

AD-A195 583

BINAURAL MASKING: AN ANALYSIS OF MODELS (U) CENTRAL INST 1/1
FOR THE DEAF ST LOUIS MO B H GILKEY 13 APR 88

AFOSR-TR-88-0660 AFOSR-86-8298

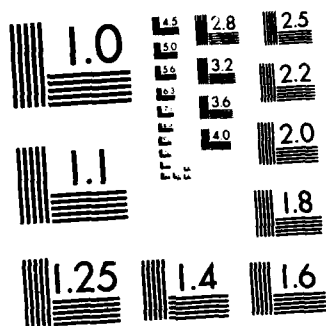
UNCLASSIFIED

F/G 6/4

NL



END
DATE
MAY
98



MICROCOPY RESOLUTION TEST CHART
 NATIONAL BUREAU OF STANDARDS-1963-A

REPORT DOCUMENTATION PAGE

AD-A195 503

ELECTE
JUN 29 1988

1b. RESTRICTIVE MARKINGS

DDC FILE COPY

3. DISTRIBUTION / AVAILABILITY OF REPORT

Approved for public release; distribution unlimited.

4b. DECLASSIFICATION / CONTINUATION SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

HS

5. MONITORING ORGANIZATION REPORT NUMBER(S)

AFOSR-TR- 88 - 0660

6a. NAME OF PERFORMING ORGANIZATION

Central Inst. for the Deaf

6b. OFFICE SYMBOL
(if applicable)

7a. NAME OF MONITORING ORGANIZATION

AFOSR/NL

6c. ADDRESS (City, State, and ZIP Code)

818 South Euclid Avenue
Saint Louis, Missouri 63110

7b. ADDRESS (City, State, and ZIP Code)

Bldg. 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

AFOSR

8b. OFFICE SYMBOL
(if applicable)

NL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

AFOSR-86-0298

8c. ADDRESS (City, State, and ZIP Code)

Bldg. 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2313	A6	

11. TITLE (Include Security Classification)

Binaural masking: An Analysis of models

12. PERSONAL AUTHOR(S)

Robert H. Gilkey

13a. TYPE OF REPORT

Annual

13b. TIME COVERED

FROM 7/15/86 TO 3/14/88

14. DATE OF REPORT (Year, Month, Day)

Apr 15, 1988

15. PAGE COUNT

13

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The goal of this project is to specify the transformations used by the auditory system in order to determine the presence of the signal in an auditory masking task, with particular emphasis on the role of processes that compare information in the frequency domain and in the time domain. Classical models that restrict analysis to a single frequency band and a single temporal window are evaluated. In addition, the role of pattern or "profile" analysis in auditory processing is being assessed by fitting more complex models to the data. The results show that similar cues determine performance in monaural and binaural masking tasks. Information remote from the signal in frequency can be used to overcome the effects of uncertainty about the stimulus level. Masking noise which does not overlap with the signal in time can either improve or degrade the detectability of the signal, depending upon the interaural phase relations of the masker and the signal. The analyses yield a quantitative description of processes that compare information within the frequency band and temporal window that contain the signal to information in other spectral/temporal regions. Other significant results include a more complete description of internal noise processes →

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT

UNCLASSIFIED/UNLIMITED SAME AS RPT D.C. USERS

21. ABSTRACT SECURITY CLASSIFICATION

U

22a. NAME OF RESPONSIBLE INDIVIDUAL

Dr. John Tangney

22b. TELEPHONE (Include Area Code)

(202) 767-5021

22c. OFFICE SYMBOL

NL

A

Concl'd
↓

19. and evidence that the external masker is not cancelled by the binaural processor.



B

AFOSK-TK- 88 - 0660

BINAURAL MASKING: AN ANALYSIS OF MODELS

Robert H. Gilkey
Central Institute for the Deaf
818 South Euclid Avenue
Saint Louis, Missouri 63110

April 15, 1988

Annual Technical Report for Period 15 July 1986 to 14 March 1988

Prepared for

Department of the Air Force
Air Force Office of Scientific Research (AFSC)
Bolling Air Force Base, DC 20332-6448

88 6 29 045
~~88 6 29 045~~

narrow frequency band centered around the signal (i.e., the critical band) and only within the brief temporal window that contains the signal. In this project we use a variety of approaches to demonstrate that these classical models are oversimplifications, to develop models that provide a more accurate description of the responses of the subject, and to delineate the relationships between the mechanisms underlying monaural and binaural masking.

III. STATUS OF THE RESEARCH

Our research on monaural masking and binaural masking proceeds in parallel, with considerable interdependence. Additional support for this research has been provided by grants from NSF (BNS-85-11768) "Analysis of models of auditory masking," period of support July 15, 1985 through July 14, 1988, R.H. Gilkey, PI, and (BNS-87-20305) "Analysis of models of auditory masking," period of support January 1, 1988 through December 31, 1988, R.H. Gilkey, PI.

Profile analysis in noise

In an experiment on diotic masking we investigated the effect of randomizing the overall level of the stimulus on the detectability of a relatively brief tonal signal as a function of the duration and the bandwidth of the masking stimulus. If the subject bases his decision on the energy in a single critical band and a single temporal integration window, it should be possible to disrupt his performance by randomizing the overall level of the stimulus (thus randomizing the energy in that critical band and temporal integration window). The approach is based on the profile analysis experiments of Green (*Am. Psychol.* 38: 133-142, 1983). Here, however, the background is random noise rather than a tone complex. When the masker is narrowband and short in duration, there is a substantial effect of randomizing overall level. However, when the masker is wideband and long in duration, the effect of randomizing overall level is negligible. Apparently, some information present in the spectral fringe (that portion of the masker that does not overlap with the signal in the frequency domain) and temporal fringe (that portion of the masker that does not overlap with the signal in the temporal domain) can be used to overcome the effects of randomizing overall level. When the bandwidth and the duration of the masker are manipulated separately, the results suggest that the information in the frequency domain is most important. These results call into question the classical critical band interpretation of auditory masking. The results were described in talks presented at CID and at the meeting of the AFOSR program in Auditory Pattern Recognition and are included in Gilkey (in W.A. Yost and C.S. Watson (eds.), Auditory Processing of Complex Sounds, 26-36, 1987).

