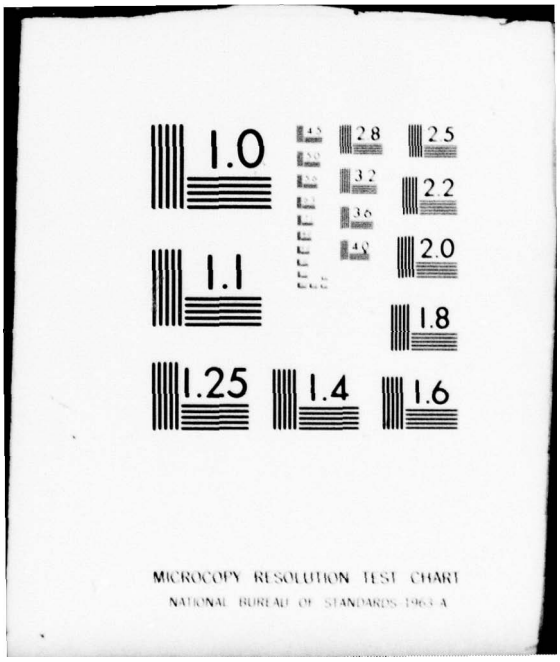
AD-A072 941 ORINCON CORP LA JOLLA CALIF F/G 17/1
FURTHER DEVELOPMENT AND NEW CONCEPTS FOR BIONIC SONAR. VOLUME 2--ETC(U)
OCT 78 R A ALTES, R W FLOYD N66001-78-C-0080
UNCLASSIFIED OC-R-78-A004-1-VOL-2 NOSC-TR-404-VOL-2 NL

1 OF 2
AD A072941




1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL III

12 B.S.

NOSC

NOSC TR 404
VOLUME 2

DDC
RECEIVED
AUG 21 1979
C

NOSC TR 404
VOLUME 2

Technical Report 404

FURTHER DEVELOPMENT AND NEW CONCEPTS FOR BIONIC SONAR

Volume 2 - Spectrogram Correlation

RA Altes
ORINCON Corporation
RW Floyd, NOSC
(Contract Monitor)

October 1978

AD A072941

DDC FILE COPY

Approved for public release; distribution unlimited

NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO, CALIFORNIA 92152

79 08 20 026

NAVAL OCEAN SYSTEMS CENTER, SAN DIEGO, CA 92152

AN ACTIVITY OF THE NAVAL MATERIAL COMMAND

RR GAVAZZI, CAPT, USN
Commander

HL BLOOD
Technical Director

ADMINISTRATIVE INFORMATION

The work reported here was performed by Richard A. Altes, ORINCON Corporation, 3366 No. Torrey Pines Ct., Suite 320, La Jolla, CA 92037 under Contract Number N-66001-78-C-0080. R. W. Floyd, Code 512, was the contract monitor.

The work was reported in three volumes. This document contains volume 2 of the report, reproduced in its entirety.

Released by

Dr. JF FISH, Head
Bioacoustics and Bionics Division

Under authority of

HO PORTER, Head
Biosciences Department

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) NOSC TR-404 Volume 2 (9) Final Rept.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) FURTHER DEVELOPMENT AND NEW CONCEPTS FOR BIONIC SONAR. VOLUME 2, SPECTROGRAM CORRELATION.	5. TYPE OF REPORT & PERIOD COVERED FINAL FY 78
	6. PERFORMING ORG. REPORT NUMBER (14) OC-R-A004-1-Vol-2
7. AUTHOR(s) -78- RA Altes, ORINCON Corporation RW Floyd, NOSC (Contract Monitor)	8. CONTRACT OR GRANT NUMBER(s) (15) N66001-78-C-0080
9. PERFORMING ORGANIZATION NAME AND ADDRESS ORINCON CORPORATION 3366 No. Torrey Pines Ct., Suite 320 La Jolla, CA 92037	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE6271; SF11-121-491
11. CONTROLLING OFFICE NAME AND ADDRESS (11) Naval Ocean Systems Center, Hawaii Laboratory PO Box 997 Kailua, Hawaii 96734	12. REPORT DATE October 1978
	13. NUMBER OF PAGES 144
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 150 p. NAVSEA 0342 (F. Romano) 2221 Jefferson Davis Highway Arlington, VA 22314	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(10) Richard A. Altes
R. W. Floyd

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(16) F11121

18. SUPPLEMENTARY NOTES

(17) SF11121491

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bionics
Optimum detectors
Signal processing
Sonar
Spectrogram correlation
Target transfer function

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

This report documents theoretical studies of models of echolocation signal processing of dolphins and the testing of the effectiveness of these models for detecting and classifying small objects in introduced noise and reverberation. In this volume it is demonstrated that spectrogram correlation is a locally optimum detection operation for low SNR signals that have been passed through a random, time-varying process in Gaussian noise. For non-Gaussian

→ next page (Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

392776

20. (Continued)

noise local optimality is preserved by correlating stored data with a modified spectrogram. Detection and estimation performance of the spectrogram process is predicted and compared with psychophysical data.

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

TABLE OF CONTENTS

Volume 2. SPECTROGRAM CORRELATION

		Page
2.1	Introduction	1
2.2	Signal Reconstruction and Detection from Spectrograms, with Applications to Theories of Hearing and Animal Echolocation.	2
2.2.1	Introduction to Section 2.2	2
2.2.2	Definitions and Properties of Time-Frequency Energy Distributions, Spectrograms, and Ambiguity Functions	6
2.2.4	Sampling the Spectrogram	11
2.2.5	Reconstruction of a Noise-Free Signal from a Noise-Free Spectrogram	15
2.2.6	Estimation of a Signal Waveform from Its Spectrogram in the Presence of Noise	18
2.2.7	Detection of Weak Signals in Noise.	23
2.2.8	Array Processing for Direction Estimation	34
2.2.9	Application to Theories of Hearing and Animal Echolocation	38
2.2.10	Summary	51
2.2.11	Conclusion	55
2.2.12	References for Section 2.2	57
2.2.13	Appendices for Section 2.2	63
2.3	Parameter Estimation from Spectrograms	99
2.3.1	Brief Summary of Section 2.3	99
2.3.2	Introduction.	99
2.3.3	A Locally Optimum Maximum Likelihood Estimator.	100
2.3.4	A Variance Bound for the Locally Optimum ML Estimate	101
2.3.5	Delay (Range) Estimation	103

TABLE OF CONTENTS (Continued)

		Page
2.3.6	Frequency (Doppler Shift) Estimation	107
2.3.7	Recommended Experiments and Comparison of Results with Existing Data.	111
2.3.8	Conclusion	113
2.3.9	References for Section 2.3	116
2.4	Spectrogram Correlation as an Ideal Detection Process for Signals that Have Been Passed Through a Time-Varying, Random Channel	118
2.4.1	Brief Summary of Section 2.4	118
2.4.2	Introduction.	119
2.4.3	Scattering Functions and Spectrograms	122
2.4.4	Detection Via Spectrogram Correlation	124
2.4.5	K-L Representations Versus Spectrograms . . .	126
2.4.6	Signal Design for Maximization of Detector Signal-to-Noise Ratio and Signal-to-Interference Ratio	135
2.4.7	Detection of a Random, Time-Varying, Range-Extended Target with a Bank of Matched Filters	138
2.4.8	Conclusion	139
2.4.9	References for Section 2.4	141
2.4.10	Appendix to Section 2.4.	142
2.5	Summary and Conclusions for Volume 2.	144

II. SPECTROGRAM CORRELATION

2.1 Introduction

The experimental results in Volume 1 indicate that a spectrogram correlation process (using principal component analysis) is an excellent target classification device. The use of a spectrogram representation is natural to a bionic sonar receiver, since the spectrogram is an idealized model of the neural transduction process in the mammalian cochlea. A spectrogram representation, however, does not necessarily imply the use of a spectrogram correlator as a signal detection or parameter estimation device.

Section 2.2 considers bionic aspects of spectrogram processing in some detail. It is shown in Section 2.2 that spectrogram correlation is in fact a locally optimum detection device when the data reaches the detector in the form of a spectrogram. The utilization of a spectrogram correlator by human listeners implies that detection of a signal with multiple harmonics depends upon the sum of the squared energies in the harmonics. This phenomenon has been observed by D. M. Green.

Section 2.3 explores the capabilities of a spectrogram correlator as a parameter estimation device. The predicted standard deviation of a spectrogram frequency measurement again corresponds to psychoacoustic data. The standard deviation of a delay measurement is also predicted, and experiments to check this prediction are suggested.

Section 2.4 investigates the spectrogram correlator as a detector for signals that have been passed through a time-varying, random channel. It is shown that the expected spectrogram of the channel output is the convolution of the channel scattering function

with the spectrogram of the input signal. In the case of low energy coherence (small SNR at any one filter output at any one time instant), the convolution result implies that a spectrogram correlator is equivalent to the usual Karhounen-Loève receiver implementation. Furthermore, the spectrogram correlator can be modified to accommodate non-Gaussian noise backgrounds by changing the power law that is used to envelope detect the responses of a bank of bandpass filters. The spectrogram correlator is therefore an ideal detector for low SNR signals that have been passed through a random channel. An example of such a channel is a range extended sonar target or an undersea propagation channel that involves refraction or reflection from surface or bottom.

2.2 Signal Reconstruction and Detection from Spectrograms, with Applications to Theories of Hearing and Animal Echolocation

2.2.1 Introduction to Section 2.2

2.2.1.A Brief Summary of Section 2.2

Properties of spectrograms and related signal representations are listed, and some new properties are derived. These properties indicate that a signal can be reconstructed from its spectrogram except for the time invariant part of the signal's phase. A detection method that uses the spectrogram itself as data is compared to a detector that operates on the reconstructed signal. The reconstruction method can also be applied to hearing aid design and to signal synthesis for auditory experimentation.

The theoretical results are used to interpret neurophysiological properties of the peripheral mammalian auditory system,

and to predict the outcome of various psychoacoustic experiments. Matched filtering of a signal that is reconstructed from a spectrogram may be relevant to animal echolocation. Detection experiments with humans, however, seem to favor the correlation of a data spectrogram with a reference spectrogram that is either prespecified (spectrogram correlator) or estimated from the data itself (spectrogram estimator-correlator).

2.2.1.B Background and Goals

A spectrogram provides a convenient signal representation by portraying a waveform as a non-negative function of instantaneous frequency and time. The convenience of the representation is probably related to the fact that signals are converted into lines on a plane, or into a three-dimensional surface that shows intensity as a function of time and frequency. Since these visual representations are similar to other images in our environment, they can be easily assimilated by a human observer.

There is evidence that spectrogram-like signal representations may be formed by the peripheral auditory system (Evans, 1975 and 1977; Siebert, 1968). To be sure, some of the elements in the biological system are nonlinear-with-memory (Kim, Molnar, and Pfeiffer, 1973), probabilistic (de Boer, 1975; de Boer and de Jongh, 1978; Siebert, 1970) or both, but we should nevertheless be able to further our understanding of audition by analysis of an idealized model with linear filters and deterministic envelope detectors.

Our primary motive in studying such a model is to try to answer questions such as the following:

1. What information about a signal is lost when a spectrogram is formed?

2. To what extent can we reconstruct a signal from its spectrogram?

3. Can a data spectrogram and a reference spectrogram be compared in order to form a nonideal detector? What is the best operation for implementing such a comparison? How much worse is this "best" nonideal detector than an ideal detector?

4. Is it possible to construct an ideal detector, i.e., a matched filter, for data that is given in the form of a spectrogram? Are bats and dolphins capable of synthesizing ideal detectors for echolocation?

5. Man-made spectrogram synthesizers do not usually incorporate a great amount of filter overlap, but the auditory spectrogram (if it exists) is formed from many filters with transfer functions that overlap very closely. What is the reason for this difference? Why is there such a high density of hair cells along the basilar membrane (Fletcher, 1953) when cochlear resonances are so broad (Pfeiffer and Kim, 1975)?

6. Can fine frequency resolution be obtained with envelope detected responses of overlapping critical bandwidth (approximately 1/3 octave wide) filters?

Is frequency discrimination capability dependent upon the sharp, high-frequency cutoffs that are sometimes observed in neural tuning curves?

7. How can a better understanding of spectrograms be applied to hearing aid design or to future experimental work in audition?

2.2.1.C Outline of Section 2.2

Section 2.2 begins with some definitions and properties of spectrograms and related signal representations. One of the most important properties is that an input signal can be reconstructed from a noise-free spectrogram, except for a multiplicative constant with unit magnitude and unknown phase. Another property indicates that only the signal autoambiguity function is needed for matched filtering, and reconstruction of the waveform itself is thus unnecessary for detection. The effect of noise upon the signal reconstruction process is investigated, and sampling considerations are discussed. A "best" nonideal detector is derived for input data that is given in the form of a spectrogram. The performance of this detector is compared with that of a "best" detector that uses the original input data rather than the spectrogram.

Section 2.2 goes on to consider some ramifications of the spectrogram as a model for the peripheral auditory system. The existence of overlapping filters, monaural phase insensitivity, and critical bands can all be explained in terms of spectrogram properties. Rapid high-frequency cutoff in neural tuning curves is evaluated from the viewpoint of spectrogram analysis. Binaural direction estimation is considered.

An important ramification of a spectrogram representation is that matched filtering can be performed upon reconstructed input data, and an ideal detector can theoretically be synthesized in an animal echolocation system. A nonideal process using spectrogram correlation can also be used, but such a process requires 5 dB more signal energy to achieve the same performance as an ideal detector.

The possibility of data reconstruction and subsequent matched filtering means that detection experiments can be realistically analyzed with a matched filter hypothesis rather than with an energy detector hypothesis. For a two-alternative, forced choice experiment, the theoretical function that predicts probability of a correct response versus signal energy does not have as steep a slope as a best fit to the data, when the theoretical curve is based upon an energy detector model. A theoretical curve with steeper slope can be obtained from a matched filter model. Masking experiments with short duration sinusoids also yield results that partially support a matched filter hypothesis.

If a spectrogram is an acceptable model for the response of the peripheral auditory system, then the fact that a signal can be almost completely reconstructed from a noise-free spectrogram suggests a new method of hearing aid design and a new signal synthesis method for future experiments in audition.

2.2.2 Definitions and Properties of Time-Frequency Energy Distributions, Spectrograms, and Ambiguity Functions

A given real signal $u_r(t)$ can be used to construct a complex waveform (Gabor, 1946; Rihaczek, 1969; Franks, 1969)

$$u(t) = 2^{-\frac{1}{2}} \left[u_r(t) + j \hat{u}_r(t) \right] \tag{1}$$

where $\hat{u}_r(t)$ is the Hilbert transform (Erdelyi, 1954) of $u_r(t)$. The waveform $u(t)$ in (1) is the so-called analytic representation of $u_r(t)$, and $\mathrm{Re}\{u(t)\} = 2^{-\frac{1}{2}} u_r(t)$. The Fourier transform of $u(t)$ is

$$U(f) = \int_{-\infty}^{\infty} u(t) \exp(-j2\pi ft)\, dt . \tag{2}$$

If $u(t)$ is analytic, $U(f)$ is identically zero for $f < 0$. The factor $2^{-\frac{1}{2}}$ in (1) is included to preserve signal energy.

The envelope of $u_r(t)$ is usually approximated by passing a rectified version of $u_r(t)$ through a low-pass filter or integrator. An exact version of the envelope, however, is

$$|u(t)| = 2^{-\frac{1}{2}} \left[u_r^2(t) + \hat{u}_r^2(t) \right]^{\frac{1}{2}} \tag{3}$$

which is the magnitude of the analytic representation of $u_r(t)$.

The time-frequency energy density function of $u(t)$ is defined (Rihaczek, 1968) as

$$e_{uu}(t, f) = u(t)\, U^*(f) \exp(-j2\pi ft) . \tag{4}$$

The narrowband ambiguity function of $u(t)$ is (Woodward, 1964)

$$X_{uu}(\tau, \phi) = \int_{-\infty}^{\infty} u(t)\, u^*(t+\tau) \exp(-j2\pi\phi t)\, dt . \tag{5}$$

Similarly, we can define the cross-energy density function and cross-ambiguity function by

$$e_{uv}(t, f) = u(t)\, V^*(f) \exp(-j2\pi ft) \tag{6}$$

$$X_{uv}(\tau, \phi) = \int_{-\infty}^{\infty} u(t)\, v^*(t+\tau) \exp(-j2\pi\phi t)\, dt\,. \tag{7}$$

A spectrogram can be obtained either by bandpass filtering or by Fourier transformation of a time-windowed signal. In the case of bandpass filtering, the short-time spectral history is (Ackroyd, 1971)

$$S_{uv}(t_1, f_1) = \left| \int_{-\infty}^{\infty} U(f)\, V(f - f_1) \exp(j2\pi ft_1)\, df \right|^2 \tag{8}$$

$$= \left| \int_{-\infty}^{\infty} u(t_1 - t)\, v(t) \exp(j2\pi f_1 t)\, dt \right|^2 \tag{9}$$

where u(t) and v(t) are analytic. $S_{uv}(t_1, f_1)$ is the squared envelope of the temporal response of a filter with transfer function $V(f - f_1)$, evaluated at time t_1.

