *Computational Auditory Scene Analysis* in 1998 by Erlbaum. The figure at the left shows the spectrogram of sine-wave speech.

Book chapter (153k pdf)

Book chapter (620k postscript)

Audio/video examples

Original paper

I have written several papers describing how to convert auditory representations into sounds. I have built models of the cochlea and central auditory processing, which I hope both explain auditory processing and will allow us to build auditory sound separation tools. These papers describe the process of converting sounds into cochleagrams and correlograms, and then converting these representations back into sounds. Unlike the printed versions of this work, the web page includes audio file examples. It includes better spectrogram inversion techniques, a description of how to invert Lyon's passive cochlear model, and a description of correlogram inversion. This material was first presented as part of the *Proceedings of the ATR Workshop on "A Biological Framework for Speech Perception and Production"* published in September 1994. A more refined version of this paper was an invited talk at the 1994 NIPS conference. The image on the left shows the spectrogram of one channel of cochlear output; one step in the correlogram inversion process.

ATR (Kyoto) Workshop Web Reprint with Sound Examples

Keynote NIPS Conference Paper (Postscript)

Pattern Playback is the term used by Frank Cooper to describe his successful efforts to paint spectrogram on plastic and then convert them into sound. I wrote of Pattern Playback techniques, from Frank Cooper's efforts to my own efforts with auditory model inversion, in a paper which was published at the *1995 IEEE International Conference on Systems, Man, and Cybernetics*. My paper is titled "Pattern Playback from 1950 to 1995". The image at the left shows a portion of one of Cooper's spectrograms.

Web Version

Postscript (1.8M)

Adobe PDF (227k)

The following are publications during my time at Apple. The *Mathematica* notebooks are designed to be self-documenting and in each case the postscript and PDF files are also available. Those files that are *Matlab* toolboxes include source and documentation All these files are available with the gracious permission of Apple.

"Auditory Model Inversion for Sound Separation" is the first paper to describe correlogram inversion techniques. We also discuss improved methods for inverting spectrograms and a cochlear model designed by Richard F. Lyon. This paper was published at ICASSP '94.

Postscript (1.5M)

Adobe PDF (243k)

Online patent

"A Perceptual Pitch Detector" is a paper that describes a model of human pitch perception. It is similar to work done by Meddis and Hewitt and published in JASA, but this paper has more real-world examples. This paper was published at ICASSP '90.

Postscript (3M)

Adobe PDF (315k)

"On the importance of time" is an invited chapter by Dick Lyon and myself in the book *Visual Representations of Speech Signals* (edited by Martin Cooke, Steve Beet and Malcolm Crawford, John Wiley & Sons). This tutorial describes the reason that we think time-domain processing is important when modeling the cochlea and higher-level processing.

Postscript

Adobe PDF

"Lyon's Cochlear Model" is a *Mathematica* notebook that describes an implementation of simple (but efficient) cochlear model designed by Richard F. Lyon. It is also known as Apple Technical Report #13.

Mathematica Notebook (1.2M)

Postscript (2.2M)

Adobe PDF (628k)

A software package called MacEar implements the latest version of Lyon's Cochlear Model. MacEar is written in very portable C for Unix and Macintosh computers. This link points to the last published version (2.2). (Note the README file included has old program results. The names of the output files have changed and there are a couple of extra channels being output. I'm sorry for the confusion.)

Unix Shell Archive with Sources

Gammatone Math is a *Mathematica* notebook that describes a new more efficient implementation of the Gammatone filters that are often used to implement critical band models. It is also known as Apple Technical Report #35.

Mathematica Notebook (327k)

Postscript (677k)

Adobe PDF (184k)

Apple Hearing Demo Reel was published as Apple Technical Report #25. It includes more than one hour of correlogram videos, including a large fraction of the ASA Auditory Demonstration CD. I have a limited number of NTSC copies left. Send email to malcolm@ieee.org to request a copy.

HTML Video Guide

PDF Video Guide (116k)

Postscript Video Guide (195k)

Signal Processing

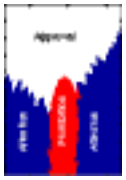
I recently finished some nice work establishing a linear operator connecting the audio and video of a speaker. A paper describing this work has been accepted for presentation at the NIPS'2000 conference.

PDF Paper (600k)

Chris Bregler, Michele Covell, and I developed a technique we call Video Rewrite to automatically synthesize video of talking heads. This technology is cool because we use a purely data driven approach (concatenative triphone video synthesis) to create new video of a person speaking. Given new audio, we concatenate the best sequence of lip images and morph them into a background sequence. We can automatically create sequences like the Kennedy and Johnson scenes in the movie "Forrest Gump."

Original SIGGRAPH '97 Paper (with examples)

Audio Visual Speech Perception Workshop



We studied how adults convey affective messages to infants using prosody. We did not attempt to recognize the words, let alone to distill more nebulous concepts such as satire or irony. We analyzed speech with low-level acoustic features and discriminated approval, attentional bids, and prohibitions from adults speaking to their infants. We built automatic classifiers to create a system, Baby Ears, that performs the task that comes so naturally to infants. The image on the left shows one of the decision surfaces which classifies approval, attention and prohibition utterances on the basis of their pitch.

