

HCS 7367

Speech Perception Lab

Dr. Peter Assmann
Nov 10, 2011

Homework 2

- Spectral shape models
 - Invariant with respect to spectral tilt?
 - Differences in speaking style (e.g., shouting)
 - Differences in vocal source properties (breathy voices)
 - Possible qualification:
 - allow for a constant spectral tilt change
 - Pre-emphasis filter
 - >> pre = [1 -0.9]; % pre-emphasis filter
 - >> y = filter(pre,1,y);

Homework write-up

- Introduction
 - Statement of problem + background information (10%)
- Method
 - Enough detail for replication purposes (25%)
- Results
 - Figures + written summary (25%)
- Discussion
 - Discuss expected / explain unexpected findings (25%)
- Appendix
 - Include all Matlab code (15%)

Oral presentations

Delete your initials from one of the two rows (keep in mind that we need at least two presentations on Nov 17 to avoid a 4-hour class on Dec 1).

Nov 17	Dec 1
JA	JA
CD	CD
AG	AG
RM	RM
JR	JR
QW	QW
SL	SL
SM	SM
BR	BR

Spectral analysis of speech

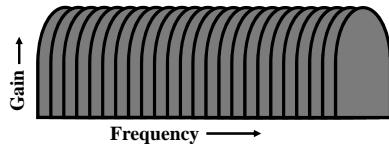
- *Why perform a frequency analyses of speech?*
 - Ear+brain carry out a form of frequency analysis
 - Relevant features of speech are more readily visible in the amplitude spectrum than in the raw waveform

Spectral analysis of speech

- But: the ear is not a spectrum analyzer.
 - **Auditory frequency selectivity** is best at low frequencies and gets progressively worse at higher frequencies.

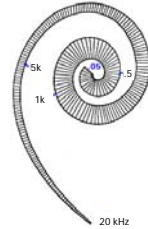
Auditory filters

- Fletcher (1940) suggested that the peripheral auditory system could be modeled as a bank of linear bandpass filters with continuously overlapping center frequencies.



Auditory filters

- Each point along the basilar membrane corresponds to a filter with a different center frequency, with center frequencies increasing roughly logarithmically from the apex to the base.



Mapping frequency to cochlear position

- Greenwood, D.D. (1990). *A cochlear frequency-position function for several species — 29 years later*. **J. Acoust. Soc. Am.** **87**, 2592–2605.

Mapping frequency to cochlear position

- Greenwood devised an almost-exponential function for mapping frequency onto position along the cochlear partition, measured in millimeters from the base.

Mapping frequency to cochlear position

- Greenwood map
$$F = A (10^{ax} - k)$$
 - F is frequency in Hz
 - x is distance in mm.
 - For humans, $A=165.4$, $a=0.06$ and $k=0.88$
 - The human cochlea is about 35 mm in length.

Mapping frequency to cochlear position

- Greenwood map
$$F = A (10^{ax} - k)$$
 - F is frequency in Hz
 - x is distance in mm.
 - For humans, $A=165.4$, $a=0.06$ and $k=0.88$
 - Different sets of constants are used for other mammals (elephant, cat, rat, mouse) with different cochlear lengths

Mapping frequency to cochlear position

- Greenwood map

$$F = A (10^{ax} - k)$$

- In Matlab:

$$F = 165.4 * (10^{(0.06 * x)} - 1);$$

- Inverse map (frequency to millimeters):

$$x = (1/0.06) * \log_{10}((F/165.4) + 1);$$

Exercise: Calculate the distance along the cochlear partition for a 500 Hz tone.

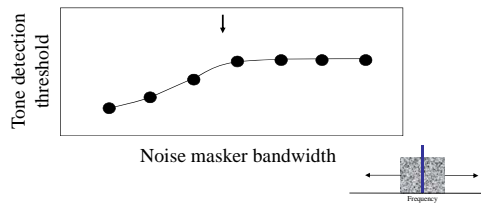
Mapping frequency to cochlear position

- Assumption: auditory critical bands correspond to equal distances along the cochlea.
- What is an auditory critical band?

Critical Bandwidth

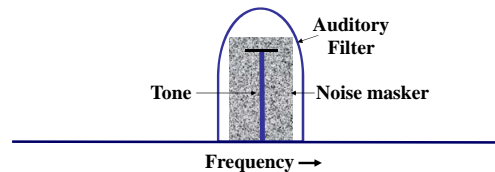
- Fletcher (1940) band-widening experiment

- The threshold for detecting a pure tone in the presence of a bandpass noise masker increases as the noise bandwidth increases, until the width of the band exceeds the *critical bandwidth* of the auditory filter.