The time-frequency energy density function is defined so as to portray the variation of u(t) in the time and frequency domains by means of a three-dimensional plot (Gabor, 1946). The spectrogram is also constructed so as to display this variation. The spectrogram is obtained by passing the signal u(t) through a bank of filters that are all shifted versions of the same basic function V(f).

Historically, the ambiguity function was motivated by the need to detect a radar signal u(t) in Gaussian noise, and to simultaneously estimate its delay τ and Doppler shift, ϕ. If $V^*(f) = U^*(f)$, the filter $V^*(f-\phi)$ tests the hypothesis that the transmitted signal spectrum U(f) has been Doppler shifted by ϕ Hz. The radar receiver implements a sequence of such hypotheses by using a bank of filters with ϕ-dependent transfer functions. When u(t) is present at the input, the filter responses as a function of range (time) and Doppler shift (frequency) are represented by the ambiguity function. This interpretation of the ambiguity function explains its close connection to time-frequency signal representations.

Some relevant properties of time-frequency energy density functions, spectrograms, and ambiguity functions are listed in Appendix A. These properties show that the autoambiguity function of an input signal can be obtained from a noise-free spectrogram by deconvolution. A correlation operation can then be implemented by integrating the product of the data autoambiguity function and the conjugated autoambiguity function of a reference signal. Alternatively, one can obtain the input signal itself (apart from a complex constant) from the data autoambiguity function.

Several practical questions should be considered before the deconvolution and detection operations are implemented, or comparisons with animal audition are made. Filters with variable bandwidth should be considered (Section 2.2.3), and the required sampling rate for deconvolution should be examined (Section 2.2.4) The sampling question is important because two different rates apply for parsimonious spectrogram representation and for deconvolution, and because the higher sampling density for deconvolution

can be related to biological data. Finally, the idealized signal reconstruction process (Section 2.2.5) should be modified to take account of noise, and an error measure (e.g., mean-square error should be obtained in order to describe the effect of noise upon the reconstructed signal (Section 2.2.6).

2.2.3 Variable Bandwidth Filters

The list of properties in Appendix A pertains to a bank of constant bandwidth filters $V(f - f_1)$, where f_1 is varied. Similar results for filters with bandwidths that depend upon f_1 can sometimes be obtained (Flaska, 1976), but the derivations are more difficult and the properties are often not as simple as those given in the Appendix.

It is desirable to exploit the elegance and simplicity of the constant bandwidth case, but the analysis should also be sufficiently general to include frequency-dependent passbands. Fortunately, this problem has a simple solution. Before entering the filter bank, received spectra can be distorted by a nonlinear frequency transformation that effectively converts the constant bandwidth filters into frequency variable ones. For example, if proportional bandwidth filtering is desired, a given signal spectrum $G(f)$ is transformed into a new signal $U(f)$ such that

$$U(f) = \exp(f/2)\, G\left[\exp(f)\right] . \tag{10}$$

Since $|U(f)|^2$ is an energy spectral density, the transformation (10) should be energy-invariant, i.e., $E_u = E_g$, or

$$\int |U(f)|^2 \, df = \int \left|G\left[\exp(f)\right]\right|^2 \exp(f) \, df . \tag{11}$$

This condition is satisfied when the factor $\exp(f/2)$ is included in (10).

U(f) is obtained by plotting $f^{\frac{1}{2}}G(f)$ on a logarithmic frequency scale. Constant-bandwidth filtering of U(f) is equivalent to proportional bandwidth filtering of G(f). Alternatively, one can use

$$U(f) = \left[\cosh(f)\right]^{\frac{1}{2}} G\left[\sinh(f)\right] \tag{12}$$

to obtain a more realistic model of mammalian audition. Constant bandwidth filtering of U(f) now corresponds to constant bandwidth filtering of G(f) at low frequencies and proportional bandwidth filtering of G(f) at higher frequencies. The transformation in (12) was suggested at a recent conference (Bullock, 1977).

If no information is lost in a signal processing operation with U(f) as the input spectrum, then the original data G(f) can be obtained from a reconstructed version of U(f).

2.2.4 Sampling the Spectrogram

2.2.4.A Minimum Sampling Rates in Time and Frequency Directions

Consider (8) with f_1 fixed. Fixing f_1 corresponds to looking at a constant-f_1 profile of the spectrogram, i.e., at the squared envelope of the time response of the filter $V(f - f_1)$. The transfer function V(f) is a low pass function with two-sided bandwidth $2B_v$, i.e., $V(f) \equiv 0$ for $|f| > B_v$. For an analytic or one-sided representation, $V(f) \equiv 0$ for $f < 0$ and $f > B_v$, and $V(f - f_1) \equiv 0$ for $f < f_1$ and $f > f_1 + B_v$.

Let $x(t, f_1)$ be the real part of the filter response when the input signal is u(t), i.e.,

$$x(t, f_1) = \mathrm{Re}\left\{ \int_0^\infty U(f)\, V(f - f_1) \exp(j2\pi ft)\, df \right\} . \qquad (13)$$

Assuming that u(t) is always accompanied by a small amount of white noise, $x(t, f_1)$ has bandwidth B_v, and it should be sampled with $2B_v$ samples/sec. The analytic representation involves a complex time function, and to completely specify the filter response, we need to obtain samples of

$$y(t, f_1) = \mathrm{Im}\left\{ \int_0^\infty U(f)\, V(f - f_1) \exp(j2\pi ft)\, df \right\} . \qquad (14)$$

The real function $y(t, f_1)$ should also be sampled with $2B_v$ samples/sec. A total of $4B_v$ real-valued samples per second or $2B_v$ complex samples per second are therefore required.

When u(t) is a random process, it can be shown that $x(t, f_1)$ and $y(t, f_1)$ are statistically uncorrelated random variables (Appendix C). The complex samples $\left[x(t, f_1) + jy(t, f_1)\right]$ and $\left[x(t - k/2B_v, f_1) + jy(t - k/2B_v, f_1)\right]$ are also uncorrelated for any integer, nonzero k (Slepian, 1954; Cook and Bernfeld, 1967).

The squared envelope of the filter output is

$$S_{uv}(t, f_1) = \left[x^2(t, f_1) + y^2(t, f_1)\right] / 2 . \qquad (15)$$

The squaring operation doubles the bandwidth of $x(t, f_1)$ and $y(t, f_1)$. The sum in (15) has the same bandwidth as x^2 or y^2, and the squared envelope can therefore be represented by $4B_v$ real-valued samples per second.

A similar argument can be used to determine the minimum sampling rate for a constant-t_1 profile of $S_{uv}(t_1, f_1)$. It must be assumed, however, that the filter impulse response v(t) is time limited as well as band limited, and this situation cannot occur. An alternate assumption is that a function is adequately specified if it can be reconstructed to within a given small number ϵ, and that, for any filter of interest, there exist sampling densities in time and frequency such that ϵ is not exceeded. In other words, we assume the existence of numbers B_v and T_v such that v(t) is adequately specified by samples that are $1/(2B_v)$ apart and V(f) is adequately specified by samples that are $1/(2T_v)$ apart. B_v has been called the bandwidth of v(t), and T_v will be called the duration of v(t).

From (9), the time width of the real or imaginary parts of a constant-t_1 response is T_v, assuming that a small amount of noise is always added to u(t), so that the input data has infinite duration. Forming the squared envelope in the frequency domain leads to a constant-t_1 spectrogram profile with a time width of $2T_v$, and $S_{uv}(t_1, f)$ should be sampled at a rate of at least $4T_v$ samples/Hz.

In summary, the spectrogram can be represented by a grid of samples that are spaced $(4B_v)^{-1}$ seconds apart in the time direction and $(4T_v)^{-1}$ Hz apart in the frequency direction. In a noise-free situation, (8) and (9) imply that samples can be spaced $\left[4 \min (B_v, B_u)\right]^{-1}$ seconds apart in time and $\left[4 \min (T_u, T_v)\right]^{-1}$ Hz apart in frequency.

2.2.4.B Sampling Rates for Two-Dimensional Deconvolution

Property 1 in Appendix A states that the spectrogram can be formed by convolution of $e_{uu}(t, f)$ with $e_{vv}(t, f)$. From the definition

of the energy density function (4), $e_{uu}(t, f)$ and $e_{vv}(t, f)$ can be adequately represented by using a sampling rate in time that is at least 2 max (B_u, B_v) samples/sec and a sampling rate in frequency that is at least 2 max (T_u, T_v) samples Hz. These sampling rates are therefore required for two-dimensional digital convolution of $e_{uu}(t, f)$ and $e_{vv}(t, f)$. The output of the convolution process will be sampled at the same rates as above, and these rates will often be much larger than the minimum required sampling rates for the spectrogram ($4B_v$ samples/sec and $4T_v$ samples/Hz). High sampling rates are also required if $S_{uv}(t, f)$ is to be subjected to a digital deconvolution operation in order to separate $e_{uu}(t, f)$ from $e_{vv}(t, f)$. The estimate of $e_{uu}(t, f)$ after deconvolution may be under sampled unless the spectrogram is sampled at 2 max (B_u, B_v) samples/sec and 2 max (T_u, T_v) samples/Hz.

One way to obtain the relatively high sampling rates that are required for deconvolution is to interpolate between the more sparse samples that are required for efficient representation of the spectrogram. Another method is to construct a spectrogram that is deliberately over-sampled. Physically, a given frequency sample is obtained by using a given value of f_1 in (8), i.e., by using a filter with the specified center frequency. If frequency interpolation is not to be used, the filter center frequencies should be separated by at most

$$\Delta f_1 = \left[2 \max (T_u, T_v)\right]^{-1} \text{ Hz} . \tag{16}$$

For non-dispersive filters,

$$B_v = \alpha T_v^{-1} . \tag{17}$$

where $\alpha \approx 1$, and if $T_u >> T_v$,

$$\Delta f_1 \Big|_{T_u >> T_v} << B_v . \tag{18}$$

Since each filter has bandwidth B_v, it follows from (18) that the filter transfer functions must overlap in order to retain information about $e_{uu}(t, f)$, if frequency interpolation is not to be used. The required degree of overlap increases as the signal duration T_u becomes longer.

A wideband noise process (e.g., thermal noise) is usually added to the signal, u(t). In this case, a maximum signal bandwidth B_u and maximum signal duration T_u must be explicitly implemented by the receiver, since the data (signal plus noise) is neither band limited nor time limited. In order to apply sampling ideas, the receiver must introduce a window function or gate in the time domain. Most transducers experience a gradual decrease in sensitivity at sufficiently high frequencies, but for sampling purposes it may be advantageous to deliberately introduce a more rapid high frequency cutoff as well as a time window.

2.2.5 Reconstruction of a Noise-Free Signal from a Noise-Free Spectrogram

2.2.5.A Deconvolution

Under noise-free conditions, we wish to reconstruct the signal u(t) from the spectrogram $S_{uv}(t_1, f_1)$ where we know the filter function v(t). From Property 8 in Appendix A, $x_{uu}(\tau, \phi)$ can be obtained from the two-dimensional Fourier transform of the spectrogram. The

Fourier transformed spectrogram is divided by $X_{vv}(-\tau, \phi)$, where $X_{vv}(-\tau, \phi)$ can be formed from our assumed knowledge of v(t), as in (5). Dividing the Fourier transformed spectrogram by $X_{vv}(-\tau, \phi)$ yields $X_{uu}(\tau, \phi)$. From Property 9 in Appendix A, we can solve for u(t) by using (A14). The solution, however, will always be multiplied by a factor $\exp(j\lambda)$, where λ is an unknown constant. Under ideal conditions, we can therefore reconstruct the input signal u(t) from its spectrogram, except for an unknown, complex factor with unity magnitude.

Multiplication of the Fourier transform of the spectrogram by $\left[X_{vv}(-\tau, \phi)\right]^{-1}$ is a form of two-dimensional filtering. The realizability of the filter $X_{vv}(-\tau, \phi)^{-1}$ depends, in part, upon the absence of zeroes in the function $X_{vv}(\tau, \phi)$. Many filter functions v(t) give rise to ambiguity zeroes on the τ-ϕ plane. One function that does not produce ambiguity zeroes is a Gaussian pulse. The ambiguity function of such a pulse is Gaussian in two dimensions (Cook and Bernfeld, 1967).

2.2.5.B Sharpening-Up the "Smearing" Function, $e_{vv}(t, f)$

From Property 9, $e_{uu}(t, f)$ contains information about the signal, but Property 1 says that $e_{uu}(t, f)$ is "smeared" by two-dimensional convolution with $e_{vv}(t, -f)$ when a spectrogram is formed.

An intuitively appealing approach to signal reconstruction is thus to sharpen up the function $e_{vv}(t, f)$ in (A1).

From Property 11, we have

$$S_{uv}(t, f) * e^{*}_{vv}(-t, f) = e_{uu}(t, f) * \left| X_{vv}(t, f) \right|^{2}. \qquad (19)$$

The operation in (19) would provide an accurate version of $e_{uu}(t, f)$ if

$$\left|X_{vv}(t, f)\right|^2 \approx \delta(t)\,\delta(f) . \tag{20}$$

Unfortunately, such a condition is impossible because

$$\left|X_{vv}(0, 0)\right|^2 = E_v^{\,2} \tag{21}$$

(Woodward, 1964; Cook and Bernfeld, 1967), where

$$E_v = \int \left|v(t)\right|^2 dt . \tag{22}$$

It is possible to obtain a squared ambiguity function with a sharp maximum of height $E_v^{\,2}$ at $t = f = 0$, but $\left|X_{vv}(t, f)\right|^2$ will be nonzero over other parts of the t, f plane because of the volume invariance property

$$\iint \left|X_{vv}(t, f)\right|^2 dt\,df = E_v^{\,2} \tag{23}$$

(Woodward, 1964; Cook and Bernfeld, 1967).

Suppose that the ambiguity volume that is not present in the sharp spike at $t = f = 0$ is evenly distributed, i.e.,

$$\left|X_{vv}(t, f)\right|^2 = (\text{a narrow spike at } t = f = 0) + \\ + (\text{a constant for all } t, f) . \tag{24}$$

This is a so-called thumbtack ambiguity function, and it can be approximated with signals $v(t)$ that have large time-bandwidth

product. If $\left|X_{vv}(t,f)\right|^2$ has the form (24), then

$$e_{uu}(t,f) * \left|X_{vv}(t,f)\right|^2 = (e_{uu}(t,f) * \text{sharp spike}) + (e_{uu}(t,f) * \text{a constant}) \approx c_1\, e_{uu}(t,f) + c_2\, E_u$$

where c_1 and c_2 are constants and E_u is the energy of the signal, as in (A9). From (19)

$$S_{uv}(t,f) * e_{vv}^{*}(-t,f) - c_2\, E_u \approx c_1\, e_{uu}(t,f) \qquad (25)$$

if $\left|X_{vv}(t,f)\right|^2$ is described by (24).

Eq. (22) states that the time-frequency energy density function of the signal, $e_{uu}(t,f)$, can be approximated without a deconvolution process if $\left|X_{vv}(t,f)\right|^2$ resembles a thumbtack. The time-bandwidth product of v(t) must be large in order to obtain a thumbtack ambiguity function, and the operation in (25) is a two-dimensional pulse compression process.

2.2.6 Estimation of a Signal Waveform from Its Spectrogram in the Presence of Noise

The problem is now to estimate the signal from a noise-corrupted spectrogram. From Property 9, a simpler version of the problem is to estimate $e_{uu}(t,f)$ or $X_{uu}(\tau,\phi)$, the signal's energy density function or its ambiguity function. The spectrogram can be corrupted by noise that is external to the receiver and also by noise that is internal to the receiver.

Figure 1 shows the formulation of the problem and the assumed form of its solution. The data consists of a signal u(t) plus a sample function of noise n(t), and the data is represented by its energy density function or by its ambiguity function, which is the Fourier transform of its energy density function (Property 3). According to Property 8 in Appendix A, the Fourier transform of the data spectrogram is obtained by forming the product $X_{u+n,\,u+n}(\tau, \phi)\, X_{vv}(-\tau, \phi)$, where $X_{vv}(\tau, \phi)$ is the ambiguity function of the filter impulse response, v(t). The external noise n(t) is characterized by the function

$$N_e(\tau, \phi) = X_{nn}(\tau, \phi) + X_{un}(\tau, \phi) + X_{nu}(\tau, \phi) \tag{26}$$

and

$$X_{u+n,\,u+n}(\tau, \phi) = X_{uu}(\tau, \phi) + N_e(\tau, \phi)\,. \tag{27}$$

In addition to the external noise n(t), Figure 1 shows a second, independent noise source $N_i(\tau, \phi)$ that is added to the Fourier transform of the data spectrogram. $N_i(\tau, \phi)$ represents a sample function of the internal noise, which can be associated, for example, with quantization errors in the representation of the spectrogram.