Web Page

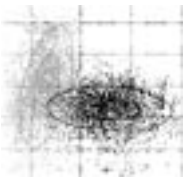
Postscript (189k)

Adobe PDF (42k)

I was able to help Michele Covell do some neat work on time-compression of audio. Lots of people know how to compress a speech utterance by a constant amount. But if you want to do better, which parts of the speech signal can be compressed the most? This paper describes a good technique and shows how to test the resulting comprehension.

Conference Paper

Technical Report with Audio Samples



Eric Scheirer and I worked on a system for discriminating between speech and music in an audio signal. This paper describes a large number of features, how they can be combined into a statistical framework, and the resulting performance on discriminating signals found on radio stations. The results are better then anybody else's results. (That comparison is not necessarily valid since there are no common testing databases. We did work hard to make our test set representative.) This paper was published at the 1997 ICASSP in Munich. The image on the left shows clouds of our data.

Web Page

Postscript (349k)

Adobe PDF (263k)

Work we've done to morph between two sounds is described in a paper at the 1996 ICASSP. This work is new because it extends previous audio morphing work to include inharmonic sounds. This paper uses results from Auditory Scene Analysis to represent, match, warp, and then interpolate between two sounds. The image on the left shows the smooth spectrogram, one of two independent representations used when morphing audio signals.

Web Page

Postscript (3M)

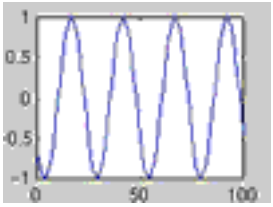
Adobe PDF (237k)Patent

I wrote an article describing my experiences writing "intelligent" signal processing documents. My Mathematica notebook "Lyon's Cochlear Model" was the first large document written with Mathematica. While I don't use Mathematica as much as I used to, I still believe that intelligent documents are a good way to publish scientific results. These ideas were also published in a book titled "Knowledge Based Signal Processing" that was published by Prentice Hall.

KBSP Book Chapter in Adobe PDF (3M)

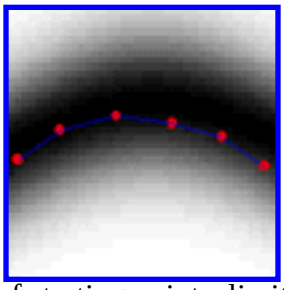
IEEE Signal Processing Article in Adobe PDF (2M)

Software Publications



I have written Matlab m-functions that read and write QuickTime movies. The WriteQTMovie code is more general than previous solutions for creating movies in Matlab. It runs on any platform that Matlab runs on. It also lets you add sound to the movie. The ReadQTMovie code reads and parses JPEG compressed moves.

Matlab Source Code

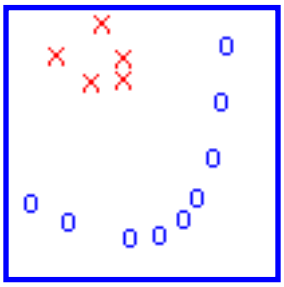


Chris Bregler and I coded an implementation of an image processing technique known as snakes. There are two m-files that implement a type of dynamic contour following popular in computer vision. First proposed by Kass, Witkin and Terzopoulos in 1987, snakes are a variational technique to find the best contour that aligns with an image. The basic routine, snake.m, aligns a sequence of points along a contour to the maximum of an array or image. Provide it with an image, a set of starting points, limits on the search space and it returns a new set of points that better align with the image. The second m-file is a demonstration script. Using your own array of image data, or a built-in default, a demo window is displayed where you can click to indicate points and see the snake program in action.

Matlab Source Code

Matlab Demonstration Source

Image showing GUI



Dick de Ridder and his colleagues wrote a nice description of a Support Vector Classifier and provided some code to demonstrate how it works. I added a Graphical User Interface (GUI) so I could play with all the options and put lots of data through it.

With the GUI, you select points with the mouse. After you tell it what kind of distance metric you want, you get several plots showing the results. The links at the right show a number of points separated by a fourth order polynomial.

Image showing points and support

Image showing distance to hyperplane

Get all the code

Michele Covell and I wrote some Matlab code to compute multi-dimensional scaling (MDS). MDS allows you to reconstruct an estimate of the position of points, given just relative distance data. These routines do both metric (where you know distances) and non-metric (where you just now the order of distances) data.

Technical report containing the code (no documentation).

Apple Publications

The SoundAndImage toolbox is a collection of *Matlab* tools to make it easier to work with sounds and images. On the Macintosh, tools are provided to record and playback sounds through the sound system, and to copy images to and from the scrapbook. For both Macintosh and Unix system, routines are provided to read and write many common sound formats (including AIFF). Only 68k MEX files are included. Users on other machines will need to recompile the software. This toolbox is published as Apple Computer Technical Report #61.

Postscript Documentation (153k)

Adobe PDF Documentation (20k)

Macintosh Archive

Filter Design is a *Mathematica* notebook that describes (and implements) many IIR filter design techniques. It was published as Apple Technical Report #34.

