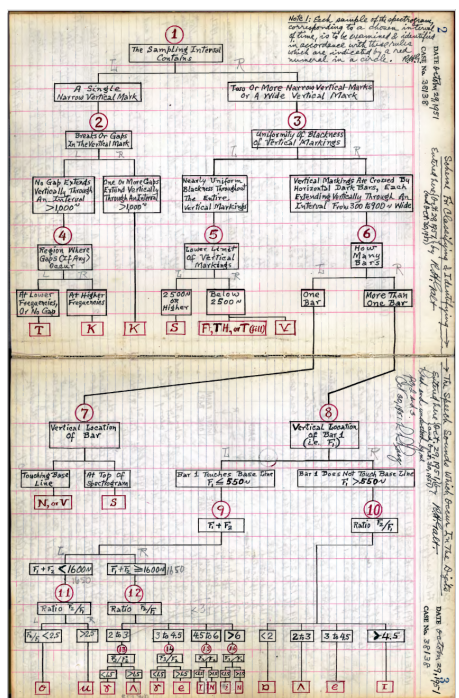
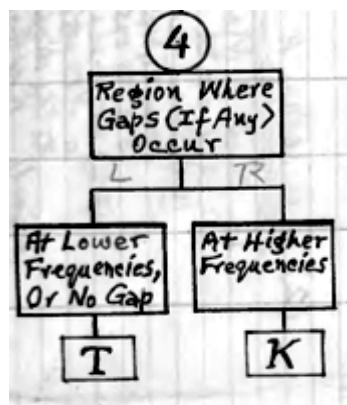
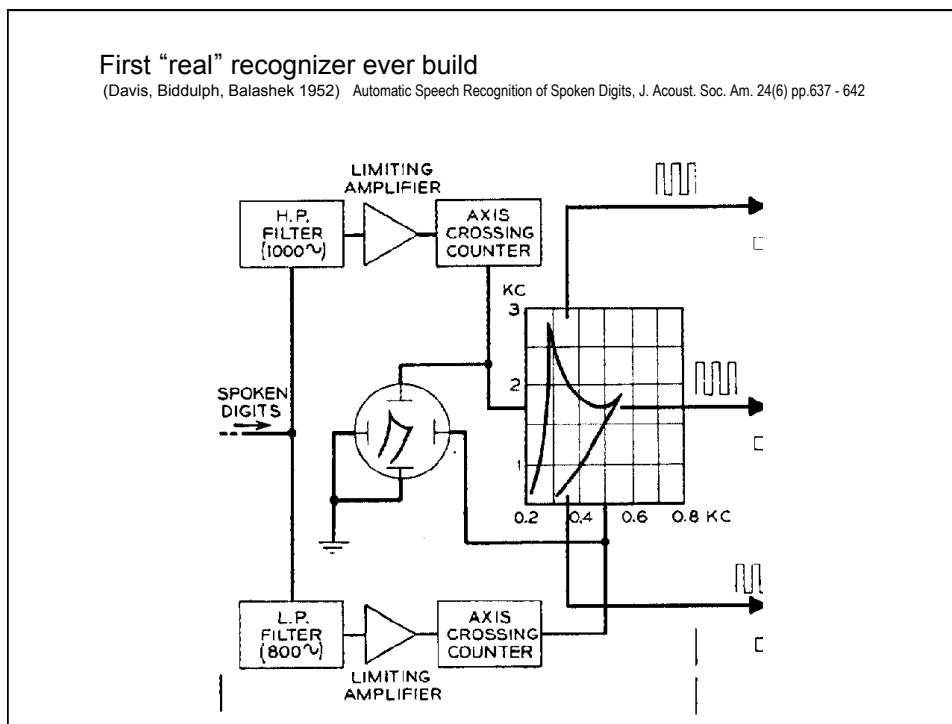


Concept of the first "real" automatic speech recognizer (R.H. Galt 1951)



First "real" recognizer ever build

(Davis, Biddulph, Balashek 1952) Automatic Speech Recognition of Spoken Digits, J. Acoust. Soc. Am. 24(6) pp.637 - 642



Note 1 (4/23/52): I called attention to the fact that in my tests, one voice was that of Balashek. The number of errors was either 2 or 1, the success score thus being either 97.5% or 98.7%. [The uncertainty was because of the use of the 1115 box in one test.] Thus my rules were almost or equally as good, with Balashek's voice, as the rules governing S. Peterson's machine. However, my rules were almost as effective for a considerable range of voices; whereas for several music voices, the machine gave scores perhaps as low as 65%.