An experiment analogous to that of Gilkey (1987) was conducted using the dichotic NOS_n stimulus configuration. Again, the effect of randomizing overall level was much smaller under

the wideband long duration condition than under the narrowband short duration condition. Again, when the effects of duration and bandwidth are investigated separately the results suggest that it is the information in the spectral fringe that is most important. The presence of an effect of randomizing overall level under the narrowband short duration conditions is difficult to explain by simple models that are based solely on interaural differences. That is, even though the overall level is randomized, the interaural differences are not. Thus, in order to explain these data a model would, at a minimum, have to postulate the presence of multiplicative internal noise. On the other hand, the fact that this effect is eliminated when the bandwidth of the masker is wide suggests that the classical critical band interpretation of auditory masking is inadequate. These results were presented at the meeting of the AFOSR program in Auditory Pattern Recognition.

Binaural temporal masking

If the interaural phase of a noise masker is switched during the observation interval from in phase ($N0$) to 180° out of phase ($N\pi$) or from $N\pi$ to $N0$, a brief interaurally out-of-phase signal ($S\pi$) will be about 15 dB more detectable in the $N0$ portion of the noise than in the $N\pi$ portion. By investigating the change in detectability as a function of the delay (Δt) between the onset of the signal and the phase transition in the noise, the temporal response of the binaural system can be evaluated. The results of this case can be contrasted with a set of conditions in which the interaural phase of the noise is held constant ($N\pi$), but the level of the noise is reduced or increased by 15 dB halfway through the observation interval. Within a model such as the E-C model (N.I. Durlach, "Binaural signal detection: Equalization and cancellation theory" in J.V. Tobias (ed.), Foundations of Modern Auditory Theory II, 371-462, 1972) the first case produces a change of level only in the binaural channel. The second case produces a change in the noise level in the monaural channel. The curves that describe the relationship between threshold and Δt can be thought of as temporal masking functions. They show, like previous curves reported in the literature, that the decay of backward masking (cases where the $N0$ segment of the noise precedes an $N\pi$ segment or where the lower intensity segment of the noise precedes the higher intensity segment) is more rapid than for forward masking. In addition, the time course of the changes in detectability under both forward and backward masking is much more gradual when there is a phase change in the noise than when there is a level change, compatible with the notion that the binaural system responds "sluggishly" to changing interaural parameters (Grantham and Wightman, *J. Acoust. Soc. Am.* 63: 511-523, 1978). A manuscript describing this research is in preparation (Kollmeier and Gilkey, 1988).

Effects of forward masker fringe

- In studying the effects of a forward masker fringe, Yost (*J. Acoust. Soc. Am.* 78: 901-907, 1985) found that the threshold for

a brief S_n signal masked by a brief N_0 masking noise was not changed when an N_n forward masker fringe was added. This result was surprising in light of results such as those of McFadden (J. Acoust. Soc. Am. 40: 1414-1419, 1966) who showed that an N_0 forward fringe substantially improved performance in an N_0S_n detection task and concluded that the system uses the forward fringe as a reference against which to detect the signal. If an N_0 forward fringe provides a useful reference, it might be expected that an N_n forward fringe would provide a detrimental reference. Yost's results also seemed to conflict with the interpretations of Kollmeier and Gilkey (op. cit.), who thought of the N_n fringe as a forward masker. They found a gradual decrease in the amount of masking as a function of Δt (the delay between the onset of the signal and the phase transition in the noise), which showed substantial effect of the N_n segment of the noise even when the signal was presented well within the N_0 segment of the noise. One possibility was that the function that relates threshold to Δt for the N_n forward fringe condition intersects with the function that relates threshold to Δt for the pulsed masker condition at $\Delta t = 0$, even though the functions are different elsewhere. To resolve these questions, the detectability of an S_n tonal signal was investigated as a function of Δt , in the presence of an N_0 "masker" that was preceded by quiet, or by an N_n "forward fringe" and followed by quiet or by an N_0 or N_n "backward fringe." The results show that the two functions were indeed different and that they did not intersect when Δt was equal to zero. Overall, the results failed to replicate those of Yost, showing instead that the presence of the N_n forward fringe reduced detectability for all subjects under a wide variety of conditions. The results are a further indication that the auditory system uses information that does not overlap with the signal in the temporal domain. These data were presented to the MLD Society and to the Acoustical Society of America (Simpson and Gilkey, J. Acoust. Soc. Am. 82: S108(A), 1987). A manuscript is in preparation (Gilkey, Simpson, and Weisenberger, 1988).

Molecular psychophysical analyses of models of masking

In most studies of auditory masking, including those described above, both the stimulus and the performance of the subjects are described by their statistical properties (e.g., the average power of the stimulus and the average probability of a correct response). The outputs of models are described by their distributional properties and the average performance of a model is fit to the average performance of a subject. Another approach was described by Green (Psychol. Rev. 71: 392-407, 1964) and referred to as "molecular" psychophysics. In this approach, reproducible noise is used as a masker, such that the stimulus can be specified exactly on every trial. Similarly, the responses of the subject are considered on a trial-by-trial basis. The outputs of models are determined for each stimulus and the fit of the model is evaluated by comparing these outputs to the associated responses of the subjects.

Gilkey and Robinson (J. Acoust. Soc. Am. 79, 1499-1510, 1986) used computer models to predict subjects' responses to the individual noise samples of Gilkey, Robinson, and Hanna (J. Acoust. Soc. Am. 78: 1207-1219, 1985). The parameters of the models were manipulated until the outputs were best able to predict the subjects' responses. The combination of a 50-Hz-wide filter followed by a half-wave rectifier and an integrator with a 100-200-ms decay constant predicted their responses relatively well. However, a model that formed a weighted combination of the outputs of several detectors, each of which processed information in a different spectral region, yielded even better predictions and suggested that subjects compare the spectrum near the signal frequency to other areas of the spectrum. A similar model that processed information over different temporal intervals suggests that subjects also compare the waveform during the signal interval to the waveform immediately before the onset of the signal.