An estimating filter $H(\tau, \phi)$ is used to form a minimum mean-square error (MMSE) estimate of the input function $X_{uu}(\tau, \phi)$.

The mean-square error is

$$\text{MSE} = E\left\{ \iint \left| X_{uu} - \left[(X_{uu} + N_e)\, X_{vv} + N_i \right] H \right|^2 d\tau\, d\phi \right\}. \tag{28}$$

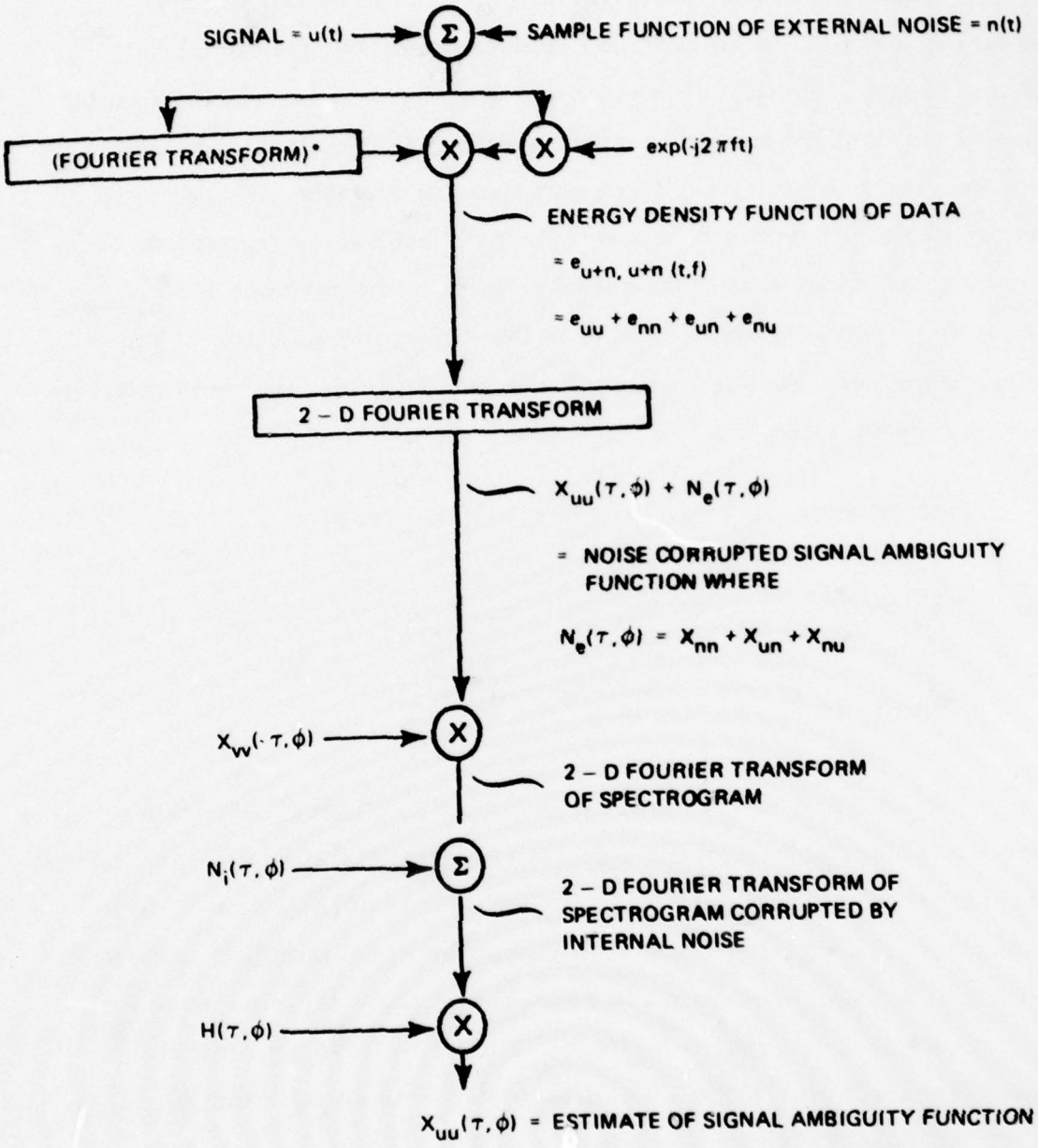


Figure 1. An estimation problem that is encountered when a signal or its autoambiguity function is reconstructed from a spectrogram.

In Appendix B, a variational approach is used to find $H(\tau, \phi)$ such that MSE is minimized. The resulting function $H_o(\tau, \phi)$ is

$$H_o(\tau, \phi) = \frac{\overline{\left|X_{uu}(\tau, \phi)\right|^2}\, X_{vv}^*(-\tau, \phi)}{\left|X_{vv}(-\tau, \phi)\right|^2\left[\overline{\left|X_{uu}(\tau, \phi)\right|^2} + P_e(\tau, \phi)\right] + P_i(\tau, \phi)} \qquad (29)$$

where $P_e(\tau, \phi) = E\{|N_e(\tau, \phi)|^2\}$, $P_i(\tau, \phi) = E\{|N_i(\tau, \phi)|^2\}$, and $\overline{\left|X_{uu}(\tau, \phi)\right|^2}$ is an expected value of $\left|X_{uu}(\tau, \phi)\right|^2$ that is based upon accumulated knowledge of the signals that are encountered in practice.

Substituting $H_o(\tau, \phi)$ from (29) back into (28), we obtain (see Appendix B)

$$MSE_o = \iint \overline{\left|X_{uu}\right|^2}\left[1 - \frac{\overline{\left|X_{uu}\right|^2}}{\overline{\left|X_{uu}\right|^2} + P_e + (P_i/\left|X_{vv}\right|^2)}\right] d\tau\, d\phi \,. \qquad (30)$$

If both P_e and P_i are zero, (29) is identical to (A13), $e_{uu}(t, f)$ is obtained by deconvolution, and $MSE_o = 0$. If P_i is zero but P_e is not, MSE_o in (30) is independent of the filter impulse response v(t). For no internal noise, any convenient filter function can be used, provided that $H_o(\tau, \phi)$ is bounded. This result is surprising, since it implies that the time and frequency resolution of the reconstructed signal energy density function $e_{uu}(t, f)$ is independent of the bandwidth of the analyzing filters! If there is no internal noise, spectrograms that are obtained with octave-wide filters yield the same mean-square error as spectrograms that are obtained with very

narrowband filters, even for narrowband signals! Without frequency interpolation, this result holds only if (16) is satisfied, and T_u will be large for narrowband signals. Although wideband filters can be used, their center frequencies must be closely spaced, and the resulting transfer functions will overlap.

The existence of internal noise can be advantageous from the viewpoint of realizability. If $P_i(\tau, \phi)$ has no zeroes, then the filter $H_o(\tau, \phi)$ will be bounded (i.e., realizable) even if $\left|X_{vv}(-\tau, \phi)\right|^2 = 0$ for many values of (τ, ϕ). The condition $P_i(\tau, \phi) \neq 0$ can therefore result in a larger set of admissible filter functions, v(t), such that $H_o(\tau, \phi)$ is bounded.

Another interesting condition occurs for relatively large internal noise power, i.e.,

$$P_i(\tau, \phi) >> \left|X_{vv}(-\tau, \phi)\right|^2 \left[\overline{\left|X_{uu}(\tau, \phi)\right|^2} + P_e(\tau, \phi)\right] . \qquad (32)$$

In this case,

$$H_o(\tau, \phi) \approx X_{vv}^{*}(-\tau, \phi) \left[\overline{\left|X_{uu}(\tau, \phi)\right|^2} / P_i(\tau, \phi)\right] . \qquad (33)$$

If $P_i(\tau, \phi) \propto \overline{\left|X_{uu}(\tau, \phi)\right|^2}$, (33) describes the "sharpening" operation in Property 11 of Appendix A and (19) – (25). If $P_i(\tau, \phi)$ is constant, (33) is a "sharpening" operation in cascade with a filter $\left|X_{uu}(\tau, \phi)\right|^2$.

In many situations, it may be desirable to reconstruct the noisy input data as accurately as possible. Filtering algorithms can then be applied to the reconstructed process. In this case, $P_e(\tau, \phi)$ is by definition equal to zero, since the desired signal is the data itself.

In summary, we can obtain an optimum (MMSE) reconstruction of $X_{uu}(\tau, \phi)$ by using the generalized deconvolution filter $H_o(\tau, \phi)$ in (29). In order that $X_{uu}(\tau, \phi)$ is not under-sampled, the spectrogram and the filter function should be sampled at 4 max (T_u, T_v) samples/Hz in the frequency direction and 4 max (B_u, B_v) samples/sec in time. In the absence of internal noise, the mean square error of the $X_{uu}(\tau, \phi)$ estimate is independent of the shape of the filter transfer function, V(f). In the presence of internal noise, the mean square error of the reconstruction process depends upon V(f). If a narrowband signal is masked by narrowband noise, (30) predicts a critical band effect, as shown in Appendix E.

2.2.7 Detection of Weak Signals in Noise

2.2.7.A Spectrogram Correlation

The spectrogram contains all the information about an input signal, except for a factor with unity magnitude and unknown phase. If a signal is known exactly, *or if some* of its characteristics (e.g., its bandwidth) are known, it should be possible to exploit this knowledge in order to detect the signal from a noisy spectrogram. The design and performance of such a detector are discussed in this section.

For detection of a known signal with spectrogram $S_{uv}(t, f)$, we can compare the likelihood ratio (Van Trees, 1968)

$$\Lambda = p(Z|S_{uv}, H_1) \,/\, p(Z|H_0) \tag{34}$$

with a threshold. In (34), $p(Z|S_{uv}, H_1)$ is the probability that the observed spectrogram $Z(t, f)$ will occur, given that the known signal,

as well as noise, is present (H_1 true), and $p(Z|H_0)$ is the probability that the observed spectrogram will occur, given that only noise is present (H_0 true).

It often happens that the form of the detector is dependent upon signal level, and there is no simple operation that can be applied to the data for all signal levels. In this case, one can resort to the use of a locally optimum detector, which is designed only for very small signal-to-noise ratio (Capon, 1961; Middleton, 1966). Although such a detector may be suboptimum for larger SNR, the use of an optimum statistic is most critical for small SNR, and suboptimum behavior for larger SNR is often acceptable.

A locally optimum detector is obtained by writing Λ in terms of an SNR parameter θ. The detection statistic is then

$$\partial \ln \Lambda(\theta)/\partial\theta \Big|_{\theta = 0} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma . \tag{35}$$

If the derivative of $\ln \Lambda$ with respect to θ is greater than a threshold γ when θ equals zero, then the test indicates that the signal is present. Otherwise, it is decided that the signal is absent.

It is shown in Appendix C that the left-hand side of (35) depends upon the quantity

$$\sum_i \sum_j z_{ij}\, s_{ij} \propto \iint Z(t, f)\, S_{uv}(t, f)\, dt\, df . \tag{36}$$

where $z_{ij} = Z(t_i, f_j)$ and $s_{ij} = S_{uv}(t_i, f_j)$ are samples of data spectrogram and of the noise-free signal spectrogram, respectively.

A locally optimum likelihood ratio test indicates that the spectrogram correlation process in (36) is an ideal method for detection of a signal with known spectrogram in additive, white, Gaussian noise, when signal-to-noise ratio is small. Samples of the data spectrogram $Z(t,f)$ are correlated with corresponding samples of the noise-free signal spectrogram $S_{uv}(t,f)$.

The spectrogram correlator can be adapted to non-Gaussian noise by using a nonlinearith other than an ideal square law device at the filter outputs. This modification is discussed by Poor and Thomas (1978).

Equation (36) implies that spectrogram correlation is ideal for low SNR, when the receiver is constrained to use the spectrogram as data. It will be shown, however, that the performance of a spectrogram correlator is inferior to that of an ideal detector which processes the data itself, when the signal is known exactly.

2.2.7.B Matched Filtering of Reconstructed Data

Pulse compression, correlation, or matched filtering is an ideal detection method that is used in many radar, sonar, and communication systems. Matched filtering is ideal under several criteria. The Bayes criterion seeks a detector that minimizes the expected risk which is incurred when costs are assigned to detection and false alarm probabilities. The Neyman-Pearson criterion seeks to maximize probability of detection for a given upper bound on false alarm probability. Both the Bayes and Neyman-Pearson criteria lead to a likelihood ratio test, and a correlation operation is an implementation of the likelihood ratio test for detection of a known signal in white, Gaussian noise. The signal-to-noise ratio

(SNR) criterion seeks a filter that maximizes output SNR for a known signal in white (not necessarily Gaussian) noise. The matched filter is the best linear filter for SNR maximization. Another criterion is not concerned with detection of the signal, but with maximum likelihood estimation of its epoch or time of arrival. In white Gaussian noise, the correlator is the required epoch estimator for a known signal.

In order to implement a matched filtering operation, it is necessary to reconstruct either the data, its energy density function, or its ambiguity function. Reconstruction of the data, rather than the signal, implies that $P_e(\tau, \phi) \equiv 0$ in (29) and (30). The filter in (29) is then used to obtain an estimate of the ambiguity function of the input data, $\hat{X}_{dd}(\tau, \phi)$, where

$$d(t) = u(t) + n(t) , \tag{37}$$

i.e., the data is the signal u(t) plus a sample function of the noise, n(t).

To implement a correlation operation, Property 15 in Appendix A can be applied. Letting $u_H(t) = u(t - \tau_H)$, where τ_H is a hypothetical delay, we have

$$\iint \hat{X}_{dd}(\tau, \phi)\, X^*_{u_H u_H}(\tau, \phi)\, d\tau\, d\phi \approx \left| \int d(t)\, u^*(t - \tau_H)\, dt \right|^2 \tag{38}$$

which equals $\left|R_u(\tau_H)\right|^2$, the signal autocorrelation function, when n(t) = 0. A similar result can be obtained from Property 14 if $\hat{e}_{dd}(t, f)$ is obtained from $\hat{X}_{dd}(\tau, \phi)$ by a two-dimensional Fourier transform.

It would appear from (38) that the unknown phase constant λ which appears in the reconstructed data (Property 9) does not affect the capability of the receiver to perform a pulse compression operation. To obtain further insight into this phenomenon, we shall consider detection of the signal from the reconstructed time function $\hat{d}(t)$, rather than from $\hat{X}_{dd}(\tau, \phi)$.

From Property 9, the reconstructed data $\hat{d}(t) \exp(j\lambda)$ can be obtained from the estimated ambiguity function $\hat{X}_{dd}(\tau, \phi)$, where $\hat{X}_{dd}(\tau, \phi)$ is constructed as in Figure 1 with $P_e(\tau, \phi) \equiv 0$. The reconstructed data $\hat{d}(t) \exp(j\lambda)$ can then be passed through a matched filter with impulse response $u^*(-t)$. The response of the filter in the absence of noise is $R_u(t) \exp j\lambda$, where $R_u(t) = X_{uu}(t, 0)$ is again the autocorrelation function of the signal. The envelope detected matched filter response is $|R_u(t)|^2$ in the absence of noise. The unknown, constant phase parameter λ is eliminated by envelope detecting the matched filter response, and pulse compression can be implemented even though λ is unknown.

2.2.7.C Estimator-Correlator Configuration for Detection of a Random Signal

Matched filtering is for detection of a known signal, but data reconstruction from the spectrogram can be used to detect a random or unknown signal as well as a known one. To detect a random signal, one can sometimes take advantage of prior information about the signal and noise in order to construct an estimate of the signal. The estimated signal is then correlated with the original data. This estimator-correlator configuration was first derived for Gaussian signals in Gaussian noise (Price, 1956; Kailath, 1960), but the

result has been generalized to include non-Gaussian signals in Gaussian noise (Kailath, 1969 and 1970; Scharf and Nolte, 1977).

In contemporary man-made systems, an estimate of the signal is usually obtained by using information about the power spectral densities of the signal and noise processes (Wiener filtering) or by using a model of the signal process that involves a linear system (state variable representation) driven by white noise (Kalman filtering).