100

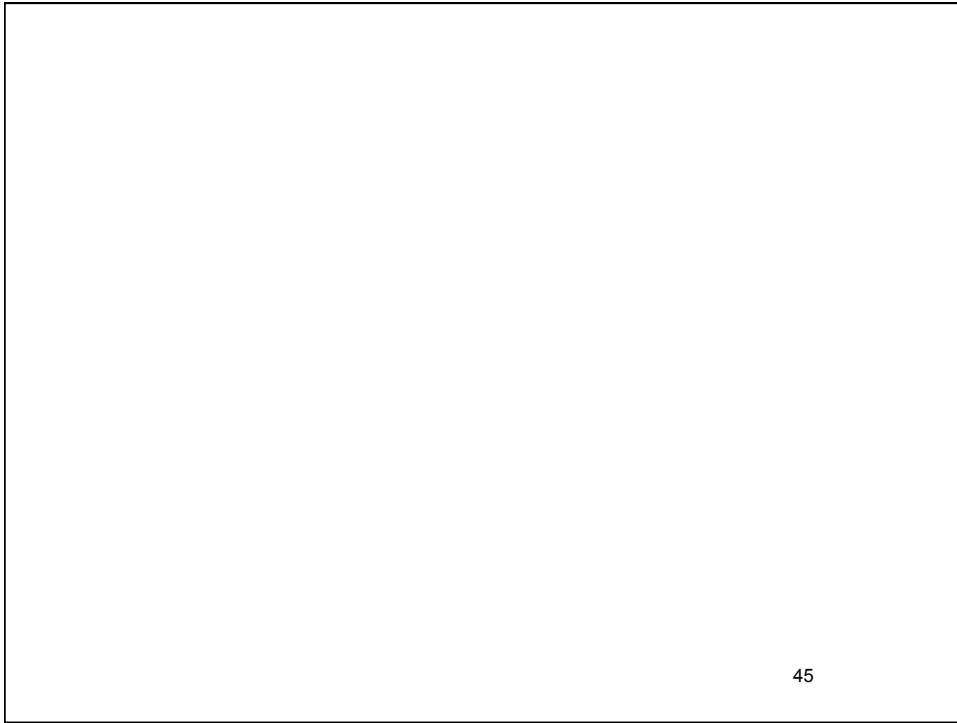
DATE 38138

CASE No. 4/29/52

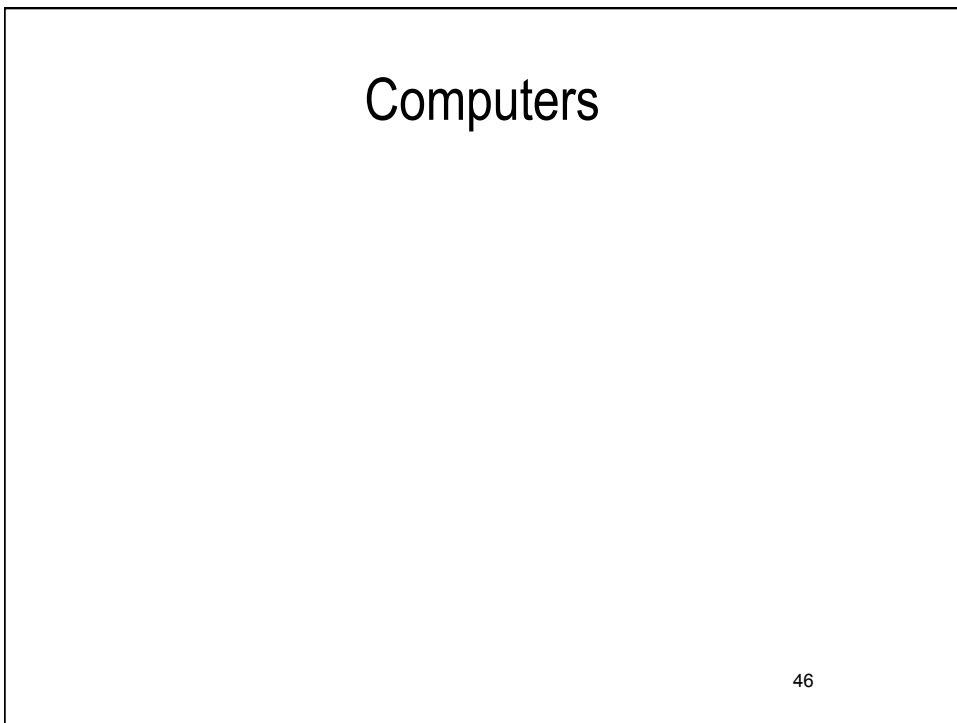
Conference 4/29/52 with Karlin

Karlin reported conclusions reached in his recent conference with Potter.

1. The V.O.T. job will not concern me any longer. This job as an exercise in simulation, will be one of the jobs in Kerster's hands.

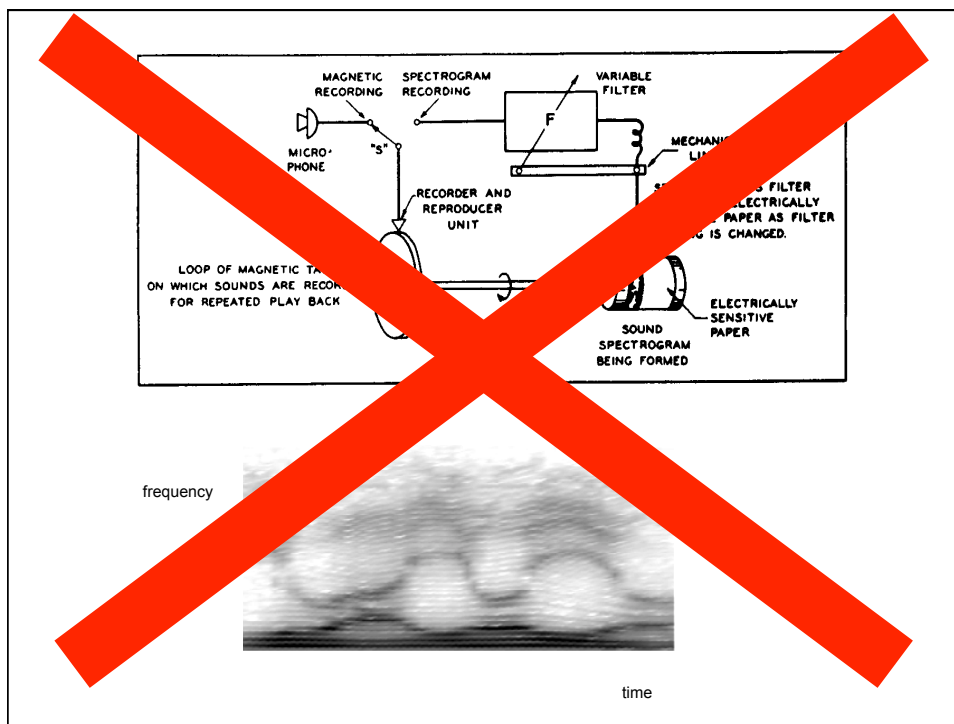


45



Computers

46



T_0

- One of Fourier ideas
 - Describe a periodic signal by an (infinite) sum of other well defined periodic signals

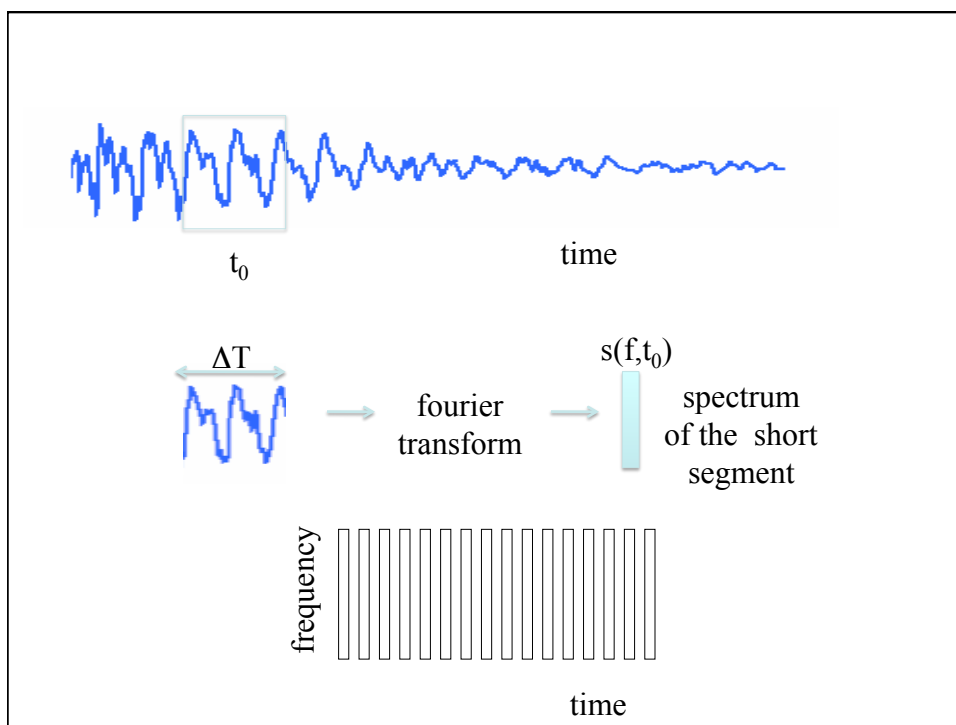
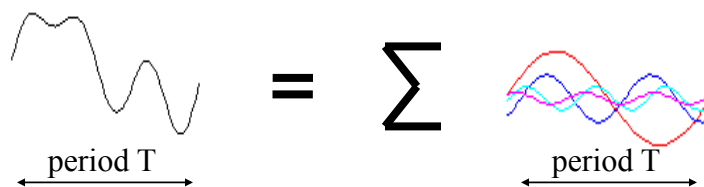
Joseph Fourier
(1768-1830)
Student of Lagrange
Adviser of e.g. Dirichlet or Navier