Gilkey and Robinson (op. cit.) investigated a fairly small set of reproducible noise samples (25 noise-alone and 100 signal-plus-noise samples). We have recently replicated and extended their finding using a larger set of reproducible noise samples (150 noise-alone and 150 signal-plus-noise samples). Again, we began with a simple model composed of a single-tuned filter centered at the signal frequency followed by a half-wave rectifier and an integrator with a leak. We sampled the output of the integrator at the end of the signal interval as the decision variable of the model. Our procedures have greater precision than those of Gilkey and Robinson and yield best-fitting bandwidths that are somewhat smaller than theirs (in the range of 26-49 Hz across subjects) and best-fitting decay constants that are somewhat shorter (in the range of 39-100 ms across subjects). We also investigated alternate or additional cues that the auditory system might be using to determine the presence of the signal. Specifically, we have considered cues related to the regularity of the envelope and the regularity of the fine structure of the waveform at the output of the model's initial filter. Preliminary results, however, suggest that these cues will not add greatly to the proportion of predicted variance. We also considered multi-channel models that combine the output of several detectors that process information in different spectral regions. The obtained spectral weighting functions, which describe how the model weights information across frequency, are quite similar to those of Gilkey and Robinson and suggest that subjects are comparing information in different spectral regions. A significant advancement has been the description of these weighting functions with a relatively simple equation, which can be interpreted as the difference between "excitatory" and "inhibitory" Gaussian-shaped weighting functions. We also examine models that combine the output of a single filter over several brief temporal windows. The obtained temporal weighting functions are quite similar to those of Gilkey and Robinson, and suggest that subjects compare information immediately before the signal onset to information during the signal interval. These results strongly question classical

critical band theory. Even though on average the subject's performance is unaffected by information that is outside a single critical band, when the responses to individual reproducible noise samples are investigated it can be seen that their responses are dependent on the pattern of spectral information across a frequency range that is much wider than a single critical band. That is, subjects are more likely to report the presence of the signal if the stimulus spectrum is "peaked" near the signal frequency, independent of the overall height of the spectrum. A talk describing some of these results was presented at the meeting of the Acoustical Society of America (Gilkey and Meyer, J. Acoust. Soc. Am. 82: S92(A), 1987) and at the meeting of the AFOSR program in Auditory Pattern Recognition. A manuscript is planned.

We have also been using the molecular psychophysical approach to examine models of binaural hearing. The method is comparable to that in the previous experiments except that the signal is presented 180° out of phase interaurally while the noise remains diotic. Initially, we have been examining a relatively small set of samples (25 noise-alone and 50 signal-plus-noise samples). Computing the output of a binaural model for a specific noise sample is not as straightforward as with a monaural model, because in most binaural models the internal noise must be added at an earlier stage and cannot be thought of as random variation at the input to the decision stage. We have implemented a computer model similar to the E-C model (Durlach, op. cit.). The model consists of two initial band-pass filters, one for each ear. The filtered waveform in the left channel is subject to random delay and multiplied by a random gain factor (i.e., "the internal noise"). The waveforms in each channel are subjected to fixed delays (equalization), and then the two channels are subtracted (cancellation). The differenced waveforms are then half-wave rectified and integrated to obtain the value of the model's decision variable. Because it is impractical to use Monte Carlo methods, the effect of the internal noise is incorporated by computing the decision variable for each of 100 equally probable combinations of random time delay and random gain factor from normal distributions of time delay and gain. One approach has been to investigate the relationship between the internal noise parameters of the model (i.e., the random time delay and gain factor) and the ratio of external to internal noise standard deviations (R) as described by Gilkey, Hanna, and Robinson (J. Acoust. Soc. Am. 69: S23(A), 1981) and others. Surprisingly, the value of R for the model is not a monotonic function of the magnitude of the internal noise parameters. Further, there are rather limited combinations of internal noise parameters that yield values of R comparable to those found with human observers.

A second approach has been to compute the average value of the model's decision variable for each waveform. This average value is used to predict the response of the subject on a waveform-by-waveform basis and vary the parameters of the model in order to produce the best fit. The parameters allowed to vary

are: the bandwidth of the initial filters, the standard deviation of a normal distribution of random delays, and the standard deviation of a normal distribution of random gains. Bandwidths fall in the range of 21 to 51 Hz, similar to the range observed in the NOS0 case. Standard deviations of delay distribution are between 119 and 200 μ s and correspond roughly to the values estimated by Colburn and Durlach ("Models of binaural interaction," in E.C. Carterette and M.P. Friedman (eds.), Handbook of Perception, 467-518, 1978). For two of three subjects, the values of the gain factor correspond relatively well to the estimates of Colburn and Durlach (op. cit.). For the third subject, however, the gain factor is near zero, a somewhat anomalous result. With these values we are able to predict between 53 and 60% of the variance in the data of the subjects, which is comparable to the single-channel model for the diotic case.

Next we formed a linear combination of the average output of the seven E-C elements, each tuned to a different frequency regions from 350 to 650 Hz. The bandwidth of the initial filter, the standard deviation of the delay distribution, and the standard deviation of the gain distribution in each channel were set to the values estimated for the single-channel model. We derived weighting functions described by the difference between excitatory and inhibitory Gaussian weighting functions. The resulting spectral weighting functions are quite similar in form to those for the diotic case and also produce a comparable increase in the proportion of predicted variance. A linear combination of the output of a single E-C element over seven different 21-ms subintervals of the signal interval also increases the proportion of predicted variance. However, the shapes of these temporal weighting functions are more inconsistent across subjects than was the case for the diotic condition.

Once again, these results indicate that classical critical band theory provides an oversimplified view of processing in auditory masking tasks. The weighting functions provide a quantitative description of the way the system compares information across frequency and across time. Even though diotic and dichotic masking have typically been assumed to be governed by quite different mechanisms, similar models can be used to predict the responses of the subjects in both cases, yielding similar results. These studies were presented at the meeting of the AFOSR program on auditory pattern recognition and a manuscript is planned.

Estimates of internal noise as a function of signal frequency

As mentioned above, the internal noise parameters of a binaural model are of critical importance. The relationship between R (the ratio of internal to external noise standard deviations) and signal frequency is being investigated under both NOS0 and NOS π conditions. The E-C model (Durlach, op. cit.) suggests that the influence of the random time delay should

increase with signal frequency. However, our initial estimates at 1500 Hz show values of R that are comparable to those observed at 500 Hz by Gilkey (op. cit.). Interpretation of these results is complicated, first by the fact that different subjects were used in the two studies and also by the fact that the relationship between the internal noise parameters of the E-C model and R is not simple. Thus we recently obtained additional measurements with a 500-Hz signal for the same subjects as our 1500-Hz measurements. The results will be analyzed to investigate further the relationship between these two concepts of internal noise.

Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model

We compared two common adaptive staircase rules, the "one up-two down" rule and the "one up-three down" rule (Levitt, J. Acoust. Soc. Am. 49: 467-477, 1971) in combination with a two-alternative forced-choice procedure and with a three-alternative forced-choice procedure. The adaptive staircase tracks were modeled as Markov chains. The model predicts that threshold estimates obtained with the adaptive techniques should be equal to those derived with equivalent "fixed signal level" techniques. However, the human data indicate that the adaptive techniques tend to yield lower thresholds. The model predicts that the standard error of a threshold estimate obtained from an adaptive technique will decrease and approach zero as the number of trials used to compute the estimate increases. The human data show greater variability than predicted and approach a nonzero value as the number of trials increases. The predictions of the model suggest that the commonly used combination of the 2AFC procedure and the "1 up 2 down" rule is the least efficient method of estimating a threshold and that the 3AFC procedure in combination with the "1 up 3 down" rule is the most efficient method. The human data are less consistent, but generally show the combination of the 2AFC procedure and the "1 up 2 down" rule to be one of the least efficient methods. A manuscript is in press (Kollmeier, Gilkey, and Sieben, J. Acoust. Soc. Am. 83: 1852-1862, 1988).

Laboratory development

As described in our original proposal, a major part of our effort during this first period of the grant (July 15, 1986 to March 15, 1988) has been spent upgrading and developing our computer facilities for laboratory control and data analysis. In October 1986 we installed a new multiuser MicroVAX II for data analysis, graphics, and signal processing to support our monaural and binaural modeling efforts. In January 1987 we replaced the aging Nova 4x computer system, which had been used for experiment control, but had become increasingly unreliable (four head crashes in December 1986). An existent SMS 1000 PDP11/73 computer was combined with newly developed hardware and software for stimulus presentation, response collection, timing, and device control. Programs were written to control specific

experiments on auditory and tactile perception, allowing rapid access to large sets of waveforms, and the ability to present four independent waveforms with 16-bit accuracy at sample rates up to 40 kHz.

The MicroVAX II has 13 megabytes of main memory, a 71-megabyte (formatted) Winchester disk and a second 300-megabyte (formatted) Winchester disk, a 95-megabyte streaming tape drive, and nine serial ports. The following devices are connected to the system: an LA210 draft printer, two Hewlett-Packard HP7475A pen plotters, a Courier 2400-baud modem, and terminals, including VT330, VT240 and HP2623A graphics terminals, and VT220 display terminals. One port is reserved for communication with the PDP11/73 computer that is used for experiment control. Principal applications of the MicroVAX II are program development, data analysis, graphics, and modeling. Fitting algorithms, signal processing subroutines, graphics software, and modeling programs have been implemented, allowing us to analyze our experimental results. Ethernet hardware and software installed on the MicroVAX II allow high-speed communication between the MicroVAX II and other computers of the Research Department, including the Speech Perception Laboratory MicroVAX II that allows access to ILS signal-processing software, and a Research Department MicroVAX II system that allows access to word processing and spreadsheet software from terminals in the Signal Detection Laboratory. For word processing we have purchased an LN03-plus laser printer. Also available on the network is an 8-port terminal server. In addition, a microwave link between CID and Washington University (and a planned high-speed light link) permits communication between the Signal Detection Laboratory and the Washington University network of computers on the main campus and medical school campus. Presently, there are over 100 computers on the network, providing access to a variety of applications. These include library search and catalog facilities, national and international mail services such as BITNET and ARPANET, and signal-processing and statistical packages.

The Scientific Micro Systems SMS-1000 Model 40 consists of a PDP11/73 processor, 4 megabytes of main memory, an 85-megabyte Winchester disk, 1.2 megabyte floppy disk drive, six serial ports, and a real-time clock. Peripherals include a VT220 system console terminal and an LA100 draft printer. Two Micro Technology Unlimited Digisound 16-bit digital-to-analog and analog-to-digital conversion subsystems provide a total of four channels of D/A and four channels of A/D, allowing the presentation of independent waveforms to a maximum of four listeners. A parallel I/O interface permits communication with subject response boxes and control of programmable attenuators and electronic switches. The SMS-1000 is used for real-time experiment control and data acquisition for the auditory and tactile experiments conducted in the Signal Detection Laboratory.

Direct memory access control of the Micro Technology Unlimited
DigiSound-16 with a Q-bus based computer

High-quality digital generation and recording of sounds is essential for many of our experiments. However, few high quality (16-bit) digital-to-analog and analog-to-digital subsystems were available for the PDP11/73 (Q-bus-based) computer, and most of these were expensive. After examining available options, we decided on the Micro Technology Unlimited (MTU) DigiSound-16. However, in order to use this system we had to overcome four problems. First, the two "stereo" channels within a single DigiSound are strobed with a 10 μ s interchannel delay, producing a detectable interaural difference. Second, while the DigiSound-16 has an 8-bit data path, the PDP11/73 has a 16-bit data path. Third, the DigiSound-16 requires the data for the two stereo channels to be interleaved in its own buffer memory, while it is typically most convenient to store the waveforms in separate arrays in computer memory. Fourth, no specific mechanism was available to allow direct memory access (DMA) control of the DigiSound 16. To overcome these problems we designed an interface that would allow as many as two DRV11-WA DMA interfaces to be connected to as many as two DigiSound-16s. This interface design is now in use on three computer systems here at CID and benefits a number of other projects, including AFOSR grant #86-0335 "Auditory-acoustic basis of consonant perception," J.D. Miller, PI. MTU has used this design to produce a commercially-available device, which we hope will be of benefit to other auditory scientists. A manuscript describing the design of this interface is in preparation (Gilkey and Partridge, 1988).

Software generation of reproducible noise

A software shift-register noise generator has been implemented to generate reproducible noise for our experiments. Given three input values, this program will generate an arbitrary length (up to 5.2 days) reproducible two-state binary noise, which, when filtered, is approximately white and Gaussian. A paper describing this program and some of the properties of the noise it produces has been published (Gilkey, Robinson, and Frank, J. Acoust. Soc. Am. 83: 829-831, 1988).