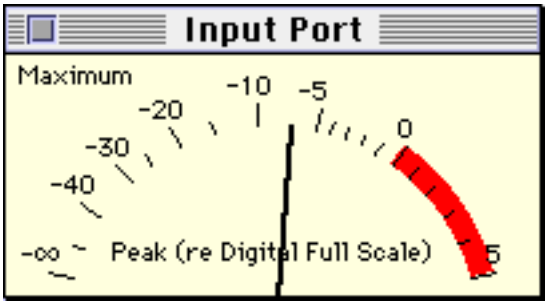
Mathematica Notebook (556k)

Postscript (1M)

Adobe PDF (212k)

I created a Hypercard stack to make it easier for people with a Macintosh and CDROM drive to interact with the Acoustical Society of America's Auditory Demonstrations CD. This CD is a wonderful collection of auditory effects and principles. The ASA Demo Hypercard stack includes the text and figures from the book and lets you browse the Audio CD.

Macintosh Archive



I wrote a program for the Macintosh 660/AV and 840/AV computers that uses the DSP (AT&T3210) to monitor audio levels. VUMeters runs on any Macintosh with the AT&T DSP chip. Source and binaries are included.

Macintosh Archive

Bill Stafford and I wrote TCPPlay to allow us to play sounds from a Unix machine over the network to the Macintosh on our desks. This archive includes Macintosh and Unix source code and the Macintosh application. There are other network audio solutions, but this works well on the Macintosh.

Macintosh Archive

Previous Publications

tr>

In a past life, I worked on medical imaging. A book on tomographic imaging (cross-sectional x-ray imaging) was published by IEEE Press: Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*, (New York : IEEE Press, c1988). The software used to generate many of the tomographic images in this book is available. The parallel beam reconstruction on the left was generated with the commands

```
gen n=100 k=100 if=lib.d.s
filt n=100 k=100
back n=100 k=100
disn min=1.0 max=1.05
```

Tomographic Software (Unix TAR format)

Tomographic Software (Shell archive)

The book is now online. Download the PDF (and republished soon as a SIAM classic!)

Code to implement the diffraction tomography algorithms in my PhD Thesis is also available.

Compressed Unix TAR File

My PhD thesis

Carl Crawford, Mani Azimi and I wrote a simple Unix plotting package called qplot. Both two-dimensional and 3d-surface plots are supported.

Compressed Unix TAR File

Now obsolete code to implement a DITroff previewer under SunView is available. This program was called *suntroff* and is an ancestor of the X Window System Troff previewer. It was written while I was an employee of Schlumberger Palo Alto Research. All files are compressed Unix TAR files.

Source

LaserWriter fonts

Complete package

Other Research Pointers

I organize the Stanford CCRMA Hearing Seminar. Just about any topic related to auditory perception is considered fair game at the seminar. An archive of seminar announcements can be found at Stanford (organized as a table) or at UCSC as a chronological listing of email announcements. Send email to `hearing-seminar-request@ccrma.stanford.edu` if you would like to be added to the mailing list.

For more Information

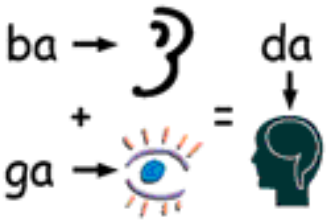
I can be reached at

Malcolm Slaney
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

The best way to reach me is to send email.

This page last updated on August 15, 2001.

Malcolm Slaney (malcolm@ieee.org)



Talking Heads:

The McGurk Effect:

hearing lips and seeing voices

"The most striking demonstration of the combined (bimodal) nature of speech understanding appeared by accident. Harry McGurk, a senior developmental psychologist at the University of Surrey in England, and his research assistant John MacDonald were studying how infants perceive speech during different periods of development. For example, they placed a videotape of a mother talking in one location while the sound of her voice played in another. For some reason, they asked their recording technician to create a videotape with the audio syllable "ba" dubbed onto a visual "ga." When they played the tape, McGurk and McDonald perceived "da." Confusion reigned until they realized that "da" resulted from a quirk in human perception, not an error on the technician's part. After testing children and adults with the dubbed tape, the psychologists reported this phenomenon in a 1976 paper humorously titled "Hearing Lips and Seeing Voices," a landmark in the field of human sensory integration. This audio-visual illusion has become known as the McGurk effect or McGurk illusion."

Dominic W. Massaro & David G. Stork, "Speech Recognition and Sensory Integration",

American Scientist, 1998, vol. 86, p. 236-244.

The McGurk effect has played an important role in audio-visual speech integration and speech reading. Related links include the following:

McGurk Effect home page

UCR Research on Audiovisual Speech Perception (L. Rosenblum)

UCSC PSL McGurk Effect Demo

Dominic W. Massaro & David G. Stork, *American Scientist*, 1998, "Speech Recognition and Sensory Integration"

SYSR Research McGurk Effect Movie

Arnte's McGurk Effect Demo

McGurk Bibliography

Harry McGurk died in April 1998.