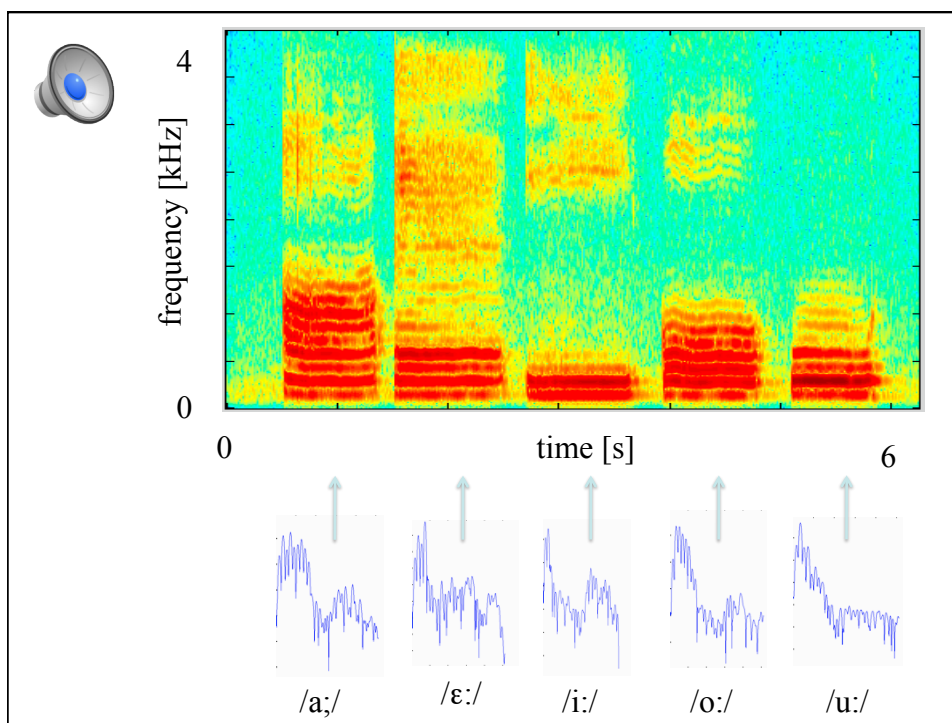
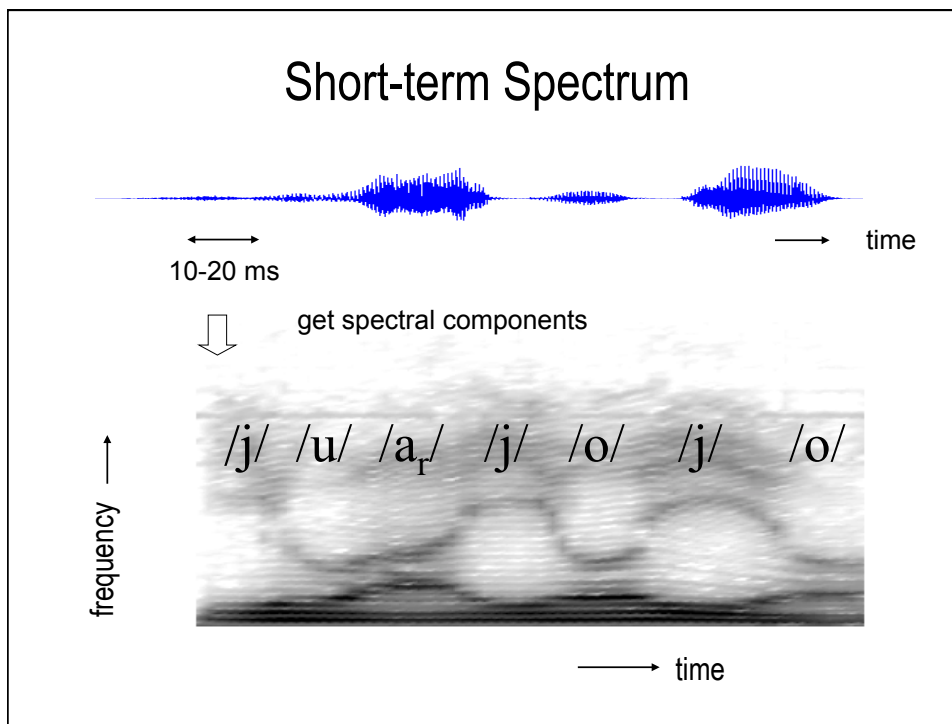
T

$= \sum$

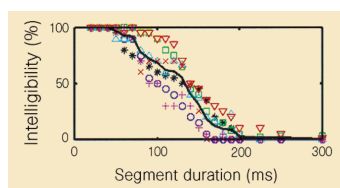
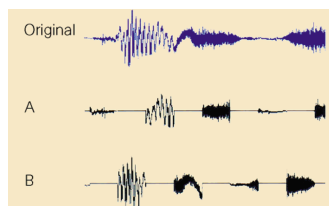
T