IV. PUBLICATION ACTIVITY

Publications

- Gilkey, R.H. (1987). "Spectral and temporal comparisons in auditory masking," in The Auditory Processing of Complex Sounds, edited by W.A. Yost and C.S. Watson (Erlbaum, New York), pp. 26-36.
- Gilkey, R.H., Robinson, D.E., and Frank, A.S. (1988). "A software pseudorandom noise generator," J. Acoust. Soc. Am. 83, 829-831.
- Kollmeier, B., Gilkey, R.H., and Sieben, U. (1988). "Adaptive staircase methods in psychophysics: An analysis using Markov theory," J. Acoust. Soc. Am. 83, 1852-1853.

Papers in preparation

- Kollmeier, B., and Gilkey, R.H. "Binaural temporal masking: Evidence for sluggishness in binaural detection," for submission to J. Acoust. Soc. Am.
- Gilkey, R.H., and Partridge, M.E. "Direct-memory-access control of the Micro Technology Unlimited Digisound-16 with a Q-bus-based computer," for submission to Behavior Res. Methods, Instrumentation and Computers.
- Gilkey, R.H., Simpson, B.D., and Weisenberger, J.M. "Effects of masker fringe on binaural detection," for submission to J. Acoust. Soc. Am.

Planned papers

- Gilkey, R.H., and Meyer, T.A. "Modeling subject responses in a reproducible noise masking task," planned for submission to J. Acoust. Soc. Am.
- Gilkey, R.H., Robinson, D.E., Hanna, T.E., and Weisenberger, J.M. "Methods of estimating internal noise," planned for submission to J. Acoust. Soc. Am.
- Gilkey, R.H. "Models of binaural masking: A molecular psychophysical approach," planned for submission to J. Acoust. Soc. Am.

V. PARTICIPATING PROFESSIONALS

Robert H. Gilkey

De Anza College, Cupertino, CA			
University of California, Berkeley, CA	B.A.	1976	Psychology
Indiana University, Bloomington, IN	Ph.D.	1981	Psychology

Dissertation title: "Molecular psychophysics and models of auditory signal detectability."

Wm. Michael Mudrovic

University of Missouri, St. Louis, MO	B.A.	1965	Spanish
University of Kansas, Lawrence, KS	Ph.D.	1976	Spanish
Washington University, St. Louis, MO	M.S.	1981	Speech and Hearing

V. INTERACTIONS

Invited papers and conference talks

- Gilkey, R.H. (1986). "Estimates of internal noise in binaural masking," MLD Society, Chicago, IL.
- Gilkey, R.H. (1986). "Binaural processing," Central Institute for the Deaf, St. Louis, MO.
- Gilkey, R.H. (1986). "Spectral and temporal comparisons in auditory masking," Central Institute for the Deaf, St. Louis, MO.
- Gilkey, R.H. (1987). "Temporal interactions in binaural masking: Effects of forward masker fringe," MLD Society, Madison, WI.
- Gilkey, R.H., and Meyer, T.A. (1987). "Modeling subject responses in a reproducible noise masking task," J. Acoust. Soc. Am. 82, S92-93(A).
- Simpson, B.D., and Gilkey, R.H. (1987). "Effects of forward masker fringe on binaural detection," J. Acoust. Soc. Am. 82, S108(A).
- Gilkey, R.H. (1987). "Binaural masking: An analysis of models," Meeting of the AFOSR Program in Auditory Pattern Recognition, Evanston, IL.
- Gilkey, R.H. (1988). "Relations between monaural and binaural masking," Psychoacoustics Laboratory, Department of Psychology, University of Florida, Gainesville, FL.

Binaural Masking: An Analysis of Models (#86-0298)
Robert H. Gilkey, PI

APPENDIX

Manuscripts included:

- 1) Gilkey, R.H. (1987). "Spectral and temporal comparisons in auditory masking," in The Auditory Processing of Complex Sounds, edited by W.A. Yost and C.S. Watson (Erlbaum, Hillsdale, NJ), pp. 26-36.
- 2) Gilkey, R.H., Robinson, D.E., and Frank, A.S. (1988). "A software pseudorandom noise generator," J. Acoust. Soc. Am. 83, 829-831.
- 3) Kollmeier, B., Gilkey, R.H., and Sieben, U.K. (1988). "Adaptive staircase methods in psychophysics: An analysis using Markov theory," J. Acoust. Soc. Am. 83, 1852-1862 (in press).
- 4) Gilkey, R.H., Simpson, B.D., and Weisenberger, J.M. "Effects of masker fringe on binaural detection," in preparation.
- 5) Gilkey, R.H., and Partridge, M.E. "Direct-memory-access control of the Micro Technology Unlimited Digisound-16 with a Q-bus-based computer," in preparation.
- 6) Kollmeier, B., and Gilkey, R.H. "Binaural temporal masking: Evidence for sluggishness in binaural detection," in preparation.

Spectral and Temporal Comparisons in Auditory Masking

Robert H. Gilkey
Signal Detection Laboratory
Central Institute for the Deaf
St. Louis, MO 63118

The critical band as it relates to simultaneous tone-in-noise masking is reevaluated in light of the results from two experiments. The first measures the detectability of a 500 Hz tone masked by white Gaussian noise when the overall level of the stimulus is randomized from interval to interval and trial to trial. The results indicate that subjects can use information that is outside the critical band centered at the signal frequency, or outside the temporal interval that contains the signal, to overcome the performance decrement caused by randomizing level. The second experiment investigates the detectability of a 500 Hz tone masked by reproducible noise. A multiple channel model, fit to the data, suggests that subjects compare information in different spectral/temporal regions of the stimulus in order to determine the presence of the signal.