Another specification of prior information about the signal process is knowledge of $\overline{|X_{uu}(\tau, \phi)|^2}$, the expected squared envelope of the ambiguity function that is constructed from a windowed version of the signal. A similar noise characterization uses

$$P_e(\tau, \phi) = \overline{|N_e(\tau, \phi)|^2},$$

where (26) shows that $P_e(\tau, \phi)$ is signal-dependent as well as noise-dependent. Given the spectrogram of the input data, $\overline{|X_{uu}(\tau, \phi)|^2}$ and $P_e(\tau, \phi)$ can be used to obtain an estimate of the signal ambiguity function $\hat{X}_{uu}(\tau, \phi)$ by using $H_0(\tau, \phi)$ in (29), and a signal estimate $\hat{u}(t) \exp(j\lambda_1)$ is obtained from $\hat{X}_{uu}(\tau, \phi)$ via Property 9. An estimate of the original data $\hat{d}(t) \exp(j\lambda_2)$ where $d(t)$ is given by (37), can be obtained from $\hat{X}_{dd}(\tau, \phi)$, where $\hat{X}_{dd}(\tau, \phi)$ is constructed from the spectrogram by letting $P_e(\tau, \phi) = 0$ in (29). Correlation of the estimated signal $\hat{u}(t) \exp(j\lambda_1)$ with the reconstructed data $\hat{d}(t) \exp(j\lambda_2)$ then gives a statistic that is often suitable for detection of a random signal. Envelope detection of the correlator output eliminates the effect of the unknown, constant phase parameters λ_1 and λ_2 that are introduced by the reconstruction processes.

An alternate form of the estimator-correlator is obtained by integrating the product $\hat{X}_{uu}(\tau, \phi)\, \hat{X}_{dd}(\tau, \phi)$, as in Property 15. In either case, knowledge of $|X_{uu}(\tau, \phi)|^2$, rather than $|U(f)|^2$ or a state variable model, is used to form an estimate of the unknown or random signal process for use in an estimator-correlator detector.

When does a receiver know the expected ambiguity function of a random signal? The answer lies in an input-output relation for time-varying random channels (Bello, 1963; Daly, 1970; Venetsanopoulos, 1978). The expected ambiguity function of the output of a time-varying random channel with transfer function $C(t, f)$ is given by the product of the expected ambiguity function of the input signal and the channel time-frequency autocorrelation function $R_C(\alpha, \beta)$, where

$$R_C(\alpha, \beta) = E\{C^*(t, f)\, C(t+\alpha,\ f+\beta)\} \quad .$$

2.2.7.D Relative Performance of a Matched Filter and a Spectrogram Correlator when a Signal Is Known Exactly

Appendix C indicates that spectrogram correlation is a locally optimum detection process if only the spectrogram of the data is available. Such a process is <u>not</u> ideal, however, if the data itself, or the data autoambiguity function, is available. When one has access to the data or to its autoambiguity function, a matched filter, or the equivalent process in (38), is the best detector for a known signal, under a variety of criteria. For low internal noise, the data or its autoambiguity function can be reconstructed from the spectrogram by two-dimensional deconvolution. Given a choice between non-ideal spectrogram correlation and a more complicated ideal operation, it is important to know the improvement in performance that can be expected if one implements the ideal detector.

In order to easily determine the performance of the spectrogram correlator, the following three assumptions will be made:

1. It will be assumed that the reference spectrogram $S(t, f)$ and the data spectrogram $Z(t, f)$ are represented by samples $s_{ij} = S(t_i, f_j)$ and $z_{ij} = Z(t_i, f_j)$ that are $(4B_v)^{-1}$ seconds apart in time and $(4T_v)^{-1}$ Hz apart in frequency.

2. It will be assumed that s_{ij} must equal either zero or one, i.e., that a one-bit quantization of the reference spectrogram is used. This assumption simplifies the calculations without leading to unrealistic results (Hagan and Farley, 1973). It will also be assumed that there are K independent samples, i.e., that $s_{ij} = 1$ over a total of K different values of i and j.

3. It will be assumed that the signal is power limited, but that it is designed for maximum energy E_u for a given peak power P_u and duration T_u. Since

$$E_u = \int_0^{T_u} |u(t)|^2 \, dt \leq \max_{0 \leq t \leq T_u} |u(t)|^2 \, T_u ,$$

maximum energy is obtained for a rectangular envelope $|u(t)|$, such that $|u(t)|^2 = P_u$ for t in $[0, T_u]$ and

$$E_u = P_u T_u . \tag{39}$$

The sampling rates in assumption 1 are such that the spectrogram can be written as a sum of orthonormal sinc functions with coefficients that correspond to the sample values. When white, Gaussian noise is present at the input to the system, the samples z_{ij} are statistically independent (Slepian, 1954).

The second assumption implies that the correlation of a measured spectrogram $Z(t,f)$ with the reference spectrogram $S(t,f)$ can be written as the sum

$$\sum_{i,j} z_{ij} s_{ij} = \sum_{(i,j)'} z_{ij} \tag{40}$$

where $(i,j)'$ is the set of i,j such that $s_{ij} = 1$.

For Gaussian input noise, the energy detected filter outputs z_{ij} are distributed as in (C7) or (C8), depending upon whether the signal is present (H_1 true) or absent (H_0 true). Since the samples z_{ij} are statistically independent, the sum in (40) has chi-square distribution under H_0 and noncentral chi-square distribution under H_1. The performance of the resulting detector for a random signal with bandwidth $B_u \leq B_v$ and duration $T_u \geq T_v$ has been analyzed by Urkowitz (1967). For this case, the sum over $(i,j)'$ in (40) becomes a sum over i, with a fixed value of j corresponding to the center frequency f_j of the appropriate filter.

The probability distributions that are obtained from (40) have K degrees of freedom, where K is the number of independent spectrogram samples that are added together in (40). For a spectrogram that is K samples in duration and one sample wide in frequency,

$$K = (4B_v)T_u . \tag{41}$$

In the paper by Urkowitz, $K = 2B_v T_u$ because the baseband filter $V(f)$ is assumed to have two-sided bandwidth B_v, rather than $2B_v$.

Although K and E_u are often treated as independent variables, they are linearly related when the signal has a rectangular envelope

with specified power, as in assumption 3. Solving (41) for T_u and substituting the result into (39), we obtain

$$E_u = K P_u / (4B_v) . \tag{42}$$

Following Urkowitz (1967), we define the signal-to-noise ratio as

$$\begin{aligned} SNR &= E_u / (N_o/2) . \\ &= K P_u / (2B_v N_o) \end{aligned} \tag{43}$$

where $N_o/2$ is the two-sided noise power spectral density. For convenience, we define

$$P_u \equiv 2B_v N_o \tag{44}$$

and

$$SNR = K . \tag{45}$$

For the constant-envelope signal, the one-bit quantization for s_{ij} (assumption 2) is no longer an approximation, but is an exact representation of the noise-free spectrogram.

The above formulation allows easy comparison of the performance of a matched filter and spectrogram correlator, since the spectrogram correlator has been made identical to an energy detector. Figure 2 illustrates probability of detection (P_D) for the matched filter (dotted lines) and energy detector (solid lines) as a function of $(SNR)^{\frac{1}{2}}$, for various values of false alarm probability (P_F). The matched filter performance was obtained from Van Trees (1968), and the energy detector performance was obtained from the ROC curves in the paper by Urkowitz (1967).

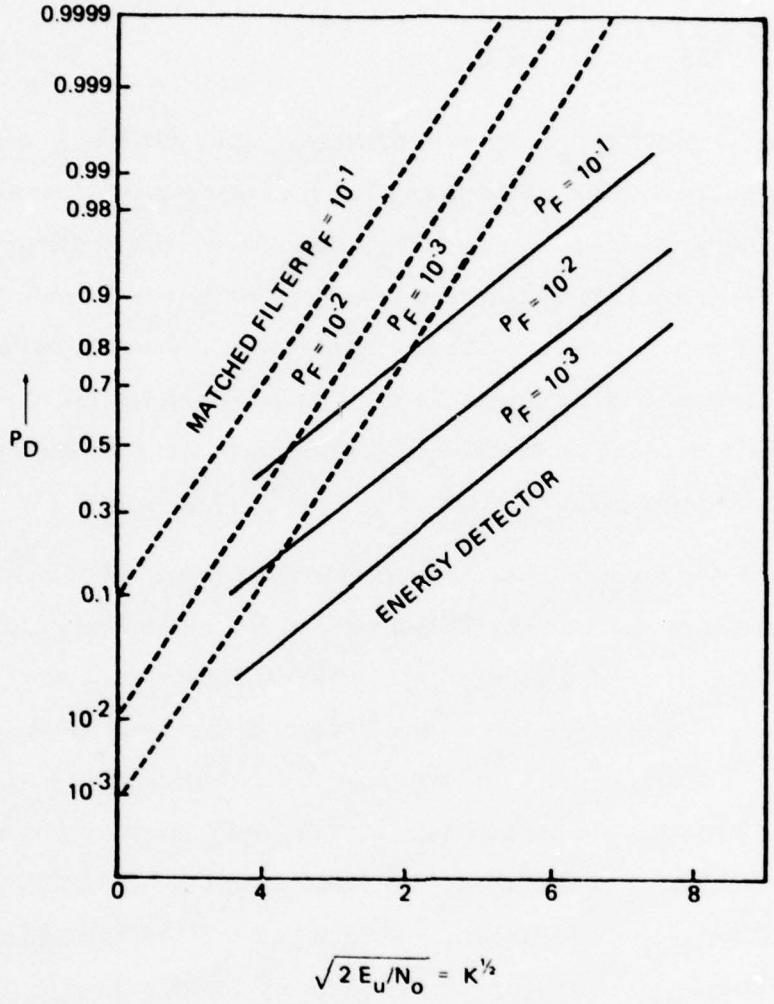


Figure 2. Probability of detection vs. square root of SNR for a signal with constant power, such that energy increases with the number of samples, K. Solid lines: Energy detector performance. Dotted lines: Matched filter performance.

Another comparative measure is obtained by plotting the factor α such that

$$\alpha (SNR)_{MF} = (SNR)_{ED} \tag{46}$$

as a function of $(SNR)_{MF}$, where $(SNR)_{MF}$ and $(SNR)_{ED}$ are the values of signal-to-noise ratio that give the same performance, i.e., the same values of P_D and P_F, for the matched filter (MF) and energy detector (ED). Since the energy detector is suboptimal, we expect that $\alpha \geq 1$, i.e., a larger SNR is required in order that an energy detector can perform as well as a matched filter. Figure 3 shows α versus $(SNR)_{MF}$. The curve is a composite of points obtained from Figure 2 for $P_F = 10^{-3}$, 10^{-2}, and 10^{-1}.

Figure 3 indicates that, for more than four observations, the performance of the energy detector can be made equal to that of the matched filter, if approximately three times as many observations are made. If the results are not very dependent upon (44), it would appear that the signal power must be increased by 5 dB in order to make the spectrogram correlator equivalent to a matched filter. The two-dimensional deconvolution process in (A13), followed by a matched filtering operation, results in an effective increase of 5 dB in signal energy, if internal noise is negligible.

2.2.8 Array Processing for Direction Estimation

Suppose that two or more spatially separated sensors are used for reception of an unknown acoustic signal, and that an estimate of the direction of the signal source is desired. Hahn and Tretter (1973) and Hahn (1975) have shown that cross-correlation

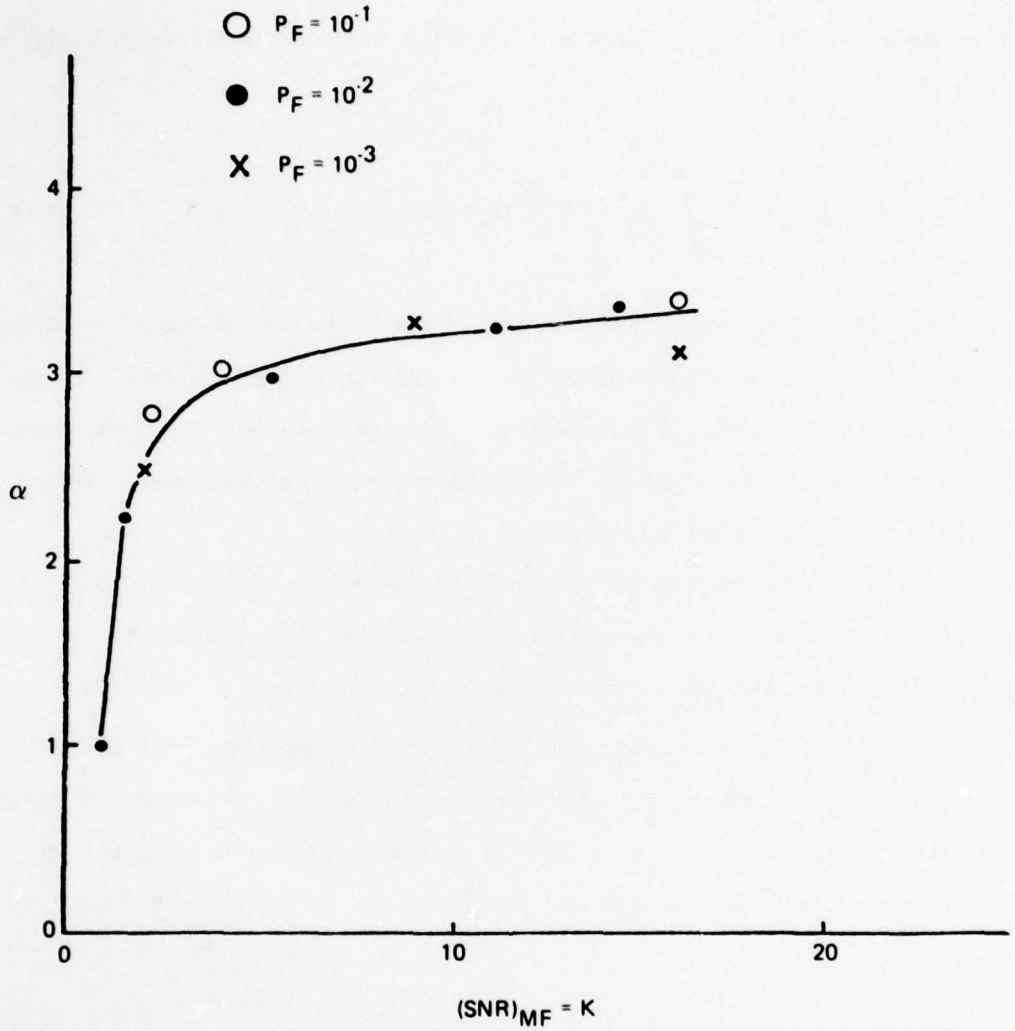


Figure 3. For a given detection perfomance (specified values of P_D and P_F), the required signal-to-noise ration for a matched filter is $(SNR)_{MF}$. For a signal with constant power, the same performance can be obtained with an energy detector, provided $(SNR)_{MF}$ is multiplied by a factor α. Solid line: α vs. $(SNR)_{MF}$ for $P_u = 2B_vN_o$.

of the outputs of pairs of sensors leads to an efficient estimate of relative delay. For a pair of sensors with outputs $d_1(t)$ and $d_2(t)$, we want to form

$$R_{d_1 d_2}(\Delta\tau_H) = \int d_1(t)\, d_2^*(t + \Delta\tau_H)\, dt \qquad (47)$$

where $\Delta\tau_H$ is any hypothesized relative delay between the two sensor responses. The cross-correlator responses are then linearly combined to form estimates of the relative delays between the responses of the various sensors and the response of the first sensor. Unspecified limits of integration correspond to $(-\infty, \infty)$.

If the signal $u(t)$ is known except for the direction of its source, one can estimate direction by means of a correlation processor. Delay variables $\tau_n(\theta_H)$ are generated from a direction hypothesis θ_H and a knowledge of the relative spatial positions of the sensors. The data $d_n(t)$ from the nth sensor is correlated with $u\left[t - \tau_n(\theta_H)\right]$, and the result is added to the correlator outputs of the other sensors. The resulting correlation function is (Altes, 1978)

$$R_{\underline{d}\,\underline{u}}(\theta_H) = \sum_n \int d_n(t)\, u_n^*\left[t - \tau_n(\theta_H)\right] dt\,. \qquad (48)$$

The operations in (47) and (48) yield efficient direction estimates in white Gaussian noise that is independent from sensor to sensor. Both of the operations involve correlation functions, not the envelopes of these functions. The inadequacy of using envelopes becomes obvious if one considers a complex sinusoid

$$d_n(t) = a(t + \tau_n) \exp\left[j2\pi f_o (t + \tau_n) \right]$$

where a(t) is a smooth, slowly varying envelope. An accurate delay estimate is available only if the phase of the correlation process is preserved. In (47), for example,

$$R_{d_1 d_2}(\Delta\tau_H) = \int a(t)\, a^*(t + \Delta\tau_H - \Delta\tau)\, dt$$

$$\exp\left\{ -j \left[2\pi f_o (\Delta\tau_H - \Delta\tau) \right] \right\}$$

where $\Delta\tau = \tau_1 - \tau_2$. The phase of the autocorrelation function is a sensitive measure of $\Delta\tau_H - \Delta\tau$, but the phase is time invariant and is eliminated when $R_{d_1 d_2}(\Delta\tau_H)$ is envelope detected.

