PROBLEM

Automatic Speech Recognition performance degrades in noise.

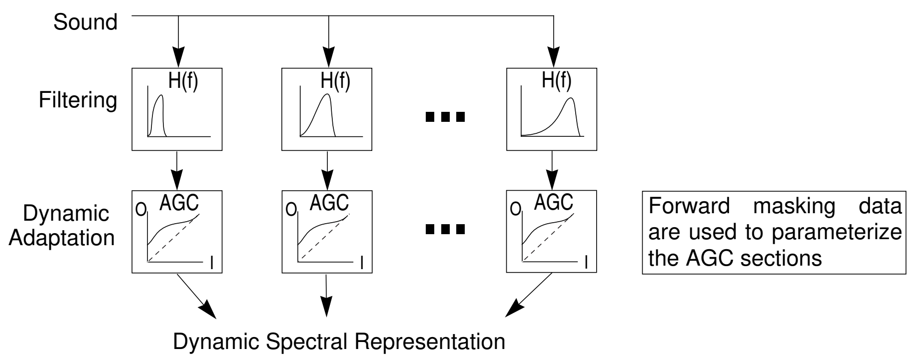
APPROACH

Obtain robust spectral features from an auditory model that emphasizes transitions and enhances spectral peaks.

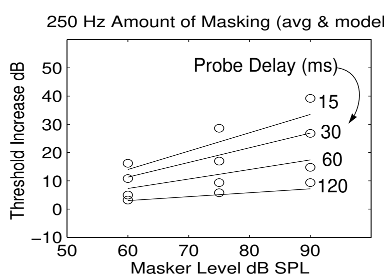
Evidence of Dynamic Audition: Physiological and Perceptual

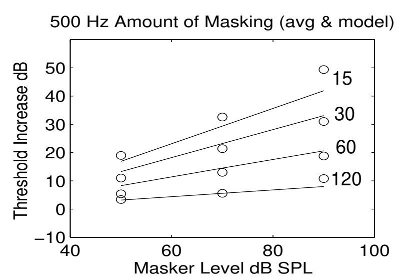
adaptation; sensitivity to onsets and transitions; forward masking

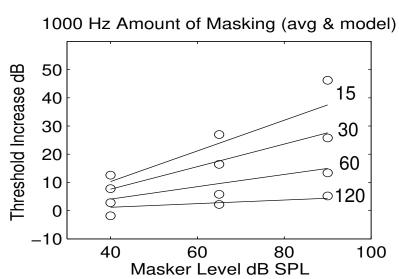
Model Overview

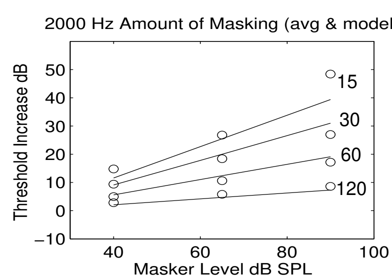


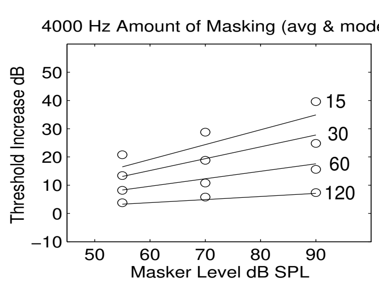
Forward Masking Data to Parameterize Adaptation