Fourier's Idea





Ignore short-term phase: Locally reversed speech



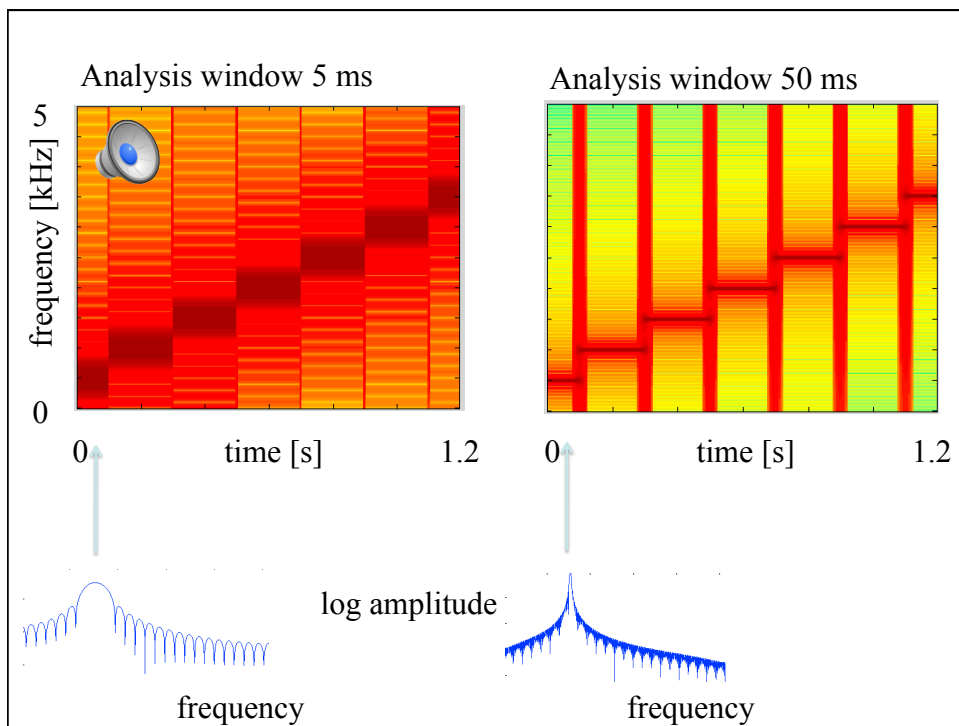
Nature 398, 760 (29 April 1999)
Cognitive restoration of reversed speech
Kouros Saberi and David R. Perrott

Uncertainty Principle

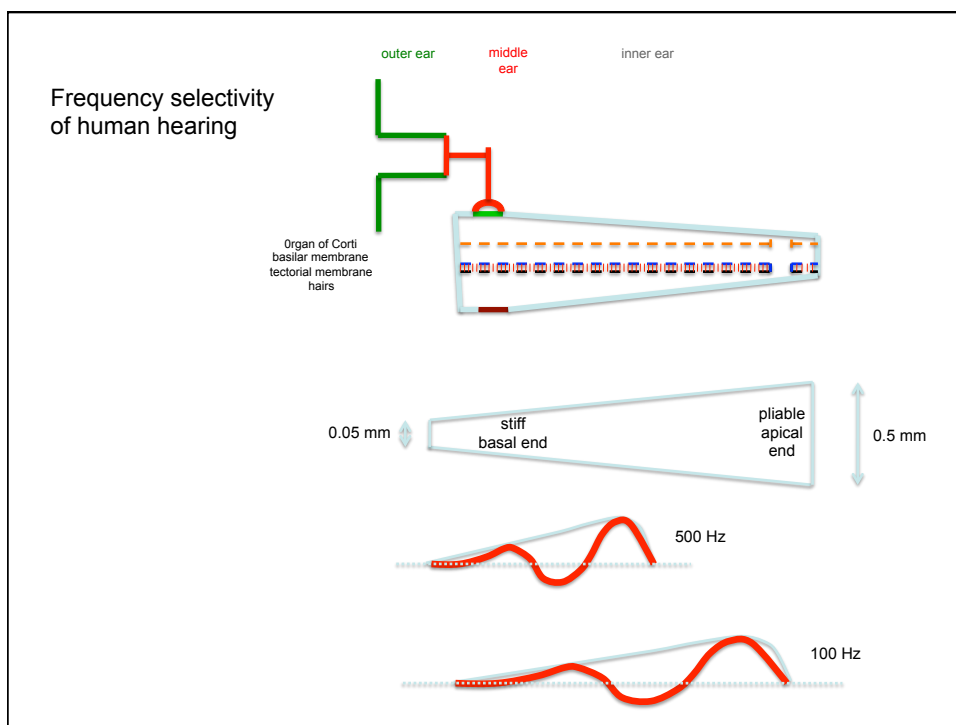
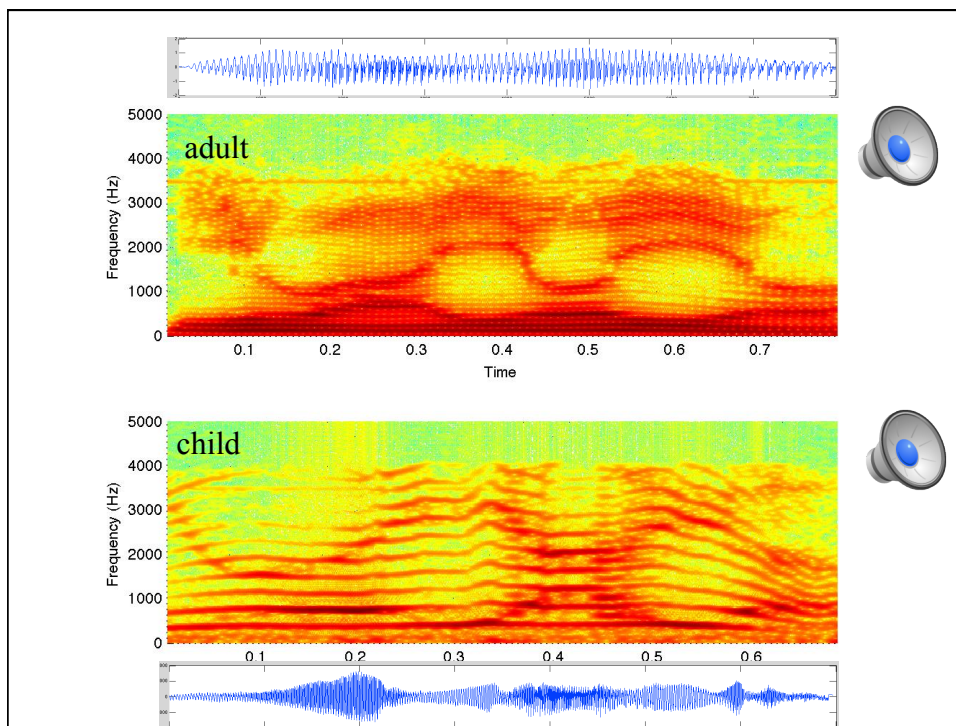
You cannot simultaneously know
the exact frequency composition
and the exact temporal locality
of an acoustic event