Operations that involve the envelopes of correlation functions, as in (A19), (A21), and (A23), cannot be used for accurate narrowband direction estimation. In particular, cross-correlation of spectrograms, as in Property 12 of Appendix A, is not suitable for array processing of narrowband signals. Spectrogram cross-correlation cannot be substituted for cross-correlation of the signals themselves because narrowband delay information is destroyed when phase is eliminated.

Since an unknown constant λ is added to the phase of a signal that is reconstructed from a spectrogram, it would appear that such reconstructed signals cannot be used for direction estimation. This observation is considered in more detail in Appendix D.

A sufficient amount of phase information can be obtained if phase, as well as amplitude, can be observed at the output of only

one filter $V(f - f_j)$, provided there is signal energy at f_j. This statement follows from the fact that λ is invariant with frequency. If λ can be measured for one frequency f_j, it is known for the whole spectrum. This idea is discussed in more detail in Appendix D. If the constant λ that is added to the phase of a reconstructed signal is known, then an input signal can be completely determined from its spectrogram. A phase measurement from a single filter then provides sufficient information such that reconstructed narrowband signals from spatially separated sensors can be used for direction estimation.

2.2.9 Application to Theories of Hearing and Animal Echolocation

2.2.9.A Overlapping Filters, Signal Duration, and Just-Noticeable Frequency Difference

Sampling rates for spectrograms were discussed in Section 2.2.4. Assuming that frequency interpolation is not used, (16) to (18) indicate that, for two-dimensional deconvolution, there must be significant overlap of the transfer functions $V(f - f_j)$. The maximum signal duration is determined by the spacing Δf_1 between center frequencies of adjacent filters. For input signals that are longer than $(2\Delta f_1)^{-1}$ seconds, the spectrogram will be undersampled insofar as deconvolution is concerned, and the input signal cannot be reconstructed. In this case, correlation processing of reconstructed signals cannot be used, but spectrogram correlation can be used instead.

The theory of signal reconstruction from spectrograms therefore leads to the following predictions:

1. There will be significant overlap between filters that are used to construct the spectrogram.

2. There will be a maximum signal duration $T_{max} = (2\Delta f_1)^{-1}$ beyond which matched filtering of reconstructed signals cannot be used. Signals that are longer than T_{max} can be detected with spectrogram correlation, which is not as efficient as matched filtering.

3. The quantity Δf_1 is the difference between center frequencies of adjacent filters, and it corresponds to the sample spacing for U(f), the Fourier transform of the reconstructed signal. The just noticeable frequency deviation of a frequency modulated (FM) signal should therefore correspond to Δf_1, where $\Delta f_1 = (2T_{max})^{-1}$.

Assuming that a spectrogram is constructed from the responses of hair cells along the basilar membrane and/or peripheral tuned neurons, there is indeed significant overlap between the filters that are used to construct the spectrogram (Evans, 1977; Pfeiffer and Kim, 1975).

Detectability versus duration-of-tone data for human listeners shows a change in slope between 107 msec and 277 msec duration (Green, Birdsall, and Tanner, 1957) for detection of a 1 kHz tone. This change could indicate a transition from matched filtering to spectrogram correlation. The Δf_1 values corresponding to T_{max} = 107 to 277 msec are Δf_1 = 1.8 Hz to 4.7 Hz. For carrier frequencies below 2 kHz, the just noticeable frequency

deviation of an FM signal is between 1 Hz and 4 Hz (Shower and Biddulph, 1931; Fletcher, 1953).

2.2.9.B Relation Between Filter Shape and Frequency Discrimination Capability

A discrepancy exists between the relatively wide bandwidths of neural tuning curves (Evans, 1977; Pfeiffer and Kim, 1975) or critical bands (Zwicker, 1961), and the ability of a listener to perform fine frequency discriminations (Shower and Biddulph, 1931; Fletcher, 1953; Schafer, et al., 1950). This discrepancy is often explained by remarking that the sharp high-frequency cutoff of neural tuning curves allows for accurate frequency discrimination, even though the tuning curves themselves have wide bandwidths. Although this explanation is certainly plausible, the theory of signal reconstruction from spectrograms provides an alternative viewpoint. For zero internal noise, (30) indicates that the accuracy of a reconstructed signal (as measured by minimum mean-square error) is independent of the shape or bandwidth of the filters $V(f - f_j)$ that are used to construct the spectrogram. If internal noise can be neglected, it follows that frequency resolution does not depend upon narrow filter bandwidths or sharp, high frequency cutoffs.

It can be conjectured that the sharp, high frequency cutoff that is often observed in neural tuning curves is such as to minimize the mean-square error in (30) when internal noise cannot be neglected. Rapid cutoff may also be useful for low pass filtering of band limited signals. Conversely, one should be ready to accept measurements that do <u>not</u> reveal a rapid, high frequency cutoff, e.g., the revcor functions of de Boer and de Jongh (1978).

2.2.9.C Critical Bands

For masking of narrowband signals by narrowband noise, there is a critical ratio of signal center frequency to noise center frequency such that the masker has little or no effect (Schafer, et al., 1950). If this ratio is approximately invariant with signal frequency, one can postulate the existence of filters with critical bandwidths that are proportional to frequency (Zwicker, 1961). In Appendix E, it is heuristically argued that a critical band effect will occur when a signal is estimated from the spectrogram of noise-corrupted data, and it is shown that critical bands can also result from a spectrogram correlation operation.

2.2.9.D Monaural Phase Deafness

By using Properties 8 and 9 in Appendix A, it has been shown that a complex analytic signal $u(t) \exp(j\lambda)$ can be reconstructed from the spectrogram $S_{uv}(t, f)$, where λ is an unknown constant. If the spectrogram model applies to human audition, it follows that people should be monaurally insensitive to an arbitrary phase shift λ. For example, $u_r(t)$, $\hat{u}_r(t)$, $-u_r(t)$, and $-\hat{u}_r(t)$ should be indistinguishable, where $\hat{u}_r(t)$ is the Hilbert transform of $u_r(t)$.

This predicted phase insensitivity is more restrictive than insensitivity to the frequency dependent phase transformation that has been suggested by Schroeder (1975). Schroeder, however, has pointed out that the experiments of Craig and Jeffress (1962) and Terhardt and Fastl (1971) indicate a sensitivity to his more general frequency dependent phase transformation. The experiments, however, do not indicate any sensitivity to the frequency invariant phase shift that transforms $U(f)$ into $U(f) \exp(j\lambda)$.

2.2.9.E Envelope Variations and Phase-Locked Neural Firing Patterns

Hartmann (1978) has recently found that human listeners are sensitive to the form of the envelope of a sinusoid, u(t). One of Hartmann's explanations for this phenomenon is that the time-window spectrogram

$$S^{T}_{u(t),\, w(-t)}(t_1, f_1) = \left| \int w(t_1 - t)\, u(t) \exp(-j2\pi f_1 t)\, dt \right|^2$$

is different for different envelopes, and that the human auditory system may transform input signals into such a function. Eq. (A25) shows that

$$S^{T}_{u(t),\, w(-t)}(t_1, f_1) = S_{uv}(t_1, f_1)$$

if

$$w(t) = v(t),$$

where $S_{uv}(t_1, f_1)$ is the frequency-window spectrogram in (8). Hartmann's results can thus be explained in terms of a frequency-window spectrogram model of audition.

The equality w(t) = v(t) also implies that one can find a more realistic weighting function model than an exponential, which is often employed for generation of time-window spectrograms (Hartmann, 1978; Fletcher, 1953; Sayers and Cherry, 1957). For example, one can use a baseband version of a reverse-correlation impulse response (de Boer and de Jongh, 1978).

Siebert (1970) has applied Cramér-Rao bounds to rate-modulated Poisson processes, in an attempt to deduce the

mechanisms for auditory frequency discrimination. He has shown that utilization of the timing information from phase-locked neural firing patterns should improve frequency discrimination capability. Anderson, et al. (1971) have discovered such phase-locked neural responses for frequencies up to 4 to 6 kHz. Despite these observations, Siebert has concluded that phase-locked neural firing patterns are not used for auditory frequency discrimination. Monaural frequency discrimination is apparently based upon the envelopes of the neural filters' responses, or upon analysis of (i) the spectrogram, (ii) a reconstructed signal ambiguity function, or (iii) the reconstructed signal itself.

2.2.9.F Energy Detection, Spectrogram Correlation, or Matched Filtering?

For signals that are longer than $(\Delta f_1)^{-1}$, where Δf_1 is the difference between center frequencies of adjacent filters, the sample density that is required for signal reconstruction exceeds the sample density in the frequency direction that is used to characterize the spectrogram. In this case, the input signal cannot be reconstructed from the spectrogram unless the data is time-gated, and spectrogram correlation becomes an attractive alternative. It has been demonstrated, however, that spectrogram correlation is sometimes the same thing as energy detection. The two operations are identical when all nonzero samples of the reference spectrogram have the same magnitude and when signal duration is prespecified.

For signal durations that are longer than ~200 msec, one must resort to spectrogram correlation. For shorter durations, however, matched filtering can theoretically be performed on a reconstructed version of the signal. Does the data support such a matched filtering hypothesis?

Gated sinusoids have often been used as stimuli for monaural detection experiments. A matched filter for a short duration sinusoid has wider bandwidth than a matched filter for a sinusoid with longer duration. The signal reconstruction/matched filter hypothesis therefore predicts that a listener's performance should be similar to that of a filter followed by a threshold device, and that the corresponding filter should have broader bandwidth as signal duration is decreased. In fact, several researchers have interpreted their masking data in this way, i.e., they have found that a filtering model is relevant, provided that the filter bandwidth broadens as the duration of a sinusoidal signal becomes smaller, for durations less than 100 msec (Srinivasan, 1971).

Although effective bandwidths increase as signal duration becomes smaller, the bandwidths are dependent upon the frequency of a tone as well as upon its duration, and the bandwidths are generally larger than those of a matched filter. For an 11-msec tone, Srinivasan (1971) has reported effective bandwidths that are roughly 165 Hz wide for a tone frequency of 880 Hz and 180 Hz wide at 1110 Hz, for a rectangular filter transfer function model. The matched filter for the tone has a bandwidth of roughly 90 Hz. A 90-Hz bandwidth is observed for a 23-msec tone, which should correspond to a 39-Hz bandwidth under a matched filter hypothesis. The observed bandwidths are therefore approximately twice the bandwidth of an ideal matched filter.

Curves showing probability of a correct response P(C) versus SNR can be analytically obtained for a two-alternative, forced-choice (2AFC) experiment, as discussed in Appendix F. The curve for a matched filter always has steeper slope than the energy detector curve. From Figure 8-3 of Green and Swets

(1966), one can see that a best fit to the experimental P(C) vs. SNR data for detection of a gated sine wave gives a curve with a slightly steeper slope than that predicted by an energy detector. This steeper slope may be indicative of an imperfect matched filter (e.g., one with too wide a bandwidth) or of an estimator-correlator configuration.

The detection performance of human observers for a sum of gated sine waves with different frequencies has been investigated by Green (1958) and by Green, McKey, and Licklider (1959). Performance is in reasonably close agreement with a detector that sums the squared energies of the M different sinusoids, i.e., a detector that computes $\sum_{j=1}^{M} E_j^2$. A matched filter would linearly combine the energies of the different signal components, i.e., its performance would depend upon $\sum_{j=1}^{M} E_j$. Multiple harmonics are therefore apparently not detected with a matched filter mechanism in humans.

A spectrogram correlator computes a weighted sum of the filter response envelopes z_{ij}, where the weights are proportional to the expected values of these responses in the absence of noise, s_{ij} (see Appendix C). When the sinusoidal signal component that excites the jth filter has a rectangular envelope, then $s_{ij} \equiv s_j$ for all time samples $i = 1, 2, \ldots, N$, and

$$\sum_{j=1}^{M} \sum_{i=1}^{N} s_{ij} z_{ij} = \sum_{j=1}^{M} s_j \left(\sum_{i=1}^{N} z_{ij} \right) .$$

When multiple components with different energy levels are present, i.e., when s_j is different for different j-values, then a spectrogram correlator is no longer identical to an energy detector, as was the case for a single sinusoid with rectangular envelope and specified duration.

Since s_j and the expected value of $\sum_{i=1}^{N} z_{ij}$ are both proportional to the energy E_j of the jth sinusoidal signal component, we have

$$\overline{\sum_{j=1}^{M}\sum_{i=1}^{N} s_{ij}\, z_{ij}} = \sum_{j=1}^{M} s_j \left(\overline{\sum_{i=1}^{N} z_{ij}}\right) \propto \sum_{j=1}^{M} E_j^{\,2}.$$

A spectrogram correlator computes a weighted sum of envelope detected filter responses, and the expected value of this weighted sum is proportional to the sum of the squared energies in the sinusoidal signal components. This behavior is similar to human detection of a superposition of gated sine waves (Green, 1958).

An estimator-correlator that uses the spectrogram as data functions like a spectrogram correlator, except that it has no prior knowledge of the noise-free filter response envelopes s_{ij}. The device forms

$$\sum_{j=1}^{M}\sum_{i=1}^{N} \hat{s}_{ij}\, z_{ij} = \sum_{j=1}^{M} \hat{s}_j \left(\sum_{i=1}^{N} z_{ij}\right)$$

where $\hat{s}_j$ is an estimate of s_j that is obtained from the data spectrogram, $Z(t_i, f_j)$. This estimate is based upon the conditional mean of the data spectrogram (the mean of the probability distribution that

describes z_{ij} when a signal is present). Since $\hat{s}_j$ should be proportional to E_j, a spectrogram estimator-correlator is again expected to behave like a detector that computes the sum of the squared energies of the sinusoidal signal components.

2.2.9.G Echolocation

Bats and dolphins echolocate with ultrasonic signals (Griffin, 1958; Evans, 1973; Airapetyants and Konstantinov, 1970). Bats, in particular, use signals that are suggestive of a pulse compression or matched filtering capability (Cahlander, 1967; Kroszczynski, 1969; Altes and Titlebaum, 1970, 1975). Echolocation performance of some bats is also indicative of such a capability (Simmons, 1973).

Ultrasonic signals are unlikely to possess sufficient energy below 6 kHz such that the unknown phase, λ, in the reconstructed waveform can be estimated, when echoes are obtained from their spectrograms. It was shown in Section VII B that pulse compression can nevertheless occur without knowledge of λ, and λ is eliminated by envelope detecting the matched filter response.

The possible existence of an ideal detector in animal echolocation can therefore be theoretically justified by a signal reconstruction argument, even if echoes are initially transformed into spectrograms by the peripheral auditory system. It was shown in Section 2.2.7.D that if a spectrogram correlation process is used instead of reconstruction/matched filtering, the penalty is an approximate 5 dB loss in signal-to-noise ratio.

Ideal detection is especially useful for sonar targets that are far away from the animal. Bats tend to use comparatively long-duration "cruising pulses" to detect such targets, i.e., the signal

energy is made relatively large even though peak power is constrained. Figure 2 shows that this strategy improves the detection performance of both matched filter and energy detector (or spectrogram correlator). If only spreading loss is considered, a 5 dB improvement in SNR increases the range of a sonar by one third, i.e., $10 \log \left[(R_1/R_2)^4 \right] = 5\,\text{dB}$ when $R_1 = 1.33\, R_2$.

In the special case of animal echolocation, a phase change that is introduced by a target can affect the energy spectral density of the signal-echo pair, when both transmitted and received pulses are processed as one waveform. For a target that is, say, 15 m from a dolphin, the transmitted signal and the echo can be processed as a single waveform if the integration time of each spectrogram filter is at least 20 msec, which seems very reasonable. Johnson and Titlebaum (1976) have used this concept to formulate the hypothesis that range measurement in animal echolocation may be associated with estimation of time separation pitch. The theoretical phase sensitivity of the process is demonstrated as follows.

Let $u(t)$ be the transmitted pulse and let the echo be $\alpha u(t - \tau) \exp(j\lambda_1)$, where α is an attenuation factor, τ is delay, and λ_1 is a phase shift that is introduced by the reflection process. The magnitude-squared Fourier transform of $[u(t) + \alpha u(t - \tau) \exp(j\lambda_1)]$ is $|U(f)|^2 [1 + \alpha^2 + 2\alpha \cos(2\pi f \tau - \lambda_1)]$, where $U(f)$ is the Fourier transform of $u(t)$. The phase factor λ_1 thus influences the shape of the energy spectral density of the signal-echo pair. The energy spectral density can be estimated from the spectrogram by using property 5 in Appendix A.

The phase parameter λ_1 should be especially useful for underwater echolocation, since it can be used to tell the difference between a target with large acoustic impedance (e.g., a rock) and

one with small impedance (e.g., an air bladder). Phase is also a measure of radial velocity if a linear period modulated signal (a typical FM bat pulse) is transmitted (Altes and Skinner, 1977). Spectrogram processing does not necessarily eliminate information about the relative phase between two pulses, and this information may be useful for echolocation.