INTRODUCTION

The concept of the critical band evolved from the work of Fletcher (1940), who investigated the detectability of pure tone signals as a function of the bandwidth of a simultaneous white Gaussian noise masker. He found that detectability decreased as the bandwidth of the masker increased, until a "critical bandwidth" was reached. Further increases in bandwidth did not affect detectability, even though the total masker energy increased. The results can be explained by assuming that the system has fixed bandwidth (approximately one critical band wide) internal filters, centered at each signal frequency, that limit the bandwidth of the effective masker. Thus, increases in the bandwidth of the external masker beyond the critical bandwidth do not change the energy in the effective masker. These basic results have been replicated in numerous studies since Fletcher's initial work (e.g., Weber, 1977), and the concept of the critical band has become a cornerstone of psychophysical thought on hearing. Despite an awareness

Gilkey - Spectral and Temporal Comparisons

that in isolation the critical band would be inadequate to explain such complex phenomena as speech and music perception, it has been assumed that the critical band could provide a front end for these complicated processes. Moreover, for simple situations, like masking, critical band theory is typically considered to be adequate in and of itself. That is, the system has been modeled as a single filter followed by a simple detector (e.g., an energy detector). Similar constructs have also been applied in the temporal domain. Although the results are less consistent, in general the subject in a masking task has been seen as processing information over a single fixed integration window, that roughly corresponds to the interval that contains the signal when it is present (e.g., Robinson and Trahiotis, 1972).

A number of more recent results suggest that even for masking tasks, critical band theory may be an oversimplification. Experiments measuring frequency selectivity in nonsimultaneous masking tasks suggest interactions among frequency regions that cannot be explained by critical band theory (e.g., Houtgast, 1972). Studies of simultaneous masking where the masker is an inharmonic tone complex whose level is randomized from trial to trial suggest that performance is augmented well outside of a single critical band (e.g., Green, 1983). Similarly, detection in certain amplitude-modulated noises suggests that performance improves as the bandwidth of the noise is increased beyond a critical bandwidth (e.g., Hall et al., 1984). Despite these results, it has still been thought that for simple stimulus configurations like those used by Fletcher, with simultaneous masking of a pure tone by an unmodulated band-limited white Gaussian noise, the critical band is an adequate construct to explain the results. That is, subjects ignore the portions of the noise that are outside the critical band and temporal interval that contains the signal. In contrast, the current chapter describes the results from two experiments that suggest that even for very simple simultaneous noise masking tasks, critical band theory may be inadequate.

EXPERIMENT 1

Experiment 1, like the Profile Analysis experiments of Green (1983), randomizes the overall level of the stimulus from trial to trial, in an attempt to force the

Gilkey - Spectral and Temporal Comparisons

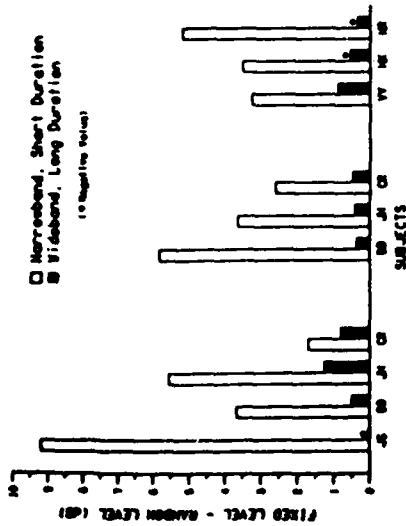


Figure 1. Effect of randomizing overall level for the NBSD masker and the WBLD masker. Left two groups of bars are for monaural presentation and the right group is for binaural presentation.

subjects away from a strategy based on energy detection in a single critical band, and into a discrimination based on spectral shape. Here, the background (i.e., the masker) is a noise, rather than a tone complex. A two-interval, forced choice adaptive staircase technique was used to obtain the threshold of a 50 msec, 500 Hz sinusoid, masked by white Gaussian noise maskers. For reasons not addressed in this chapter, both monaural and binaural (diotic) stimulus configurations were employed. With each stimulus configuration both wide-band long-duration (WBLD) maskers and narrow-band short-duration (NBSD) maskers were investigated. The bandwidth of the WBLD maskers was 100 to 2000 Hz; the duration was 300 ms for binaural presentations and 654 ms for monaural presentations. The bandwidth of the NBSD maskers was 50 Hz, centered at 500 Hz; the duration was 56 ms for both monaural and binaural presentations. Detection was investigated under a fixed level condition in which the overall level of the stimulus was fixed, and a random level condition in which overall level was randomized, from interval to interval and trial to trial, over a 30 dB range (binaural presentations) or a 40 dB range (monaural presentations).

Figure 1 shows the difference in threshold attenuation between the fixed and the random level

Gilkey - Spectral and Temporal Comparisons

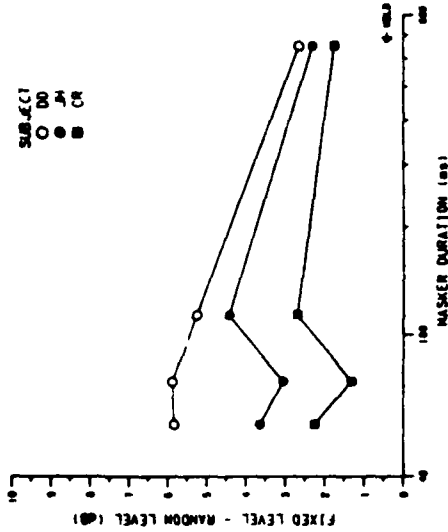


Figure 2. Effect of randomizing level as a function of masker duration. Monaural presentation.

conditions for several subjects. The left two groups of bars show the data for monaural presentations and the right group of bars shows the data for binaural presentations. The asterisks mark data where performance under the random level condition was actually better than that under the fixed level condition. For the rest of the bars, the typical case, performance under the fixed level condition was superior. With the WBLD masker (filled bars) the effect of randomizing level is negligible; the average difference is only about .4 dB. If subjects were monitoring the energy in a single critical band and a single integration window under the random level condition, their performance should be much worse than under the fixed level condition. The small differences suggest that the subjects are employing some other strategy, perhaps using information in the spectral and temporal fringe, to avoid the effects of randomizing level. If so, then it should be possible to remove the spectral and temporal fringe, and observe a decrement in performance. Results for the NBSD masker are shown by the open bars. As can be seen, there is a much greater effect of randomizing level when the spectral fringe and temporal fringe are reduced. The average of these open bars is about 4.4 dB, suggesting that subjects are attending to the spectral and temporal fringe of the WBLD masker.