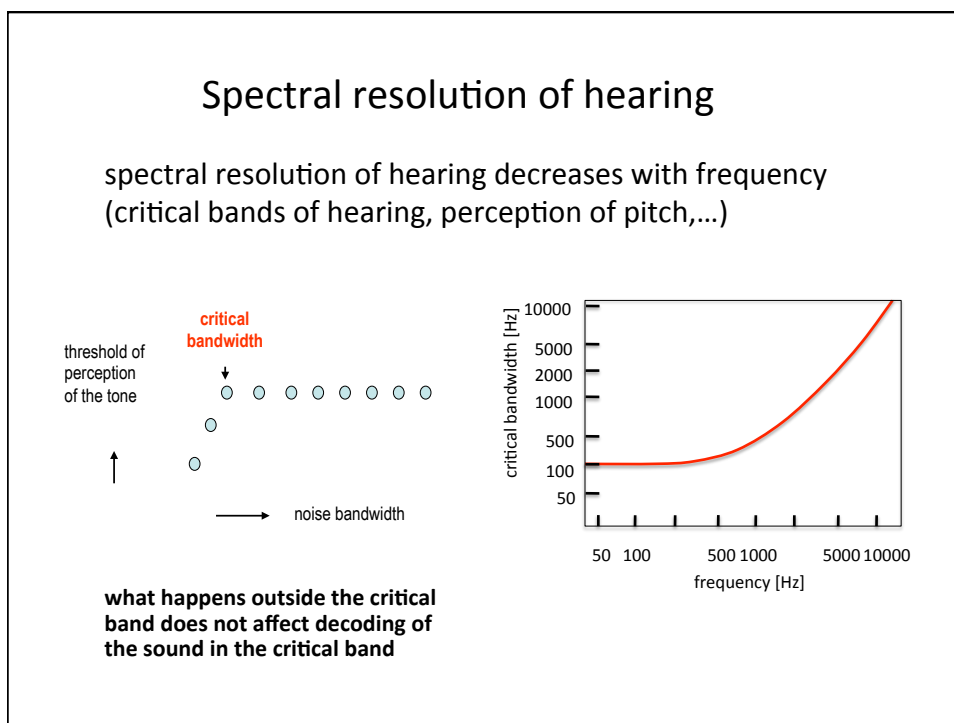
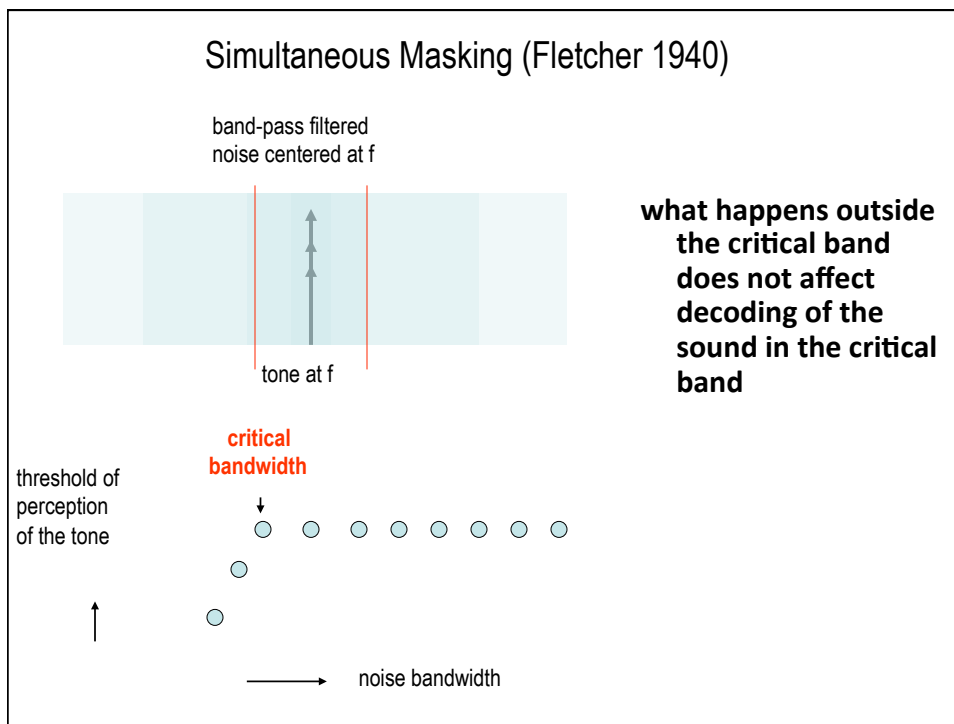
Dennis Gabor 1947

Higher spectral resolution needs longer
analysis windows and hence yields lower
temporal resolution