2.2.9.H Binaural Processing for Direction Estimation

In Section 2.2.8, it was shown that direction estimation of narrowband signals is theoretically dependent upon a measurement of relative phase between sensors, i.e., between the two ears. Although this phase information is destroyed in the formation of a separate spectrogram at each ear, it can be regained if the response of one or more filters (or tuned neurons) can be observed before envelope detection takes place (Appendix D).

The localization experiments of Sayers and Cherry (1957) indicate that binaural direction estimation in humans can be accomplished with sine waves that have equal amplitude at both ears. Since sensitivity to a phase difference between the ears is required for such an estimate, binaural processing in humans must utilize phase-locked responses such as those measured by Anderson, et al. (1971).

Phase-locked responses occur only at low frequencies (less than 4 to 6 kHz). It is shown in Appendix D, however, that a phase measurement at a single, low frequency sufficies to completely characterize the input signal at all frequencies. The observed phase sensitivity in binaural interaction can thus be explained by cross-correlation of signals that are reconstructed from spectrograms, provided that low-frequency, tuned-neuron

responses are used to solve for the unknown phase parameters in the reconstructed signals.

2.2.9.I Implementation of a Two-Dimensional Deconvolution Operation

Reconstruction of an input signal or its ambiguity function from a spectrogram requires two-dimensional deconvolution. Most engineers think of implementing such an operation by means of Fourier transforms. Approximate deconvolution techniques, however, are not nearly so complicated (Papoulis, 1972a). The apparent complexity of the reconstruction operation, then, is not necessarily a deterrent to its implementation in biological systems. Another approximate reconstruction method has been used in speech vocoding (Flanagan, 1965).

Efficient reconstruction methods and the accuracy of approximate methods may depend upon the form of the signal itself. Are echolocation signals designed so that echo spectrograms can be easily transformed back into the original echo data? Are some signals (e.g., frequency modulated "chirps") more easily detected than others (e.g., sinusoids)?

2.2.9.J Design of Hearing Aids

Signal reconstruction from spectrograms should be useful for hearing aid design. Suppose that the auditory system actually forms a spectrogram of an input signal. A corresponding spectrogram can be synthesized with man-made filters that are as similar as possible to those of the listener. The synthesized spectrogram can then be altered so as to improve the listener's internal representation of the signal. A new signal is constructed from the altered

spectrogram by using the reconstruction operations in Properties 8 and 9 of Appendix A. The new signal is then presented to the listener.

The philosophy for this kind of hearing aid design is to transform the signal into a realistic perceptual space, to operate upon the transformed signal in this perceptual space, and then to convert the result back into an acoustic waveform for presentation to the listener. The last step in this process depends upon the reconstruction of a signal from its spectrogram.

2.2.9.K Time and Place Theories of Audition

In order to reconstruct a signal from its spectrogram, one must have a record of the temporal variation of the envelope of each filter's response. Neither time variation of the envelope nor place (filter center frequency) information can be discarded, if there is to be sufficient information for signal reconstruction.

2.2.10 Summary

Properties and interrelationships of various waveform descriptions, i.e., the complex analytic signal representation, the spectrogram, the time-frequency energy density function, and the ambiguity function, have been assembled in Appendix A. By combining several of these properties, we have found that a signal can be reconstructed from its spectrogram, except for a complex factor with unit magnitude, $\exp(j\lambda)$. The time invariant (or frequency invariant) part of a signal's phase function is then the only information that is lost when the signal is converted into a spectrogram. Variable bandwidth filters can be taken into account by a nonlinear transformation of the frequency scale before a spectrogram is formed.

Practical reconstruction of a signal from its spectrogram is affected by sampling rates and by additive noise. The spectrogram must usually be sampled at a much higher rate for signal reconstruction than for an efficient representation of the spectrogram itself. The reconstruction method can be modified by using a two-dimensional filter that depends upon internal and external noise, rather than the deconvolution operation that is used for noise-free data.

It is obvious that a nonideal detection process can be performed by comparing a data spectrogram with the spectrogram of a known signal. For low signal-to-noise ratio, the best way to implement this comparison is by correlation of data and reference spectrograms. In order for such a spectrogram correlator to perform as well as a matched filter, the energy of the input signal must be increased by about 5 dB.

A matched filter can be applied to signals that have been reconstructed from a spectrogram. An equivalent process is to integrate the product of the data autoambiguity function and the conjugated autoambiguity function of the reference signal. Complete reconstruction of the input data is therefore unnecessary once the data ambiguity function has been estimated.

When prior statistical information about a random signal is available, an estimator-correlator is often superior to an ordinary energy detector. Signals that are estimated from noise-corrupted spectrograms are especially useful for an estimator-correlator configuration if the prior information is given as the ensemble average of the envelope of the signal's autoambiguity function. Prior information of this type is available when a signal has been passed through a time-varying random filter with known time-frequency autocorrelation function.

From the viewpoint of optimum array processing, a spectrogram representation is inadequate, since it does not provide an efficient direction estimate if the signal has narrow bandwidth. A complete description of the input signal is necessary in this case. Such a description can be obtained by combining the spectrogram with a phase measurement. The phase measurement is obtained from the pre-detected response of one of the filters that is used to construct the spectrogram.

The above theoretical concepts have been compared with some facts and theories about hearing and animal echolocation:

The sampling rates for reconstruction of an input signal require overlapping neural or hair cell tuning curves, if these tuned neurons are analogous to the filters and envelope detectors that are used for construction of a spectrogram. A fixed sampling rate in the frequency direction also results in a maximum signal duration beyond which the input signal cannot be reconstructed and coherent processing cannot be applied. The predicted overlap of neural tuning curves and a change in listener performance for detection of signals longer than 200 msec have both been observed in mammalian audition.

Another interesting prediction is that, for negligible internal noise, tuning curves with narrow bandwidths or sharp, high frequency cutoffs are not necessary for accurate frequency discrimination. This prediction is relevant because tuning curves with sharp cutoffs are not always experimentally observed.

For masking of a narrowband signal with narrowband noise, both spectrogram correlation and signal reconstruction operations are associated with critical band effects. For masking of a gated sinusoid by wideband noise, the reconstruction/matched filter

hypothesis predicts an increase in effective filter bandwidth when signal duration is decreased, and this bandwidth increase has been experimentally observed (although the observed bandwidths are too wide).

A natural consequence of the reconstruction hypothesis is monaural insensitivity to the phase of a complex, analytic waveform. Binaural processing in humans, however, is phase sensitive. This phenomenon can only be explained by assuming that at least one predetected filter output is available. This assumption is justified for low frequency neural filters, since phase-locked discharges have been observed below 4 to 6 kHz. A broadband signal with low frequency components or a low frequency narrowband signal can thus be completely reconstructed for binaural processing.

A spectrogram representation implies that sinusoids with different envelopes should be distinguishable. This envelope sensitivity has been experimentally observed.

Another consequence of signal reconstruction is the possibility that matched filtering, rather than energy detection, can be used by a listener in a two-alternative, forced choice between signal plus noise and noise alone. In fact, the experimental results for 2 AFC detection of a single tone seem to indicate an imperfect matched filter model, or perhaps an estimator-correlator model. Human detection performance for a sum of sinusoids with different frequencies, however, seems to rule out a matched filter and instead favors a spectrogram correlation or spectrogram estimator-correlator process.

In animal echolocation, a spectrogram-like signal representation seems to imply that detection is best modelled by a spectrogram

correlator rather than by an ideal detector. We have demonstrated, however, that matched filtering is still a possibility. A 5 dB SNR improvement (or an increase in maximum range by a factor of 1.3) is obtained if the ideal processor is implemented and the signal is known exactly.

One of the most exciting applications of the analysis is a new concept for the design of hearing aids. A hearing aid would operate on the spectrogram of an input signal, and the altered spectrogram would be transformed back into a new acoustic signal for the benefit of a listener.

2.2.11 Conclusion

In designing an efficient speech encoder/synthesizer, a hearing aid, or a sonar that is built to simulate animal echolocation, it is important to know what information is destroyed by the peripheral auditory system. This information can then be neglected in the corresponding man-made systems and can lead to a significant reduction in complexity.

Much of the neurophysiological and psychoacoustic data indicate that the peripheral auditory system can be represented by a spectrogram synthesizer with some nonlinear filters and probabilistic components. Given a linear, deterministic version of this model, we have shown that the central nervous system probably has sufficient information to reconstruct the input signal, except for a multiplicative constant with unit magnitude and unknown phase.

Unfortunately, this result does little to simplify the design of man-made devices, since it implies that very little information may actually be destroyed by peripheral sound transduction and

neural encoding in mammals. In a simulation of an animal sonar, for example, a matched filter may still be relevant.

The results of the analysis, however, can be viewed in a more optimistic light. If a spectrogram is indeed analogous to the representation of a sound that is sent to the brain, then signal processing operations that are performed on the spectrogram should be directly related to sound perception. Although the concept of operating on a spectrogram is not new, we have found a precisely defined method for converting an altered spectrogram back into an acoustic signal that can be presented to a listener. This signal reconstruction method may lead to better hearing aid design. It may also lead to a new method for auditory experimentation, where signals and maskers are synthesized as spectrograms, i.e., in terms of their probable representations in the central nervous system. Such a method should provide further insight into the detection and recognition of complex acoustic signals by the brain.

2.2.12 References for Section 2.2

Abramowitz, M., and Stegun, I. A. (1964). Handbook of Mathematical Functions, NBS AMS 55, Washington, D. C., U. S. Gov't. Printing Office, 375.

Ackroyd, M. H. (1971). "Short-time spectra and time-frequency energy distributions," J. Acous. Soc. Am. 50, 1229-1231.

Airapetyants, E. Sh., and Konstantinov, A. I. (1970). Echolocation in Animals. Engl. Trans. (1973), Jerusalem, Keter Press.

Altes, R. A. (1978). "Angle estimation and binaural processing in animal echolocation," J. Acous. Soc. Amer. 63, 155-173.

Altes, R. A. (1975). "Wide-band systems and Gaussianity," IEEE Trans. Inform. Theory IT-21, 679-682.

Altes, R. A., and Skinner, D. P. (1977). "Sonar velocity resolution with a linear-period-modulated pulse," J. Acous. Soc. Am. 61, 1019-1030.

Altes, R. A., and Titlebaum, E. L. (1970). "Bat signals as optimally doppler tolerant waveforms," J. Acous. Soc. Amer. 48, 1014-1020.

Altes, R. A., and Titlebaum, E. L. (1975). "Graphical derivations of signals for radar, sonar, and communication," IEEE Trans. Aerospace and Electron. Sys. AES-11, 38-44.

Anderson, D. J. (1971). "A quantification for the desynchronization of auditory nerve discharges with increasing stimulus frequency," J. Acous. Soc. Amer. 49, 113A.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (1971). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine wave stimulus: frequency and intensity effects," J. Acous. Soc. Amer. 49, 1131-1139.

Bello, P. (1963). "Characterization of randomly time-varying linear channels," IEEE Trans. on Comm. Sys. CS-11, 360-393.

de Boer, E. (1975). "Synthetic whole-nerve action potentials for the cat," J. Acous. Soc. Amer. 58, 1030-1045.

de Boer, E., and de Jongh, H. R. (1978). "On cochlear encoding: potentialities and limitations of the reverse-correlation technique," J. Acous. Soc. Am. 63, 115-135.

Bullock, T. H., Ed. (1977). Recognition of Complex Acoustic Signals. Berlin: Abakon Verlagsgesellschaft.

Cahlender, D. A. (1967). "Echolocation with wide-band waveforms," in Animal Sonar Systems, ed. by R. G. Busnel, Jouy-en-Josas, France, Lab. de Phys. Acous., Vol. II, 1052-1081.

Capon, J. (1961). "On the asymptotic efficiency of locally optimum detectors," IRE Trans. on Inform. Theory IT-7, 67-71.

Cook, C. E., and Bernfeld, M. (1967). Radar Signals. New York: Academic, 160-164, 88.

Craig, J. H., and Jeffress, L. A. (1962). "Effects of phase on the quality of a two-component tone," J. Acous. Soc. Am. 34, 1752-1760.

Daly, R. F. (1970). "Signal design for efficient detection in dispersive channels," IEEE Trans. on Inform. Theory IT-16, 206-213.

Davenport, W. B., and Root, W. L. (1958). Random Signals and Noise. New York: McGraw-Hill, 157.

Erdélyi, A. (1954). Tables of Integral Transforms, Vol. II. New York: McGraw-Hill, 243.

Evans, E. F. (1977). "Peripheral processing of complex sounds," in Recognition of Complex Acoustic Signals, T. H. Bullock ed. Berlin: Abakon Verlagsgesellschaft, 145-159.

Evans, E. F. (1975). "Cochlear nerve and cochlear nucleus," in Handbook of Sensory Physiology, vol. 5, part II (W. D. Keidel and W. D. Neff, eds.). Berlin: Springer-Verlag, pp. 1-108.

Evans, W. E. (1973). "Echolocation by marine delphinids and one species of fresh-water dolphin," J. Acous. Soc. Amer. 54, 191-199.

Flanagan, J. L. (1965). Speech Analysis, Synthesis, and Perception. Berlin: Springer-Verlag, 140-163, 323-327.

Flaska, M. D. (1976). "Cross correlation of short-time spectral histories," J. Acous. Soc. Amer. 59, 381-388.

Fletcher, H. (1953). Speech and Hearing in Communication. Princeton, Von Nostrand, 173-174.

Franks, L. E. (1969). Signal Theory. Englewood Cliffs: Prentice-Hall, pp. 73-80

Gabor, D. (1946). "Theory of Communication," JIEE (London) 93, 429-457

Green, D. M. (1958). "Detection of multiple component signals in noise," J. Acous. Soc. Am. 30, 904-911.

Green, D. M., Birdsall, T. G., and Tanner, W. P. (1957). "Signal detection as a function of signal intensity and duration," J. Acous. Soc. Amer. 29, 523-531.

Green, D. M., McKey, M. J., and Licklider, J. C. R. (1959). "Detection of a pulsed sinusoid in noise as a function of frequency," J. Acous. Soc. Am. 31, 1446-1452.

Green, D. M., and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. New York, Wiley, 209-232.

Griffin, D. R. (1958). Listening in the Dark. New Haven, Yale Univ. Press.

Hagan, J. B., and Farley, D. T. (1973). "Digital-correlation techniques in radio science," Radio Science 8, 775-784.

Hahn, W. R., and Tretter, S. A. (1973). "Optimum processing for delay-vector estimation in passive signal arrays," IEEE Trans. Inform. Theory IT-19, 608-614.

Hahn, W.R. (1975). "Optimum signal processing for passive sonar range and bearing estimation," J. Acous. Soc. Am. 58, 201-207.

Hartmann, W. M. (1978). "The effect of amplitude envelope on the pitch of sine wave tones," J. Acous. Soc. Am. 63, 1105-1113.

Johnson, R. A., and Titlebaum, E. L. (1976). "Energy spectrum analysis: A model of echolocation processing," J. Acous. Soc. Am. 60, 484-491.

Kailath, T. (1960). "Correlation detection of signals perturbed by a random channel," IRE Trans. on Inform. Theory, IT-6, 361-366.

Kailath, T. (1969). "A general likelihood-ratio formula for random signals in Gaussian noise," IEEE Trans. Inform. Theory IT-15, 350-361.

Kailath, T. (1970). "The innovations approach to detection and estimation theory," Proc. IEEE 58, 680-695.

Kim, D. O., Molnar, C. E., and Pfeiffer, R. R. (1973). "A system of nonlinear differential equations modelling basilar-membrane motion," J. Acous. Soc. Amer. 54, 1517-1529.

Kroszczynski, J. J. (1969). "Pulse compression by means of linear-period modulation," Proc. IEEE 57, 1260-1266.

McGill, W. J. (1968). "Variations on Marrill's detection formula," J. Acous.Soc. Am. 43, 70-73.

Middleton, D. (1966). "Canonically optimum threshold detection," IEEE Trans. Inform. Theory IT-12, 230-243.

Papoulis, A. (1965). Probability, Random Variables, and Stochastic Processes. New York, McGraw-Hill, 129-196.

Papoulis, A. (1968). Systems and Transforms with Applications in Optics. New York, McGraw-Hill, p. 174.

Papoulis, A. (1972a). "Approximations of point spreads for deconvolution," J. Opt. Soc. Am. 62, 77-80.

Papoulis, A. (1972b). "Narrowband systems and Gaussianity," IEEE Trans. Inform. Theory IT-18, 20-27.

Pfeiffer, R. R., and Kim, D. O. (1975). "Cochlear nerve fiber responses: Distribution along the cochlear partition," J. Acous. Soc. Amer. 58, 867-869.