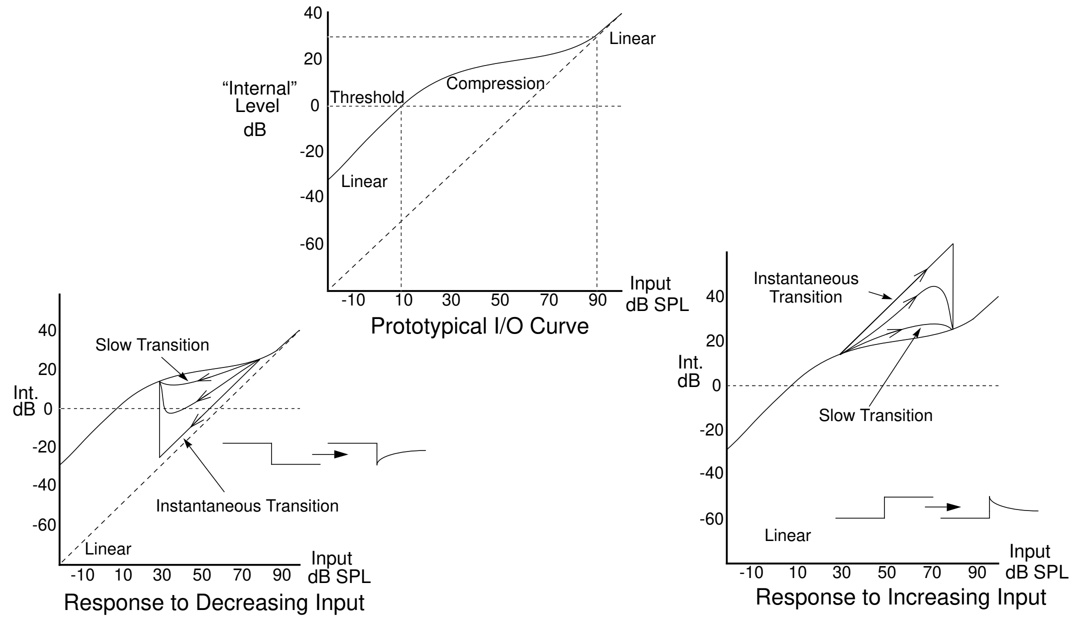




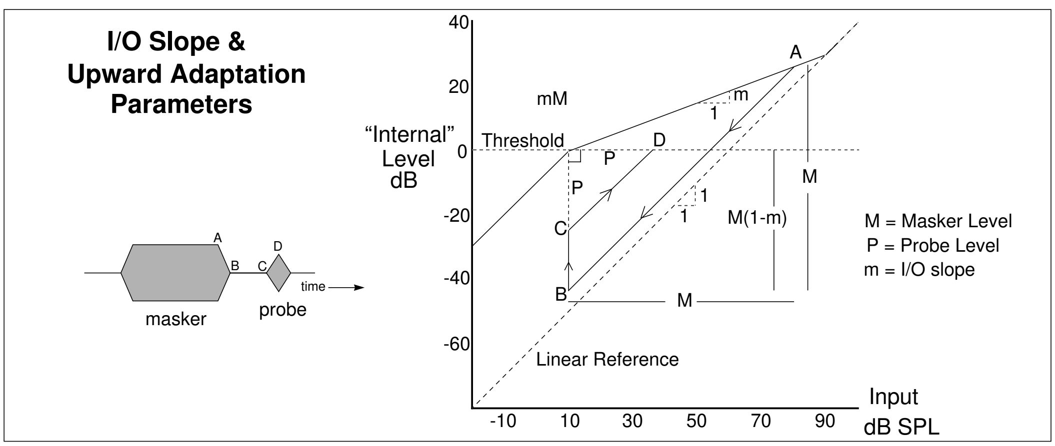


Amount of masking as a function of the masker level with contours of constant probe delay (circles), together with the model prediction of the amount of forward masking