Gilkey - Spectral and Temporal Comparisons

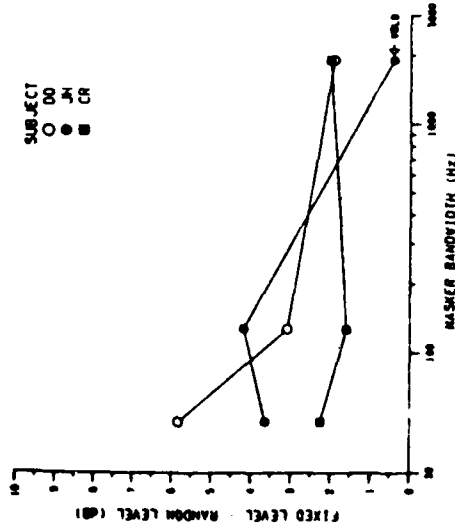


Figure 3. Effect of randomizing level as a function of masker bandwidth. Monaural presentation.

Additional conditions varied masker bandwidth and masker duration separately to determine the individual effects of the spectral fringe and the temporal fringe. Figure 2 shows the difference in threshold attenuation between the fixed level and the random level condition, as a function of the duration of the masker. In all cases the masker is narrow-band, approximately 50 Hz wide. As can be seen, on average there is a reduction in the effect of randomizing level as duration is increased, although for two subjects there is not a great difference between the longest and shortest maskers. For no subject is the information in the temporal fringe of the long duration masker sufficient to reduce the performance difference to that observed with the WBLD masker of Figure 1 (shown here as the arrow in the lower left-hand corner). Figure 3 shows the effect of manipulating the bandwidth of a 56 msec duration masker. Again, there is a reduction in the effect of randomizing level as the bandwidth is increased. For one subject the performance difference is the same as with the WBLD masker, indicating that the increase in bandwidth was sufficient to overcome the effects of randomizing level. For the other two subjects, the effect of randomizing level is greater than with the WBLD masker. Thus, most subjects can use information in either the temporal fringe or the spectral fringe to overcome the effects of

Gilkey - Spectral and Temporal Comparisons

randomizing overall level, and there is an added benefit of having both spectral and temporal fringe present.

EXPERIMENT 2 - (from Gilkey et al., 1985; Gilkey and Robinson, 1986)

Experiment 2 provides some indication of the mechanisms that might be underlying the effects observed in Experiment 1. This experiment differs from traditional masking experiments in that reproducible noise samples are used as maskers. The procedures are analogous to those of Ahumada et al. (1975), and others. The detectability of a 100 msec, 500 Hz sinusoid masked by each of 25 reproducible noise samples was investigated. The noise samples were 148 msec long, and had a bandwidth of 100-3000 Hz. A total of 125 different waveforms were employed - 25 noise-alone and 100 signal-plus-noise (25 at each of four signal starting phase angles). The proportion of "yes" responses to each waveform was estimated, based on approximately 100 single-interval yes/no trials per sample. All of the 25 noise samples were presented during each block of 100 trials, but the order of presentation was randomized across blocks. Subjects did not know that the samples were being repeated, and they did not receive trial-by-trial feedback. Thus, they believed they were listening to truly random noise, and did not learn individual samples.

Figure 4 shows the results for each of the 25 noise samples for one subject and a single signal starting phase in ROC space (i.e., $P(y)$ on signal-plus-noise trials as a function of $P(y)$ on noise-alone trials). As can be seen, all samples are not the same, but are distributed fairly broadly throughout this space. These differences across samples can be used to test detection models by presenting these same noise samples to a computer model. The parameters of the model can be adjusted until the model has the same $P(y)$ to the individual samples as do the subjects. Initial results were obtained for the simple model illustrated in Figure 5, which is based on Jeffress (1968). It is composed of a critical band-like filter, followed by a nonlinearity and an integrator with a leak. A sampling strategy is applied to the slowly-varying waveform at the output of the integrator to reduce it to a single number, the model's decision variable, x' . Although it is beyond

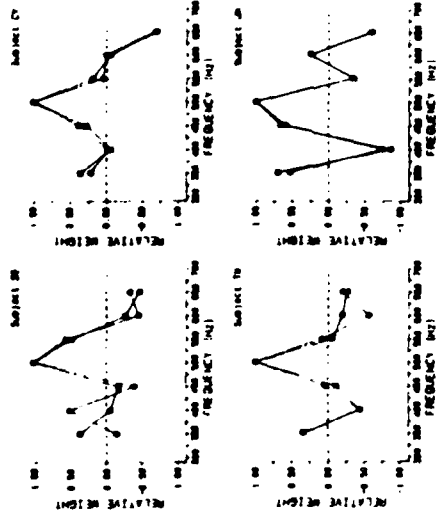


Figure 6. Spectral weighting functions for a multiple-detector model. Square and circular symbols are for $10 \log(z/Ng)$ of 8.5 and 11.5, respectively. (From Gilkey and Robinson, 1986, by permission of the Journal of the Acoustical Society of America.)

filter tuned to different frequency regions, at 50 Hz intervals between 350 and 650 Hz. A new decision variable was defined as a linear combination of the seven values of X' . The coefficients, or weights, applied to each X' in the linear combinations were manipulated until the combined decision variable again led to $P(y)$ s for the model that agreed most closely with the subjects. The curves in Figure 6 show the way the spectral information was weighted by the model that best fit the data of each of four subjects. If a single channel model was adequate, then we would expect to find positive weights for the channel centered at the signal frequency, and weights of zero for all of the other channels (i.e., equivalent to the single channel model of Figure 5). If, on the other hand, the subject looks broadly through this frequency region and responds to an increase in the output of any of the detectors, positive weights would be expected for all of the detectors. Neither of these patterns was found for any of the subjects. Instead, a maximum positive weight is assigned to the signal frequency, but negative weights are assigned to certain other frequencies. The results for all subjects show negative weights at the highest frequencies, and some show negative weights for a band in

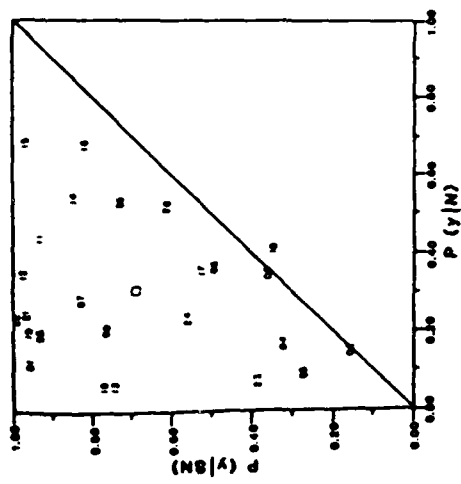


Figure 4. Hit and false-alarm proportions for individual noise samples shown in ROC space. Data are for subject SG, with a signal starting phase of 90. (From Gilkey et al., 1985, by permission of the Journal of the Acoustical Society of America.)

the scope of this chapter to provide detailed results of the fitting process, it is the case that a 50 Hz wide filter, followed by a half-wave rectifier, combined with integration and sampling strategy stages that approximate true integration over the signal interval, predicts the responses of all subjects fairly well. When the d' of the subjects is about 1.8, the model predicts between 43% and 72% of the variance in the $P(y)$ s, depending on the subject.