Poor, H. V., and Thomas, J. B. (1978). "Locally optimum detection of discrete-time stochastic signals in non-Gaussian noise," J. Acous. Soc. Am. 63, 75-80.

Price, R. (1956). "Optimum detection of random signals in noise, with application to scatter-multipath communication," IRE Trans. Inform. Theory. IT-2, pp. 125-135.

Rice, S. O. (1954). "Mathematical analysis of random noise," in Noise and Stochastic Processes, N. Wax, Ed. NewYork, Dover.

Rihaczek, A. W. (1968). "Signal energy distribution in time and frequency," IEEE Trans. on Inform. Theory IT-14, 369-374.

Richaczek, A. W. (2969). Principles of High-Resolution Radar. New York: McGraw-Hill, 10-40.

Sayers, B. McA., and Cherry, E. C. (1957). "Mechanism of binaural fusion in the hearing of speech," J. Acous. Soc. Am. 29, 973-987.

Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O. (1950). "The frequency selectivity of the ear as determined by masking experiments," J. Acous. Soc. Amer. 52, 490-496.

Scharf, L. L., and Nolte, L. W. (1977). "Likelihood ratios for sequential hypothesis testing on Markov sequences," IEEE Trans. Inform. Theory IT-23, 101-209.

Schroeder, M R. (1975). "Models of Hearing," Proc. IEEE 63, 1332-1350.

Shnidman, D. A. (1975). "Solution to a class of optimization problems with amplitude constraints," IEEE Trans. on Comm. COM-23, 979-983.

Shower, E. G., and Biddulph, R. (1931). "Differential pitch sensitivity of the ear," J. Acous. Soc. Am. 3, 275-287.

Siebert, W. M. (1958). "Studies of Woodward's uncertainty function," Quar. Progress Rept., MIT, 90-94.

Siebert, W. M. (1968). "Stimulus transformations in the peripheral auditory system," in Recognizing Patterns, P. A. Kolers and M. Eden, eds. Cambridge, MA., MIT Press, 104-133.

Siebert, W. M. (1970). "Frequency discrimination in the auditory system: place or periodicity mechanisms?" Proc. IEEE 58, 723-730.

Simmons, J. A. (1973). "The resolution of target range by echo-locating bats," J. Acous. Soc. Amer. 54, 157-173.

Slepian, D. (1954). "Estimation of signal parameters in the presence of noise," IRE Trans. Inform. Theory IT-1, 68-89.

Srinivasan, R. (1971). "Auditory critical bandwidth for short-duration signals," J. Acous. Soc. Am. 50, 616-622.

Stutt, C. A. (1964). "Some results on real-part/imaginary-part and magnitude-phase relations in ambiguity functions," IEEE Trans. Inform. Theory IT-10, 321-327.

Taylor, M. M., and Forbes, S. M. (1969). "Monaural detection with contralateral cue (MDCC). I. Better than energy detector performance by human observers," J. Acous. Soc. Am. 46, 1519-1526.

Terhardt, E., and Fastl, H. (1971). "Zum einfluss von störtönen und störgeräuschen auf die tonhöhe von sinustönen," Acustica, 25, 53-61.

Titlebaum, E. L., and DeClaris, N. (1966). "Linear transformations of the ambiguity function," IEEE Trans. Inform. Theory IT-12, 120-125.

Urkowitz, H. (1967). "Energy detection of unknown deterministic signals," Proc. IEEE 55, 523-531.

Van Trees, H. L. (1968). Detection, estimation, and modulation theory, part I. New York, Wiley, pp. 19-86.

Venetsanopoulos, A. N. (1978). "Modelling of the sea surface scattering channel and undersea communications," in Communication Systems and Random Process Theory, J. K. Skwirzynski, ed. Netherlands: Sijthoff and Noordhoff, 511-531.

Woodward, P. M. (1964). Probability and Information Theory with Applications to Radar. Oxford: Pergamon, 115-125.

Zwicker, E. (1961). "Subdivisions of the audible frequency range into critical bands," J. Acous. Soc. Amer. 33, 248.

2.2.13 Appendices for Section 2.2

APPENDIX A: PROPERTIES OF SPECTROGRAMS AND RELATED FUNCTIONS

Some properties of time-frequency energy density functions, spectrograms, and ambiguity functions are listed below. Where limits on integrals are omitted, the limits are $(-\infty, \infty)$.

Property 1 (Ackroyd, 1971).

$$S_{uv}(t_1, f_1) = e_{uu}(t, f) * e_{vv}(t, -f) \tag{A1}$$

where $(*)$ denotes two-dimensional convolution, i.e.,

$$S_{uv}(t_1, f_1) = \iint e_{uu}(t, f)\, e_{vv}(t_1 - t, f - f_1)\, dt\, df .$$

Property 2 (Ackroyd, 1971).

$$\left| X_{uv}(-\tau, \phi) \right|^2 = e_{uu}(t, f) * e^*_{vv}(-t, -f) \tag{A2}$$

$$= \iint e_{uu}(t, f)\, e^*_{vv}(t+\tau,\ f-\phi)\, dt\, df .$$

Property 3 (Stutt, 1964; Rihaczeck, 1968).

$$F\left\{ e_{uv}(t, f);\ \tau, \phi \right\} = X_{uv}(\tau, \phi) \tag{A3}$$

where $F\left\{ e_{uv}(t, f);\ \tau, \phi \right\}$ denotes a two-dimensional Fourier transform that maps $e_{uv}(t, f)$ onto the τ, ϕ plane, i.e.,

$$\iint e_{uv}(t, f)\, e^{-j2\pi(\phi t + \tau f)}\, dt\, df = X_{uv}(\tau, \phi) .$$

Property 4.

From (6) – (9),

$$\left| X_{uv}(t_1, f_1) \right|^2 = S_{u(t)v^*(-t)}(-t_1, f_1) \tag{A4}$$

$$S_{uv}(t_1, f_1) = \left| X_{u(t)v^*(-t)}(-t_1, f_1) \right|^2 . \tag{A5}$$

Since many properties of the magnitude-squared ambiguity function have been derived (Stutt, 1964; Cook and Bernfeld, 1967), the above identity leads to similar properties of the spectrogram. Properties 5 to 7, for example, are shared by $S_{uv}(t, f)$ and $\left| X_{uv}(t, f) \right|^2$.

Property 5.

Smeared energy density spectra and time envelopes can be obtained from marginals of the spectrogram, i.e.,

$$\int S_{uv}(t_1, f_1)\, dt_1 = \int \left| U(f) \right|^2 \left| V(f - f_1) \right|^2 df \tag{A6}$$

$$\int S_{uv}(t_1, f_1)\, df_1 = \int \left| u(t) \right|^2 \left| v(t_1 - t) \right|^2 dt . \tag{A7}$$

These relations are verified by substitution of (8) into the left-hand side of (A6), and by substituting (9) into the left-hand side of (A7).

Property 6.

$$\iint S_{uv}(t_1, f_1)\, dt_1\, df_1 = E_u E_v \tag{A8}$$

where

$$E_u \equiv \int \left|U(f)\right|^2 df = \int \left|u(t)\right|^2 dt$$

$$E_v \equiv \int \left|V(f)\right|^2 df = \int \left|v(t)\right|^2 dt \,. \qquad \text{(A9)}$$

This property states that the volume under the spectrogram depends strictly upon the energy of the signal and filter functions. The volume is independent of the form of these functions. To verify (A8), integrate (A6) with respect to f_1, or integrate (A7) with respect to t_1.

Property 7.

$$S_{uv}(t_1, f_1) \leq E_u E_v \qquad \text{(A10)}$$

with equality if and only if there exist values of t_1 and f_1 such that

$$V(f-f_1) \exp(j2\pi f t_1) = U^*(f) \,. \qquad \text{(A11)}$$

The magnitude of the spectrogram never exceeds the product $E_u E_v$, and the upper bound is only obtained if one of the filters that is used to form the spectrogram is matched to the signal. Eq. (A10) is obtained by applying the Schwarz inequality to $S_{uv}(t_1, f_1)$, as defined by (8).

Property 8.

$$F\{S_{uv}(t,f);\ \tau,\phi\} = X_{uu}(\tau,\phi)\, X_{vv}(-\tau,\phi) \qquad \text{(A12)}$$

or

$$X_{uu}(\tau,\phi) = F\{S_{uv}(t,f);\ \tau,\phi\} / X_{vv}(-\tau,\phi)\ . \qquad \text{(A13)}$$

Eqs. (A12) and (A13) follow directly from (A1) and (A3). Eq. (A13) implies that, in the absence of noise, $X_{uu}(\tau,\phi)$ can be obtained from a spectrogram, if the filter function v(t) is known. The significance of this result is enhanced by the following property and by Property 14.

Property 9 (Siebert, 1958; Titlebaum and DeClaris, 1966).

$X_{u_1u_1}(\tau,\phi) = X_{u_2u_2}(\tau,\phi)$ if and only if

$u_1(t) = u_2(t) \exp(j\lambda)$, where λ is real.

This observation was originally derived from the separation equation

$$\int X_{uu}(\tau-t,\phi) \exp(j2\pi t\phi)\, d\phi = u(t)\, u^*(\tau)\ . \qquad \text{(A14)}$$

The result can also be derived from Property 3 by taking $F\{X_{uu}(\tau,\phi);\ t,f\} = e_{uu}(t,f)$, and by forming

$$\int e_{uu}(t,f) \exp[j2\pi f(t-\tau)]\, df = u(t)\, u^*(\tau)\ . \qquad \text{(A15)}$$

In either case, $\left|u(\tau_o)\right|$ is obtained by setting $t = \tau = \tau_o$, where τ_o is a constant. Then $u(t)\left[u^*(\tau_o)/\left|u(\tau_o)\right|\right]$ is the desired waveform u(t), multiplied by a complex constant with unity magnitude.

It follows that one can obtain u(t) exp (jλ) from $X_{uu}(\tau, \phi)$ or $e_{uu}(t, f)$, where λ is a real, unknown constant.

Property 10 (Stutt, 1964).

The counterpart of Property 8 for the magnitude-squared ambiguity function is

$$F\left\{\left|X_{uv}(-\tau, \phi)\right|^2;\ t, f\right\} = X_{uu}(t, f)\, X^*_{vv}(t, f) \qquad \text{(A16)}$$

i.e.,

$$\iint \left|X_{uv}(-\tau, \phi)\right|^2 e^{-j2\pi(f\tau + t\phi)}\, d\tau d\phi = X_{uu}(t, f)\, X^*_{vv}(t, f)\ .$$

In particular,

$$F\left\{\left|X_{uu}(-\tau, \phi)\right|^2;\ t, f\right\} = \left|X_{uu}(t, f)\right|^2\ ,$$

i.e., the magnitude-squared autoambiguity function is its own Fourier transform. Eq. (A2) can be obtained directly from (A3) and (A16) by using the fact that multiplication of two functions is the same as convolution of their Fourier transforms.

Property 11.

Multiplying both sides of (A12) by $X^*_{vv}(-\tau, \phi)$, we have

$$F\left\{S_{uv}(t,f);\ \tau,\phi\right\} X_{vv}^*(-\tau,\phi) = X_{uu}(\tau,\phi)\left|X_{vv}(-\tau,\phi)\right|^2 .$$

Using Property 10 and (A3),

$$S_{uv}(t,f) * e_{vv}^*(-t,f) = e_{uu}(t,f) * \left|X_{vv}(t,f)\right|^2 . \qquad (A17)$$

The above relation can also be obtained from (A1) and (A2), and this alternate derivation leads to a generalized version of (A17):

$$S_{u_1v_1}(t,f) * e^*_{v_2v_2}(-t,f) = e_{u_1u_1}(t,f) * \left|X_{v_1v_2}(t,f)\right|^2$$

or

$$\iint S_{u_1v_1}(t,f)\, e^*_{v_2v_2}(t-\tau,\phi-f)\, dt\, df =$$

$$\iint e_{u_1u_1}(t,f)\left|X_{v_1v_2}(\tau-t,\phi-f)\right|^2 dt\, df . \qquad (A18)$$

Property 12.

Two different short-time spectral histories can be compared by means of the cross-correlation operation

$$\iint S_{u_1v_1}(t,f)\, S_{u_2v_2}(t+\tau,f+\phi)\, dt\, df = S_{u_1v_1}{}^*(-t,-f) * S_{u_2v_2}(t,f)$$

$$= e_{u_1u_1}{}^*(-t,-f) * e_{v_1v_1}{}^*(-t,f) * e_{u_2u_2}(t,f) * e_{v_2v_2}(t,-f)$$

$$= \left| X_{u_2 u_1}(-t, f) \right|^2 * \left| X_{v_2 v_1}(t, -f) \right|^2$$

$$= \iint \left| X_{u_2 u_1}(t, f) \right|^2 \left| X_{v_2 v_1}(\tau + t, f - \phi) \right|^2 dt\, df \,. \qquad \text{(A19)}$$

The above sequence of equalities made use of (A1) and (A2). Equation (A19) is similar to a result for wideband auto-ambiguity functions that was published by Flaska (1976), with credit to C. E. Persons.

Property 13.

Another way to compare two different signals is to integrate the product of their time-frequency energy density functions. By using (6) and Parseval's theorem, it is easy to prove that

$$\iint e_{u_1 v_1}(t, f)\, e^*_{u_2 v_2}(t, f)\, dt\, df$$

$$= \int u_1(t)\, u_2^*(t)\, dt \int v_1^*(t)\, v_2(t)\, dt \,. \qquad \text{(A20)}$$

An important special case of (A20) is obtained by letting $v_1(t) = u_1(t)$ and $v_2(t) = u_2(t) = u_1(t-\tau) \exp(j2\pi\phi t)$. In this case, (A20) becomes

$$\iint e_{u_1 u_1}(t, f)\, e^*_{u_2 u_2}(t, f)\, dt\, df = \left| X_{u_1 u_1}(\tau, \phi) \right|^2 \qquad \text{(A21)}$$

Property 14.

By using Property 3 and applying Parseval's theorem, (A20) can be written

$$\iint X_{u_1 v_1}(\tau, \phi)\; X^*_{u_2 v_2}(\tau, \phi)\, d\tau\, d\phi$$

$$= \int u_1(t)\, u_2^*(t)\, dt \int v_1^*(t)\, v_2(t)\, dt\,. \qquad \text{(A22)}$$

The above equation can also be verified by direct substitution of (7) into the left-hand side. Once again, if $v_1(t) = u_1(t)$ and $v_2(t) = u_2(t) = u_1(t-\tau_1)\exp(j2\pi\phi_1 t)$, we have

$$\iint X_{u_1 u_1}(\tau, \phi)\, X^*_{u_2 u_2}(\tau, \phi)\, d\tau\, d\phi = \left| X_{u_1 u_1}(\tau_1, \phi_1) \right|^2 .$$

Similarly, if $v_1(t) = u_1(t)$ and $v_2(t) = u_2(t)$ in (A22), we have

$$\iint X_{u_1 u_1}(\tau, \phi)\, X^*_{u_2 u_2}(\tau, \phi)\, d\tau\, d\phi = \left| \int u_1(t) u_2^*(t)\, dt \right|^2 . \qquad \text{(A23)}$$

The significance of (A23) is that two signals can be correlated by operating upon their autoambiguity functions, rather than upon the signals themselves. Eq. (A23) can also be obtained from Stutt's Fourier transform equation, (A16).

Property 15.

Because the time-window version of the spectrogram is commonly used in spectrogram representations (Flanagan, 1965) and in

auditory theory (Hartmann, 1978), it is useful to relate the time-window spectrogram $S^T_{uv}(t_1, f_1)$ to the frequency window spectrogram that is defined in (8) and (9). The time-window spectrogram is obtained by multiplying the data u(t) by a delayed window function $w(t-t_1)$ and by taking the Fourier transform of the resulting product, i.e.,

$$S^T_{uw}(t_1, f_1) = \left| \int u(t)\, w(t-t_1) \exp(-j2\pi f_1 t)\, dt \right|^2 . \qquad \text{(A24)}$$

By changing variables in (9), the frequency-window (bank-of-filters) spectrogram can be written

$$S_{uv}(t_1, f_1) = \left| \int_{-\infty}^{\infty} u(t)\, v(t_1-t) \exp(-j2\pi f_1 t)\, dt \right|^2 . \qquad \text{(A25)}$$

Comparison of (A24) and (A25) indicates that

$$S^T_{uw}(t_1, f_1) = S_{uv}(t_1, f_1) \quad \text{if}$$

$$w(t) = v(-t) , \qquad \text{(A26)}$$

where v(t) is the impulse response of the basic filter function that is used to construct the frequency-window spectrogram.

APPENDIX B: ESTIMATION FROM A NOISY SPECTROGRAM

The problem is to derive a filter for estimating a signal's ambiguity function from a noise-corrupted spectrogram.