Adaptation Stage Functionality

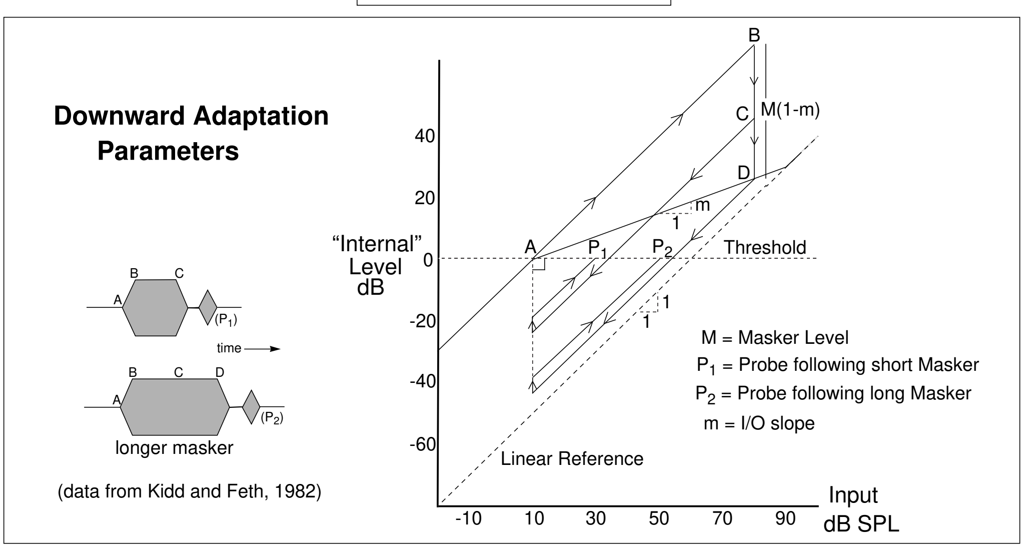


From Forward Masking Data to Model Parameters



$$P = M(1-m)\,a^{n}$$

a = upward incremental adaptation,
n = probe delay (relative to sample rate)



$$\Delta P = M(1-m)\,b^{nd}a^{nu}$$

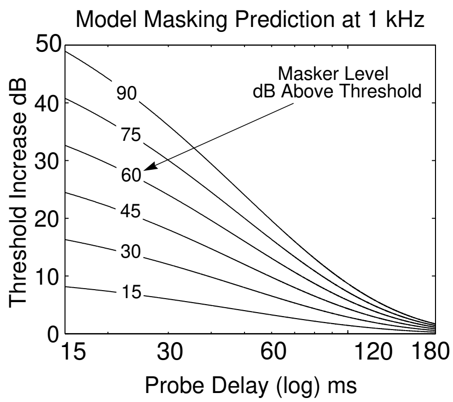
b = downward incremental adaptation,
nu = probe delay,
nd = short masker duration

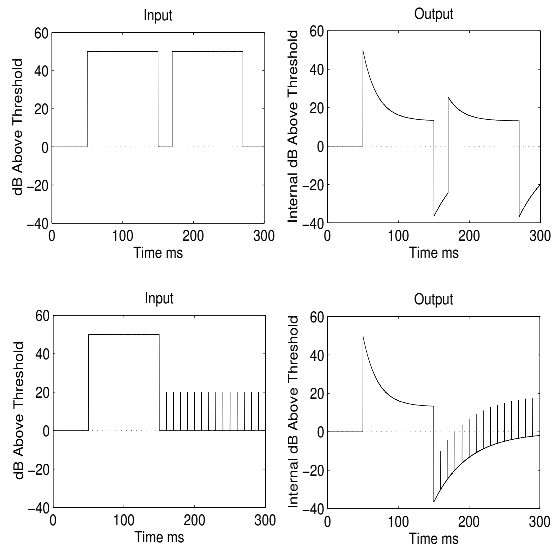
Model Parameters

Freq. Hz	Slope *m*	*a*	*b*	release (ms)	attack (ms)
250	0.19	0.864	0.474	68	13
500	0.20	0.854	0.510	63	15
1000	0.26	0.816	0.543	49	16
2000	0.29	0.851	0.525	62	16
4000	0.34	0.858	0.507	65	15

The model parameters a and b are with reference to a 100 Hz spectral sample rate (10 ms spectral sample interval).