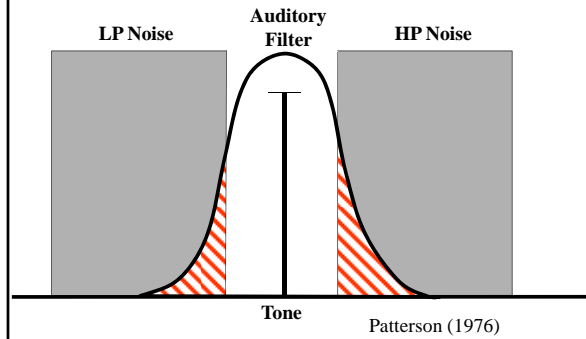
Power spectrum model of masking

- Detection of probe tone in the presence of a noise masker depends on the relative power of probe and noise passed by the auditory filter centered on the tone (Patterson, 1976).



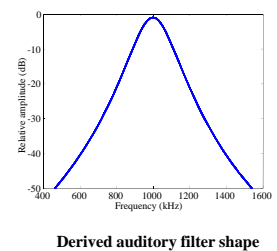
Notched noise method

Measure tone threshold as a function of notch width

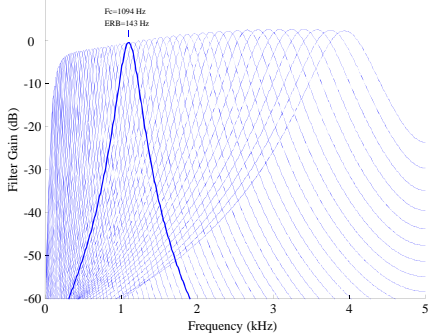


Notched noise method

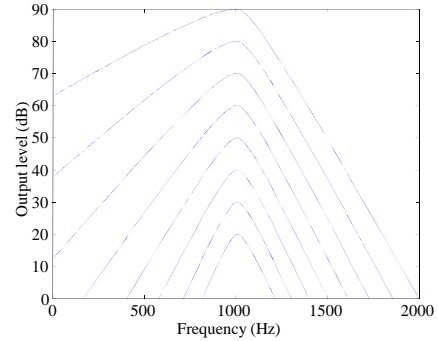
- Patterson (1976) estimated auditory filter shapes from the function relating **tone threshold** to **notch width**.
- The derived filters have a rounded top and steep skirts, with bandwidths 10-15% of filter center frequency.



Auditory filter shapes as a function of frequency

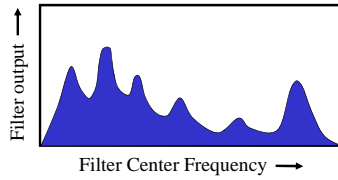


Auditory filter shapes as a function of level

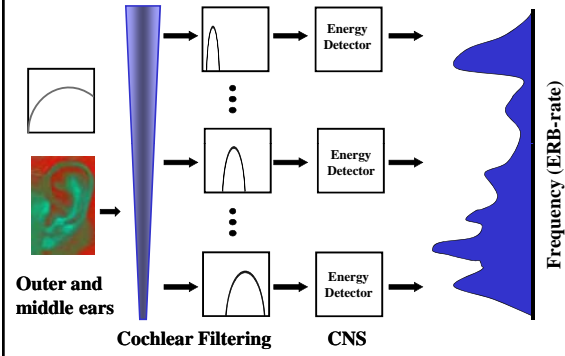


Excitation patterns

- **Auditory excitation patterns** show the composite output of a bank of simulated auditory filters as a function of filter center frequency.



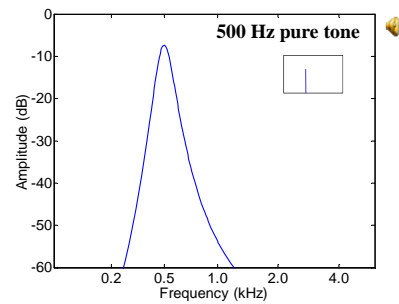
Excitation patterns



Excitation patterns

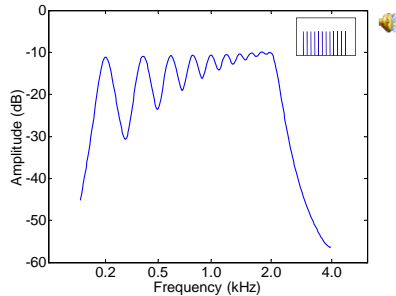
- Excitation patterns provide a good model of auditory frequency selectivity and masking: frequency components that are resolved by the auditory system produce distinct peaks in the excitation pattern.