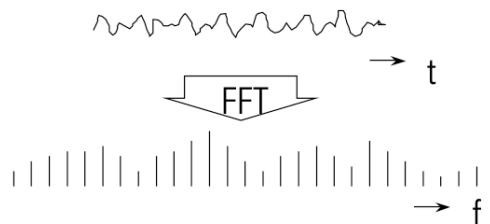


Different Speakers

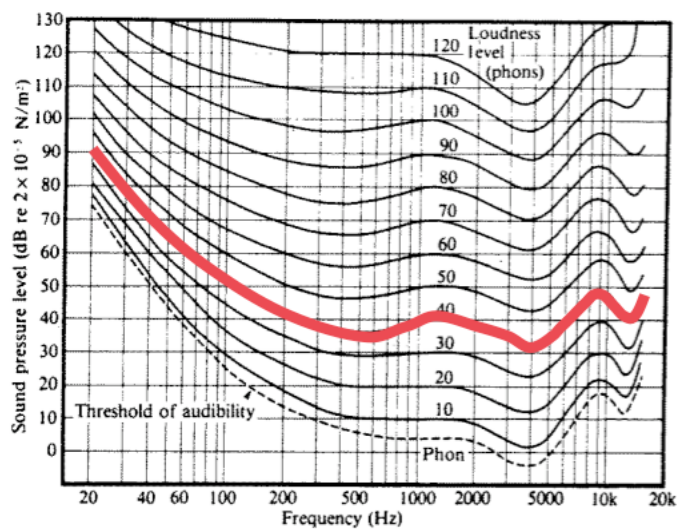


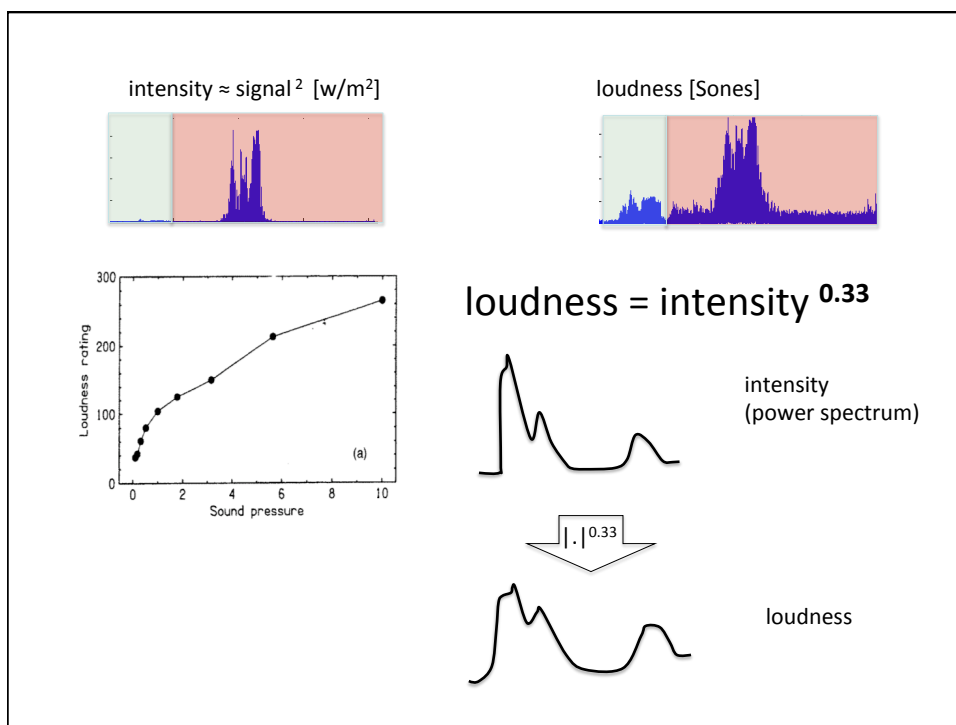
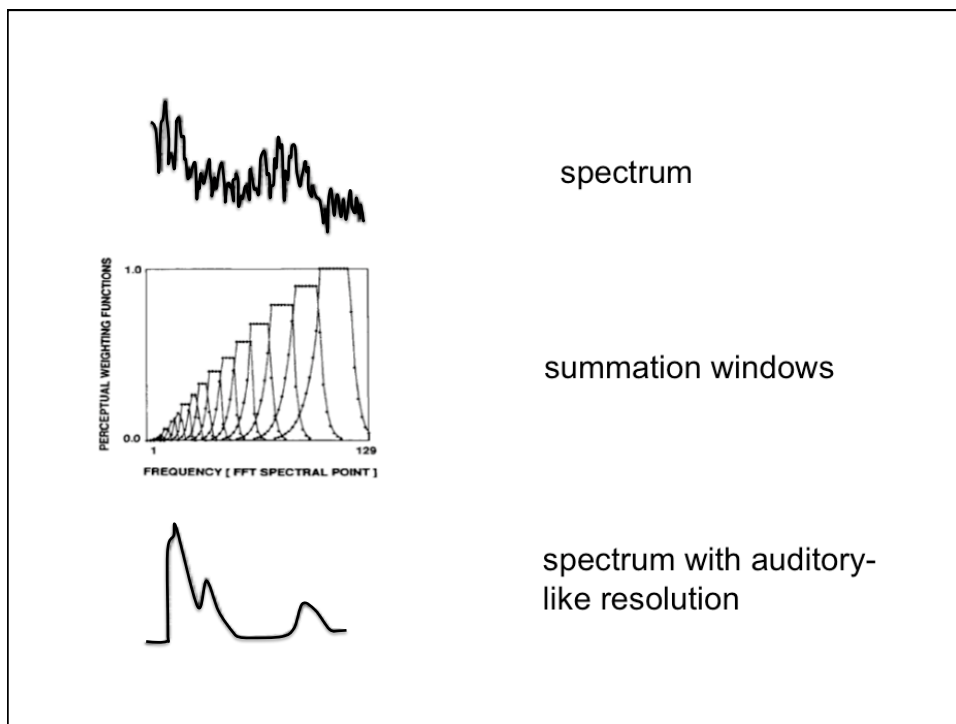


“Poor man’s”
way of emulating
critical-band
filter



Sensitivity of hearing depends on frequency

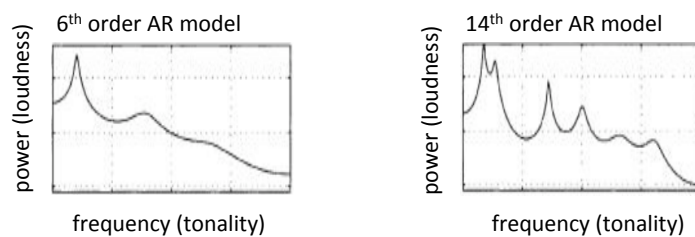




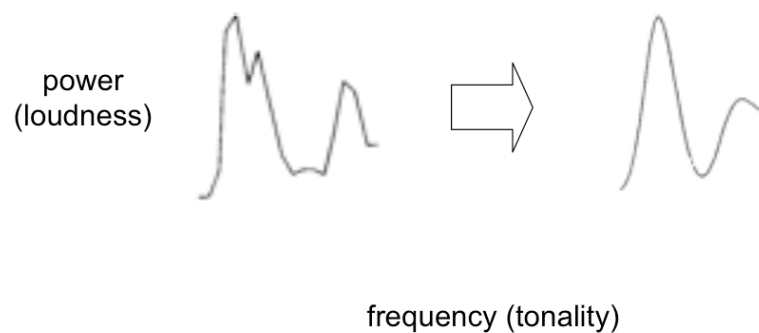
Not all spectral details are important

a) compute Fourier transform of the auditory spectrum and truncate it (cepstrum)

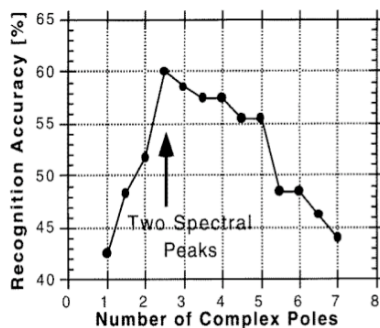
b) approximate the auditory spectrum by an autoregressive model



Perceptual Linear Prediction (PLP) Autoregressive fit to the auditory-like spectrum



Optimal Amount of Spectral Smoothing (order of PLP autoregressive model)



- cross-speaker ASR (trained on one speaker and tested on another)
- all speaker-dependent information harmful