Because the proportion of predictable variance is large relative to these values, we next combined seven detectors of this form, each with a 50 Hz wide initial

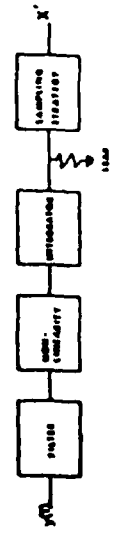


Figure 5. A block diagram of the electrical analog model of Jeffress, 1968. (From Gilkey and Robinson, 1986, by permission of the Journal of the Acoustical Society of America.)

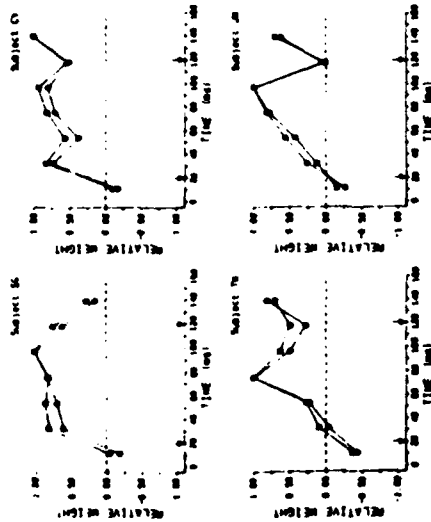


Figure 7. Temporal weighting functions for a multiple-detector model. Square and circular symbols are for $10 \log(E/N_0)$ of 8.5 and 11.5, respectively. (From Gilkey and Robinson, 1986, by permission of the Journal of the Acoustical Society of America.)

the low frequency region. It is as if subjects are comparing information across frequency regions. That is, as energy increases in the regions of negative weight, they are less likely to say yes, unless the energy also increases in the regions of positive weight. A similar analysis was performed in the temporal domain. A linear combination was formed of the average output over seven 21-ms sub-intervals of the stimulus interval and the best-fitting weights were determined, as shown in Figure 7. Here, the noise begins at 9 and ends at 148 msec. The small arrows show the onset and offset of the signal. For all four subjects there is a slight negative weight given to the interval immediately before the signal onset. Other regions of the functions show generally positive weights. The small negative weight at the beginning is at least suggestive of some sort of comparison process. But as seen in Experiment 1, the spectral weighting seems to be more important. The average increase in the proportion of predicted variance with the frequency weights is about 8%, and the average increase with the temporal weights is about 6%.

CONCLUSION

The results of both of these experiments suggest that subjects can and do use information in the spectral and temporal fringe of the masker waveform in order to make decisions regarding the presence of a signal. The results of both experiments conflict with simple single filter models of auditory detection. It could be argued that the first experiment, as well as the experiments of Green (1983), does not represent a typical masking situation because the act of randomizing overall level degrades the quality of the information in the critical band centered on the signal frequency. Therefore, subjects are forced to use other information in order to achieve the same level of performance. That is, they do not use the strategy they would in a typical masking task, such as the task investigated by Fletcher.

Similar arguments apply to the experiment of Hall et al. (1984). Although the information in the critical band centered on the signal frequency was not degraded, additional information was provided in other critical bands. There is no particular reason to expect the subjects to ignore this additional information. In contrast, the second experiment, from the subject's point of view, was a typical masking situation. There should be no particular advantage for the subjects in attending to information outside the critical band or outside the temporal integration window that contains the signal. Nevertheless, the results of fitting the model suggest that subjects are using this additional information. The results of both of these experiments suggest that even in very simple simultaneous tone-in-noise masking tasks, subjects use listening strategies that are more complicated than would be suggested by a traditional interpretation of the critical band.

ACKNOWLEDGMENTS

Work was supported by NSF (BNS-77-17398, BNS-85-11786), and NIH (NS-93856). Additional support was provided by APOSR (86-NL-949). The author is indebted to D.E. Robinson, T.E. Hanna, S.F. Cooper, B.D. Simpson, T.A. Meyer, and J.M. Weisenberger.

Gilkey - Spectral and Temporal Comparisons

REFERENCES

- Ahumada, A., Marken, R., and Sandusky, A. (1975). "Time and frequency analyses of auditory signal detection," *J. Acoust. Soc. Am.* **57**, 385-390.
- Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.* **12**, 47-65. Reprinted in *Psychological Acoustics*, edited by E. D. Schubert (Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1979), pp. 221-230.
- Gilkey, R.H., Robinson, D.E., and Hanna, T.E. (1985). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," *J. Acoust. Soc. Am.* **78**, 1207-1219.
- Gilkey, R.H. and Robinson, D.E. (1986). "Models of auditory masking: A molecular psychophysical approach," *J. Acoust. Soc. Am.* **79**, 1499-1510.
- Green, D.M. (1983). "Profile analysis, a different view of auditory intensity discrimination," *Am. Psychol.* **38**, 133-142.
- Hall, J.W., Haggard, M.P., and Fernandes, M.A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50-56.
- Houtgast, T. (1972). "Psychophysical evidence for lateral inhibition in hearing," *J. Acoust. Soc. Am.* **51**, 1885-1894.
- Jeffress, L.A. (1968). "Mathematical and electrical models of auditory detection," *J. Acoust. Soc. Am.* **44**, 187-203.
- Robinson, D.E. and Trahiotis, C. (1972). "Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions," *Percept. Psychophys.* **12**, 333-334.
- Weber, D.L. (1978). "Suppression and critical bands in band-limiting experiments," *J. Acoust. Soc. Am.* **64**, 141-150.

LMED
88