Predictions, and Example Responses





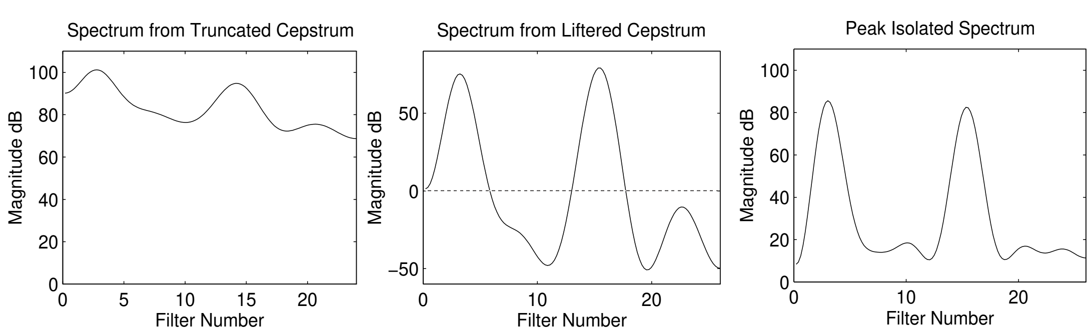
Spectral Sharpening: Physiological and Perceptual Evidence

two-tone suppression; lateral inhibition; Klatt's synthesizer experiments

Direct Approach: Cepstral Bandpass Lifter

raised-sin lifter
removes DC level of log spectrum; de-emphasizes slow and fast changes with frequency
no effect for a diagonal covariance continuous density estimation; affects distance metric only

Keep Only the Peaks

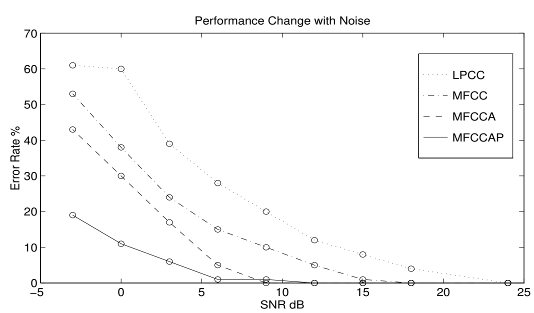


filter output ➛ cepstrum ➛ lifter ➛ spectrum ➛ half-wave rectify ➛ cepstrum

Word Recognition Evaluations: Does it help in noise?

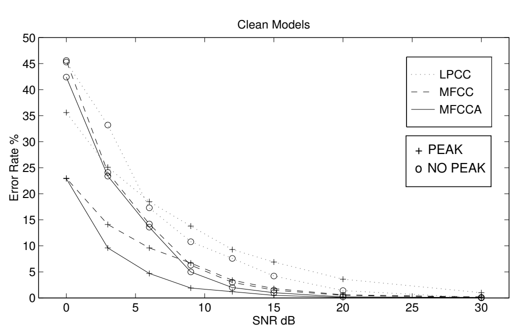
DTW System

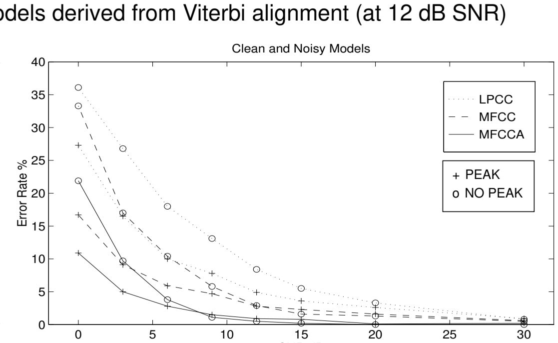
discrete digit isolation and recognition task; single speaker; isolated clean tokens for templates



HMM System

discrete digit isolation and recognition task; data from TI-46; 8-male speakers; speaker-independent
6-state models; continuous-density, single Gaussian; diagonal covariance; fixed global variance
training from isolated digits; Viterbi initialization followed by full forward-backward estimation
first evaluation with clean training data
second evaluation with clean and noisy models derived from Viterbi alignment (at 12 dB SNR)





A Model of Dynamic Auditory Perception and its Application to Robust Word Recognition

Brian Strope and Abeer Alwan

Department of Electrical Engineering, University of California at Los Angeles
This work was supported in part by NIDCD Grant No. 1 R29 DC 02033-01A1.

bps@ucla.edu ***http://delphi.icsl.ucla.edu***