Excitation patterns

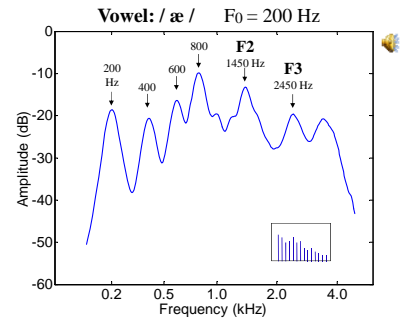


Excitation patterns

Complex tone, equal amplitude harmonics



Excitation patterns

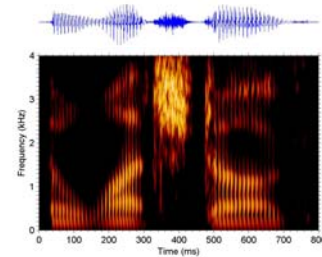


Demo: harmonic synthesis

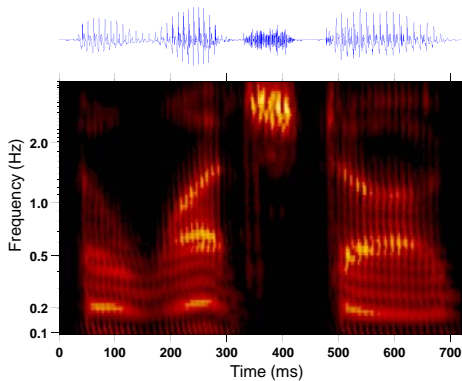
- Additive harmonic synthesis: vowel /i/ 🗣️
- Cumulative sum of harmonics: vowel /i/ 🗣️
- Additive synthesis: “wheel” 🗣️
- Cumulative sum of partials: 🗣️

Speech spectrogram

- *Running amplitude spectra* (codes amplitude changes in different frequency bands over time).



Auditory filterbank spectrogram



Vowels in noise

- Extract a brief segment from near the midpoint of one of the 12 vowels recorded in lab 1.
- For each segment, create a sample of white noise of the **same duration** and **equal rms level** (0 dB SNR).
- Add the noise to the vowel and listen to the mixture. How does white noise at a 0 dB signal-to-noise ratio affect vowel quality?

Vowel representations

- To generate the noise, use the built-in Matlab random number generator, `rand.m`. Since this generates numbers uniformly distributed between 0 and 1, you need to subtract 0.5 to distribute them around zero:
 - » `noise=rand(size(y)) - 0.5;`
- To equate the rms levels of vowel and noise, use the function `rms.m` on the class web page:
 - » `noise = noise .* (rms(y) ./ rms(noise));`

Vowels in noise

- Use the functions `fp.m` and `lpcp.m` to compute FFT and LPC spectra for the vowel alone and the vowel+noise mixture. (Note: `lpcp.m` incorporates 6 dB/oct pre-emphasis).
- What happens to the spectrum when white noise is added? How are the peaks in the LPC spectrum affected by the noise? Are some formant peaks and some vowels affected more than others? Why?

Vowel representations

- Next, compute **auditory excitation patterns** for each vowel alone and vowel+noise mixture.
- Estimate the frequencies of resolved peaks in the excitation pattern. At low frequencies, these will correspond to **harmonics** of the fundamental. At higher frequencies, they may represent **formants** or clusters of formants.
- Compare the excitation patterns in quiet and in noise. What happens to the peaks?

Excitation patterns

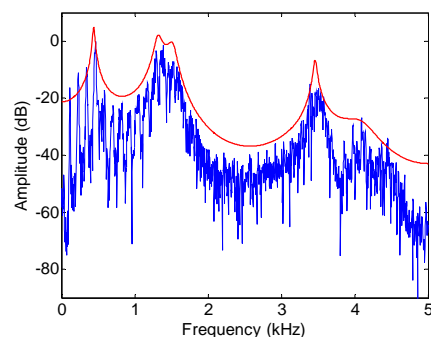
- To equate the rms level of the vowel and noise, use the function `rms.m` on the class web page:
<http://www.utdallas.edu/~assmann/hcs7367/rms.m>
- To calculate the excitation patterns, use the function `gtep.m`:
<http://www.utdallas.edu/~assmann/hcs7367/gtep.m>
- You'll also need Malcolm Slaney's Auditory Toolbox:
<http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010/AuditoryToolbox.zip>
<http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010/>