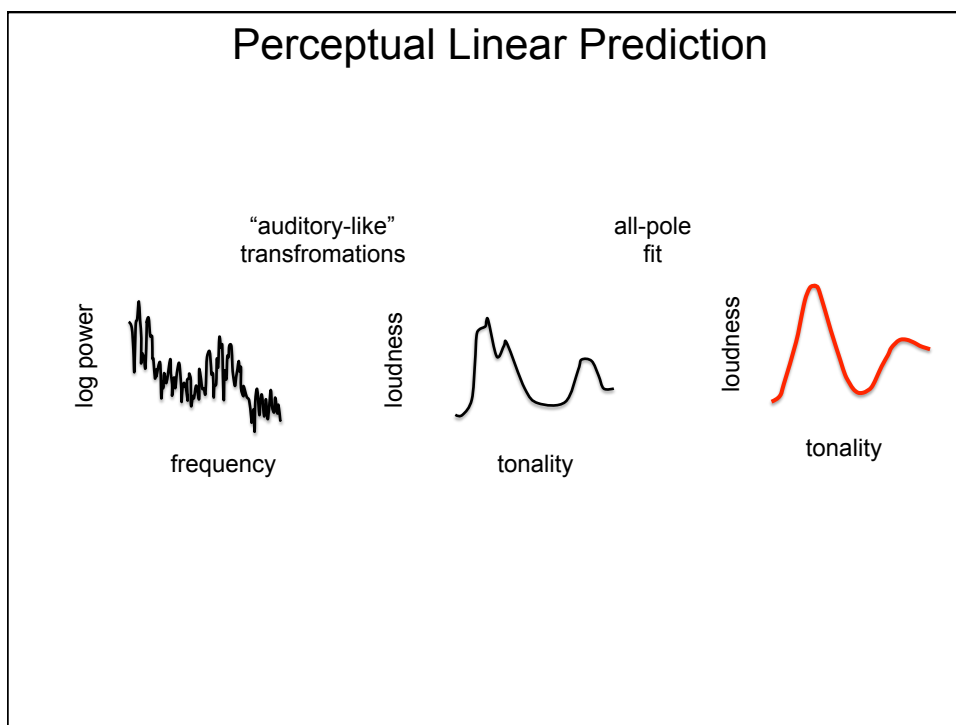
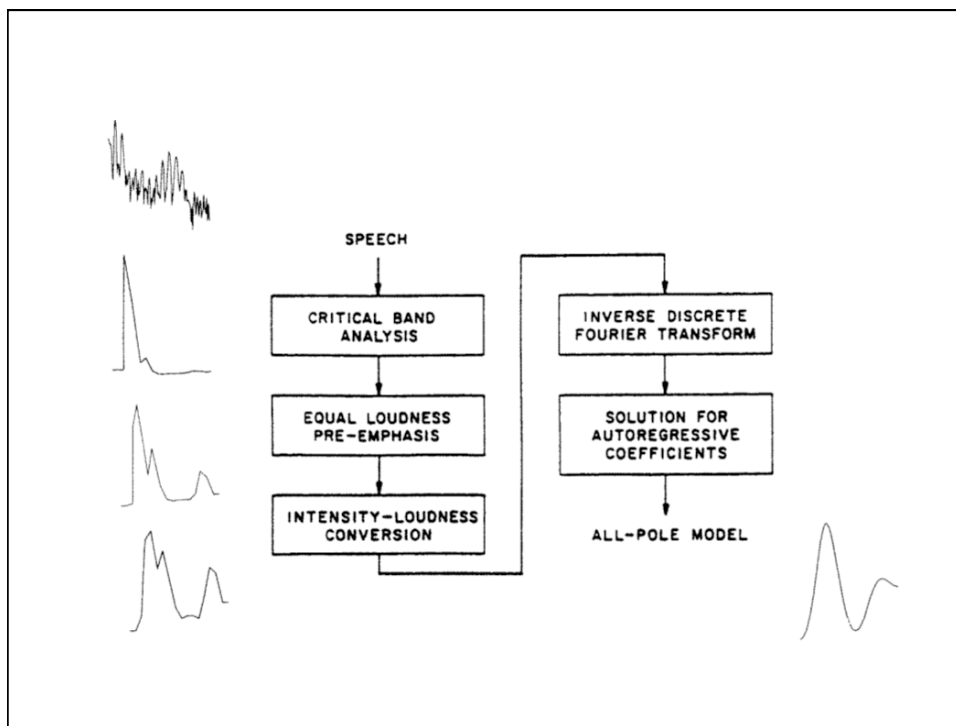
Perceptual Linear Prediction

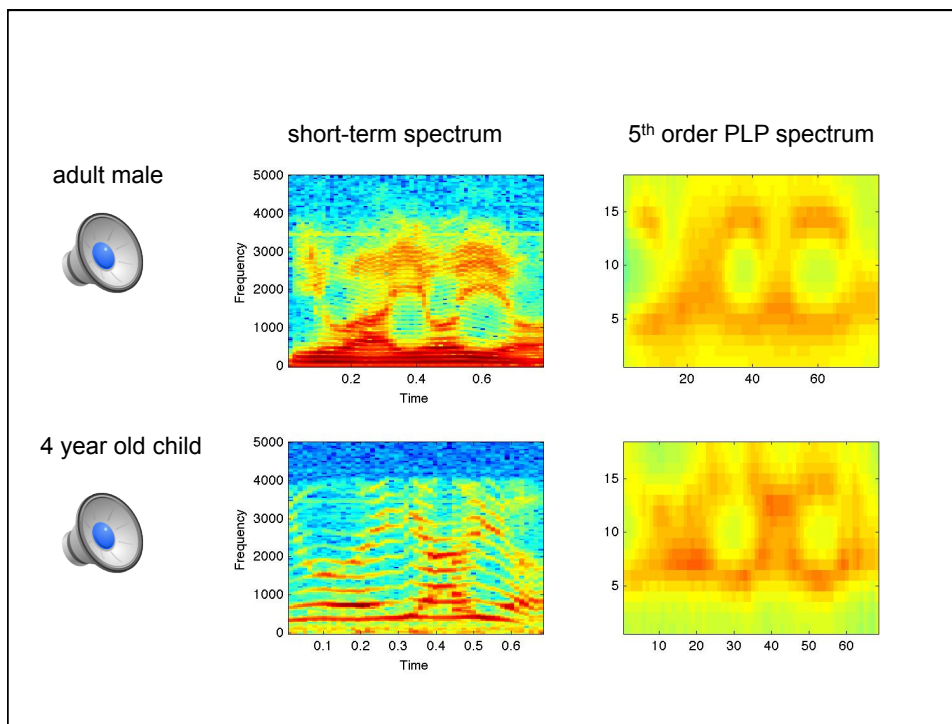
Limited spectral resolution

formant clusters as may be interpreted by auditory perception

Perceptual Linear Prediction (PLP)

critical-band (Bark) spectral analysis
 loudness domain (cubic root of intensity)
 equal loudness curve (at 40 dB)
 autoregressive spectral fit (fits well at peaks)





H6. Effect of the spectral model order in automatic speech recognition.
Kazuhiro Tsuga and Hynek Hermansky (Speech Technology Laboratory, 3888 State Street, Santa Barbara, CA 93105)

It has been observed that in speaker-independent multi-template digit recognition, the 5th-order perceptually based LP (PLP) analysis method yields about 40% lower error rates than does the standard 14th-order LP.