Excitation patterns

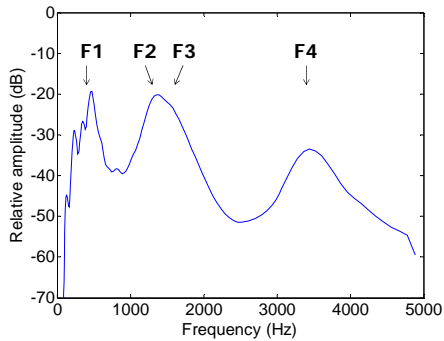
```
>> help gtep
gtep: Gammatone filter bank excitation pattern
Usage: [cf,expat]=gtep(y,rate,nfilt,flo,fhi,absflag);
cf: filter center frequencies on ERB-rate scale
expat: excitation pattern
y: waveform vector
rate: sample rate in Hz (default 10000 Hz)
nfilt: number of filter channels (default 128)
flo, fhi: CF limits (default 80 and rate/2=5000 Hz)
absflag: set to 0 if absolute threshold compensation is
NOT desired. (default 1)
See also: MakeERBFilters, ERBFilterBank
```

LPC and FFT spectra

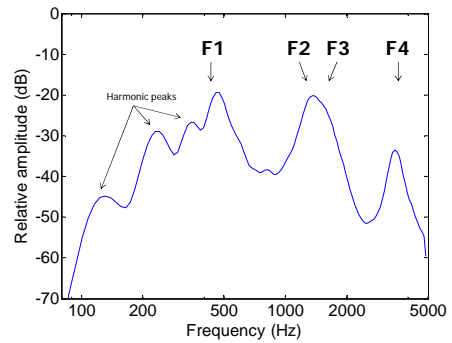
"herd"



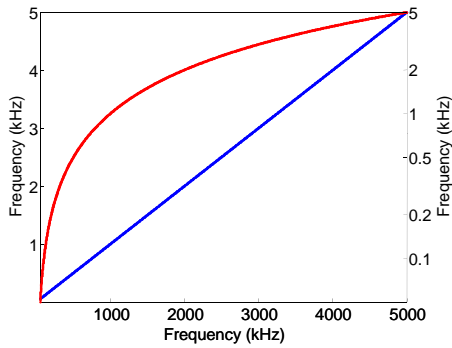
Excitation pattern



Excitation pattern: log frequency scale



Log scale



ERB-rate scale

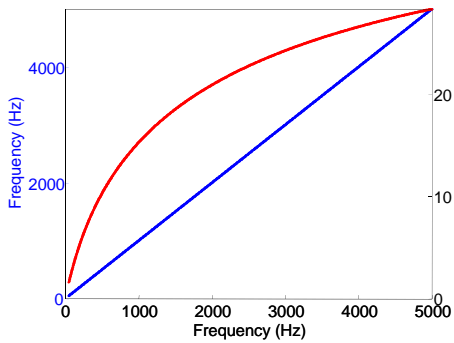
- Moore and Glasberg (1989) auditory model
- ERB: Equivalent Rectangular Bandwidth
- ERB units provide approximately equal distances along the basilar membrane

$$E = 16.7 \log_{10} (1 + f / 165.4)$$

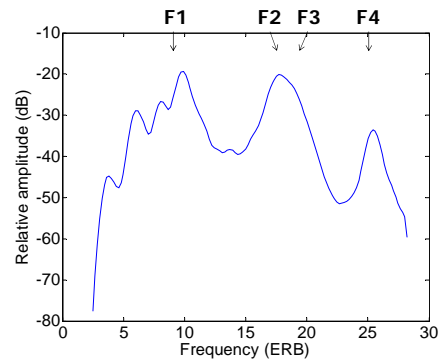
and

$$f = 165.4(10^{0.06E} - 1)$$

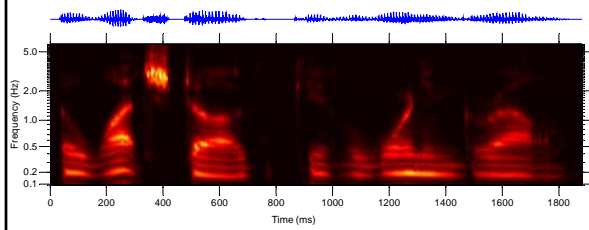
ERB-rate scale



Excitation pattern: ERB scale

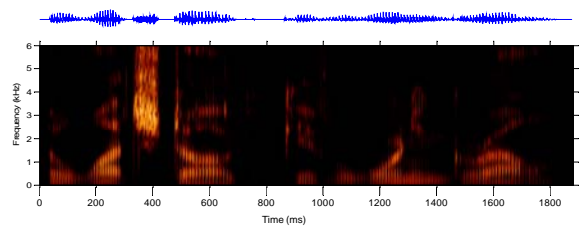


Gammatone filter bank spectrogram



- Low-frequency features represent harmonics >> `gtsp(y,rate,128);`
- High-frequency features represent formants >> `colormap(hot);`

FFT spectrogram



Correlogram representation

- `acgmovie_demo`