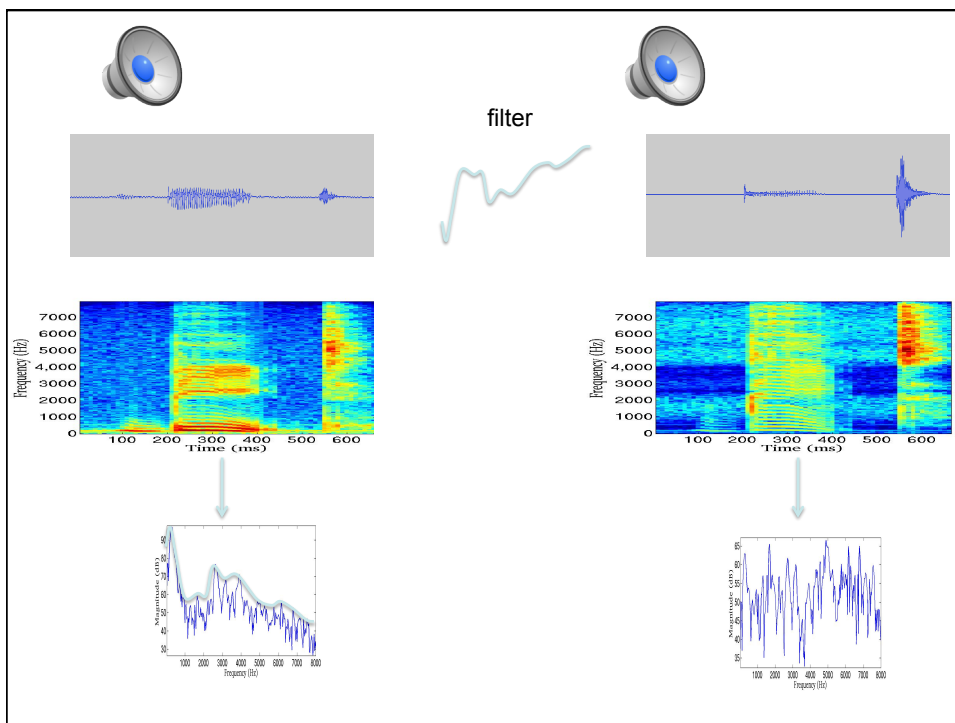
J. Acoust. Soc. Am. Volume 80, Issue S1, pp. S18-S18 (1986)

RE: TWO INVENTIONS BY DR. H. HERMANSKY ET AL
 WE HAVE CAREFULLY STUDIED WHETHER TO FILE U.S. PATENT APPLICATIONS FOR THESE INVENTIONS. AS MR. NYUJI TELEPHONED YOU, OUR CONCLUSION IS:

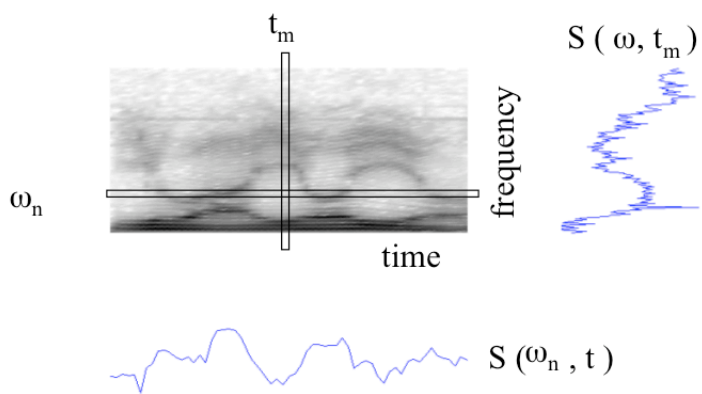
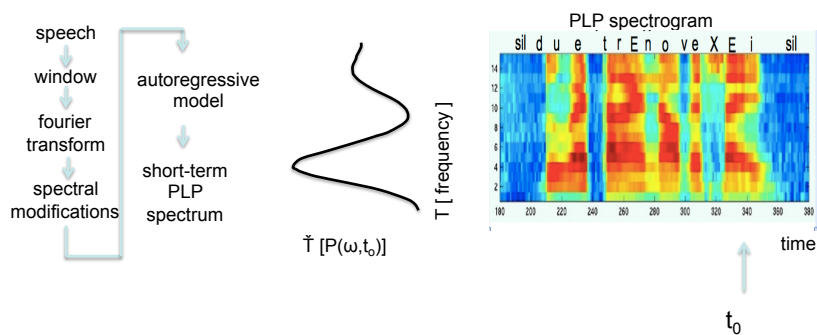
(1) ''PERCEPTUALLY BASED LINEAR PREDICTIVE ANALYSIS OF SPEECH''.
 WE CANNOT SEE ANY PRESENT OR FUTURE PRODUCTS TO WHICH THIS INVENTION IS PRESUMED TO BE APPLIED. SO, THIS INVENTION DOES NOT HAVE ENOUGH PRACTICAL VALUE TO BE APPLIED FOR A U.S. PATENT.

Different Communications Channels

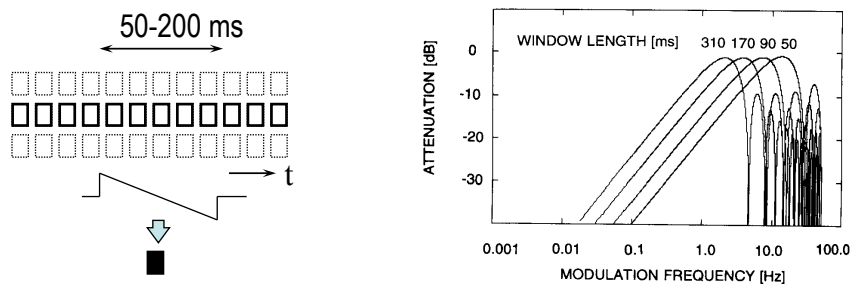
73



Short-Term Features



Delta Features



- linear combination of several short-term features
- equivalent to FIR filtering of feature trajectory
- selective band-pass with 6 dB/oct slope