From Figure 1, we have an estimate

$$\hat{X}_{uu}(\tau, \phi) = H(\tau, \phi)\left\{\left[X_{uu}(\tau, \phi) + N_e(\tau, \phi)\right]X_{vv}(-\tau, \phi) + N_i(\tau, \phi)\right\}. \tag{B1}$$

The mean-square error that is associated with this estimate is

$$MSE(\epsilon) = E\left\{\iint\left|X_{uu} - H\left[(X_{uu}+N_e)\,X_{vv}+N_i\right]\right|^2 d\tau\, d\phi\right\} \tag{B2}$$

where

$$H(\tau, \phi) = H_0(\tau, \phi) + \epsilon\ \eta(\tau, \phi) \tag{B3}$$

and $\eta(\tau, \phi)$ is an arbitrary, piecewise-smooth function.

Define

$$B(\tau, \phi) \equiv \left[X_{uu}(\tau, \phi) + N_e(\tau, \phi)\right]\ X_{vv}(-\tau, \phi) + N_i(\tau, \phi), \tag{B4}$$

so that

$$MSE(\epsilon) = E\left\{\iint\left|X_{uu} - BH\right|^2 d\tau d\phi\right\}. \tag{B5}$$

In (B2) and (B5), the expectation operator is an ensemble average with respect to N_e, N_i and all admissible input signals, u(t).

Differentiating MSE(ϵ) with respect to ϵ in (B5) yields

$$\partial MSE(\epsilon)/\partial\epsilon \Big|_{\epsilon=0} = 2\,Re \iint \eta^* \left[H_0 E\left\{ \left| B \right|^2 \right\} - E\left\{ X_{uu} B^* \right\} \right] d\tau\, d\phi \tag{B6}$$

and

$$\partial MSE(\epsilon)/\partial\epsilon \Big|_{\epsilon=0} = 0 \tag{B7}$$

for any $\eta(\tau, \phi)$ if

$$H_0 = E\left\{ X_{uu} B^* \right\} / E\left\{ \left| B \right|^2 \right\}$$

$$= \frac{\overline{\left| X_{uu}(\tau, \phi) \right|^2}\, X_{vv}^*(-\tau, \phi)}{\left| X_{vv}(-\tau, \phi) \right|^2 \left[\overline{\left| X_{uu}(\tau, \phi) \right|^2} + P_e(\tau, \phi) \right] + P_i(\tau, \phi)} \tag{B8}$$

where it has been assumed that

$$E\left\{ N_e(\tau, \phi) \right\} = 0, \qquad E\left\{ \left| N_e(\tau, \phi) \right|^2 \right\} = P_e(\tau, \phi)$$

$$E\left\{ N_i(\tau, \phi) \right\} = 0, \qquad E\left\{ \left| N_i(\tau, \phi) \right|^2 \right\} = P_i(\tau, \phi)$$

$$E\left\{ N_e^*(\tau, \phi)\, N_i(\tau, \phi) \right\} = 0$$

$$E\left\{ \left| X_{uu}(\tau, \phi) \right|^2 \right\} = \overline{\left| X_{uu}(\tau, \phi) \right|^2} .$$

Substituting (B8) into (B5), we have

$$MSE_0 = E\left\{\iint\left[\left|X_{uu}\right|^2 + \left|BH_0\right|^2 - 2Re\left\{X_{uu}^{*}\, BH_0\right\}\right] d\tau\, d\phi\right\}$$

$$= \iint\left[\overline{\left|X_{uu}\right|^2} + \left|H_0\right|^2 E\left\{\left|B\right|^2\right\} - 2Re\left\{H_0 E\left\{X_{uu}^{*} B\right\}\right\}\right] d\tau\, d\phi \qquad (B9)$$

where

$$\left|H_0\right|^2 E\left\{\left|B\right|^2\right\} = \left|E\left\{X_{uu}^{*} B\right\}\right|^2 / E\left\{\left|B\right|^2\right\}$$

$$E\left\{X_{uu}^{*} B\right\} = \overline{\left|X_{uu}\right|^2}\, X_{vv}$$

$$H_0 E\left\{X_{uu}^{*} B\right\} = \overline{\left|X_{uu}\right|^4}\, \left|X_{vv}\right|^2 / E\left\{\left|B\right|^2\right\} .$$

Therefore,

$$MSE_0 = \iint \overline{\left|X_{uu}\right|^2}\left[1 - \overline{\left|X_{uu}\right|^2}\, \left|X_{vv}\right|^2 / E\left\{\left|B\right|^2\right\}\right] d\tau\, d\phi$$

$$= \iint \overline{\left|X_{uu}\right|^2}\left[1 - \frac{\overline{\left|X_{uu}\right|^2}}{\overline{\left|X_{uu}\right|^2} + P_e + \left(P_i / \left|X_{vv}\right|^2\right)}\right] d\tau\, d\phi . \qquad (B10)$$

For given noise distributions $P_e(\tau, \phi)$ and $P_i(\tau, \phi)$, what is the best analyzing filter V(f)? If $P_i(\tau, \phi) = 0$, MSE_0 is independent of V(f), but if $P_i(\tau, \phi) \neq 0$, we can find $|X_{vv}(-\tau, \phi)|^2$ such that MSE_0 is minimized. In order to obtain a solution to this problem, we add a constraint that the volume of $|X_{vv}(-\tau, \phi)|^2$ equal E_v^2 as in (20). Then

$$MSE_0 = \iint \overline{|X_{uu}|^2} \left[1 - \frac{\overline{|X_{uu}|^2}}{\overline{|X_{uu}|^2} + P_e + \left(P_i / |X_{vv}|^2 \right)} \right] d\tau \, d\phi$$

$$+ \lambda_v \left[\iint |X_{vv}|^2 \, d\tau \, d\phi - E_v^2 \right] .$$

The Euler condition is

$$\frac{\partial}{\partial |X_{vv}|^2} \left\{ - \left[\overline{|X_{uu}|^2} \right]^2 \left[\overline{|X_{uu}|^2} + P_e + \left(P_i / |X_{vv}|^2 \right) \right]^{-1} - \lambda_v |X_{vv}|^2 \right\} = 0$$

or

$$|X_{vv}|^2 = \frac{\sqrt{P_i / \lambda_v} \left[|X_{uu}|^2 - \sqrt{\lambda_v P_i} \right]}{\overline{|X_{uu}|^2} + P_e} .$$

Since $|X_{vv}|^2 \geq 0$, we have (Shnidman, 1975)

$$|X_{vv}|^2 = \sqrt{P_i/\lambda_v}\ \max\left[\frac{\overline{|X_{uu}|^2} - \sqrt{\lambda_v P_i}}{\overline{|X_{uu}|^2} + P_e},\ 0\right]. \qquad \text{(B11)}$$

This solution represents a minimum value of MSE_0, since the second variation of the functional with respect to $|X_{vv}|^2$ is positive if $P_i > 0$.

APPENDIX C: DETECTION BY CORRELATION OF SPECTROGRAMS

1. Detector Configuration

In order to derive a detector configuration, we must first obtain the probability density function (pdf) of a noisy spectrogram. The data is assumed to consist of signal plus white Gaussian noise with power spectral density $N_o/2$. The data is first passed through a bank of linear filters with transfer functions $V(f - f_j)$, $j = 1, 2, \cdots, M$. The real part of the jth filter output is Gaussian (Davenport and Root, 1958) with power spectral density $(N_o/2) \left| V(f - f_j) \right|^2$. The variance of this process is

$$\sigma^2 = (N_o/2) E_v . \tag{C1}$$

where E_v was defined in (A9).

For Gaussian input noise, the real and imaginary parts of each filter output are statistically independent and Gaussian. This statement follows from the following three observations:

(i) Linear filtering preserves Gaussianity.

(ii) Uncorrelated Gaussian random variables are statistically independent.

(iii) For any random process $x(t)$ that has a Hilbert transform $\hat{x}(t)$,

$$E\left\{x(t)\,\hat{x}(t)\right\} = 0 , \tag{C2}$$

i.e., $x(t)$ and $\hat{x}(t)$ are uncorrelated (Papoulis, 1968).

For non-Gaussian input noise, the above result will still hold if we can modify observation (i) to read "linear filtering causes non-Gaussian signals to become Gaussian." This modified statement is approximately true of narrowband filters (Papoulis, 1972b) and wide-band filters with large time-bandwidth product (Altes, 1975).

Let x_{ij} denote the response $x(t)$ of the jth filter at time i, and let y_{ij} be a sample at time i of the Hilbert transform of the jth filter response. The envelope of the filter response is formed as in (3), and we are interested in the pdf of the squared envelope, which is a sample of the spectrogram:

$$Z(t_i, f_j) = z_{ij} = (x_{ij}/\sqrt{2})^2 + (y_{ij}/\sqrt{2})^2. \tag{C3}$$

The random variables $x = x_{ij}/\sqrt{2}$ and $y = y_{ij}/\sqrt{2}$ are Gaussian and independent, with joint pdf

$$p_{xy}(x, y) = (\pi\sigma^2)^{-1} \exp\left\{-\left[(x - \eta_1)^2 + (y - \eta_2)^2\right]/\sigma^2\right\} \tag{C4}$$

where the noise-free spectrogram sample is

$$s_{ij} = \eta_1^2 + \eta_2^2 \ . \tag{C5}$$

To find the pdf of $Z(t_i, f_j)$ in (C3), we can use some well-known results. First, let

$$w = (x^2 + y^2)^{\frac{1}{2}} \ .$$

From Papoulis (1965), p. 196, and (C4) and (C5),

$$p_w(w) = (2w/\sigma^2) \exp\left[-(w^2 + s_{ij})/\sigma^2\right] I_o(2w\sqrt{s_{ij}}/\sigma^2) \qquad (C6)$$

for $w \geq 0$, and $f_w(w) = 0$ for $w < 0$. Since

$$z_{ij} = w^2 ,$$

we have, from Papoulis (1965), p. 129,

$$p_{z_{ij}}(z_{ij} \mid s_{ij}, H_1) = \sigma^{-2} \exp\left[-(z_{ij} + s_{ij})/\sigma^2\right]$$

$$I_o(2\sqrt{z_{ij}\, s_{ij}}/\sigma^2) . \qquad (C7)$$

When the signal is absent, $s_{ij} = 0$ and

$$p_{z_{ij}}(z_{ij} \mid H_0) = \sigma^{-2} \exp(-z_{ij}/\sigma^2) . \qquad (C8)$$

In (C7) and (C8), $I_o(\cdot)$ is a modified Bessel function of zero order.

Eq. (C7) is the same as the pdf obtained by McGill (1968) from the Rayleigh-Rice distribution (C6). Rice's derivation of (C6), however, is based upon the assumption that the filter responses are narrowband (Rice, 1954). Eq. (C2) has allowed us to discard the narrowband assumption.

Eqs. (C7) and C8) specify the pdf for one sample of the spectrogram. Assuming that samples are sufficiently far apart to be statistically independent when only noise is present, we can construct the pdf of an entire spectrogram by forming the products of the pdf's of the individual, independent samples, and the likelihood ratio is

$$\Lambda = \left[\prod_{i=1}^{N} \prod_{j=1}^{M} p(z_{ij} \mid s_{ij}, H_1) \right] \Big/ \left[\prod_{i=1}^{N} \prod_{j=1}^{M} p(z_{ij} \mid H_0) \right]$$

$$= \exp\left(- \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij}/\sigma^2 \right) \prod_{i=1}^{N} \prod_{j=1}^{M} I_o\left(2\sqrt{z_{ij}\, s_{ij}}/\sigma^2 \right) . \quad \text{(C9)}$$

A locally optimum detector for very small signal-to-noise ratio (SNR) can be obtained from (C9). The locally optimum detection statistic is determined by calculating $\partial \ell n \Lambda / \partial \theta$ at $\theta = 0$, where θ is an SNR parameter (Capon, 1961; Middleton, 1966). If the signal u(t) is amplified by a factor A, i.e., if the signal is written $A^2 u(t)$, then s_{ij} in (C9) is replaced by $A^2 s_{ij}$. The SNR parameter θ can then be defined as

$$\theta = A^2/\sigma^2 , \quad \text{(C10)}$$

and

$$\ell n \Lambda = \sum_{i=1}^{N} \sum_{j=1}^{M} \left[-\theta s_{ij} + \ell n I_o\left(2 \sqrt{(z_{ij}/\sigma^2)(\theta\, s_{ij})} \right) \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} \left\{ -\theta s_{ij} + \ell n \left[1 + \theta s_{ij} z_{ij}/\sigma^2 \right.\right.$$

$$\left.\left. + (\theta\, s_{ij} z_{ij}/\sigma^2)^2/4 + \ldots \right] \right\} . \quad \text{(C11)}$$

The locally optimum detector compares the quantity

$$\left.\frac{\partial \ell n \Lambda}{\partial \theta}\right|_{\theta = 0} = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(-s_{ij} + s_{ij} z_{ij} / \sigma^2\right) \tag{C12}$$

with a threshold. If the signal u(t) and the filter function v(t) have unit energy, then from (A8) the term

$$\sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij} \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_{uv}(t, f)\, dt\, df = E_u E_v \tag{C13}$$

is the same for all situations, and this term can be incorporated into the threshold. The threshold can also be multiplied by σ^2, and the locally optimum test for distinguishing between signal plus noise (H_1) or noise alone (H_0) is

$$\sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij} z_{ij} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{C14}$$

where γ is a threshold.

The locally optimum detector for very small SNR correlates sampled versions of the noise-free signal spectrogram $S_{uv}(t, f)$ and the data spectrogram, $Z_{uv}(t, f)$.

APPENDIX D: PHASE INFORMATION FOR DIRECTION ESTIMATES

A. Destruction of Relative Phase Information in Narrowband Signals

From Property 9 in Appendix A, a signal $u_1(t)$ that is reconstructed from its spectrogram is multiplied by a constant

$$\exp(j\lambda_1) = u_1^*(\tau_o) / \left|u_1(\tau_o)\right| \tag{D1}$$

where τ_o is a time that can be chosen by the receiver.

Let $u_1(t)$ be a narrowband signal from sensor No. 1, where

$$u_1(t) = a(t) \exp(j2\pi f_o t) .$$

From (D1), the phase of the reconstructed version of $u_1(t)$ is

$$\theta_1(t) = 2\pi f_o(t - \tau_o) . \tag{D2}$$

Let $u_2(t)$ be a delayed version of $u_1(t)$ that is observed at sensor No. 2, i.e.,

$$u_2(t) = u_1(t - \tau) = a(t - \tau) \exp\left[j2\pi f_o(t - \tau)\right] .$$

If an estimate of $u_2(t)$ is reconstructed from its spectrogram, its phase is

$$\theta_2(t) = 2\pi f_o(t - \tau) - 2\pi f_o(\tau_o - \tau) \tag{D3}$$

$$= 2\pi f_o(t - \tau_o) .$$

The relative delay parameter τ, which should be used to provide accurate narrowband direction estimates, is eliminated from $\theta_2(t)$ by the reconstruction process, and $\theta_1(t) = \theta_2(t)$ from (D2) and (D3). Narrowband signals that have been reconstructed from spectrograms therefore cannot provide accurate relative delay estimates for direction measurement.

B. Sufficient Phase Information for Complete Reconstruction of a Signal from Its Spectrogram

Two-dimensional deconvolution, together with the operations in Property 9 of Appendix A, yield a signal estimate

$$\tilde{u}(t) = u(t) \exp(j\lambda)$$

with Fourier transform

$$\tilde{U}(f) = U(f) \exp(j\lambda) . \qquad \text{(D4)}$$

The contribution of the jth filter to the spectrogram is

$$S(t, f_j) = |a_j(t)|^2$$

where

$$a_j(t) = \int U(f)\, V(f - f_j) \exp(j2\pi f t)\, df ,$$

and where $V(f)$ is the known transfer function of a low pass filter.

If $a_j(t)$ as well as $|a_j(t)|^2$ can be observed, then the Fourier transform

$$A_j(f) = U(f)\, V(f - f_j) \qquad \text{(D5)}$$

can be computed. The signal spectrum sampled at frequency f_j can be determined from (D5), i.e.,

$$U(f_j) = A_j(f_j) / V(0) . \tag{D6}$$

Combining (D4) and (D6),

$$\tilde{U}(f_j) \exp(-j\lambda) = A_j(f_j) / V(0)$$

and

$$\exp(j\lambda) = \tilde{U}(f_j) V(0) / A_j(f_j) . \tag{D7}$$

By observing the phase, as well as the amplitude, at the output of a single filter, we can deduce the value of the parameter λ that is added to the phase of the reconstructed spectrum $\tilde{U}(f_j)$. Knowledge of λ allows us to completely determine the signal u(t) from its spectrogram. A complete reconstruction of u(t) is useful for direction estimation with narrowband signals.

APPENDIX E: CRITICAL BAND EFFECTS

A. Reconstruction Error for Narrowband Signals in Narrowband Noise

Suppose that $P_i(\tau, \phi)$ in (B10) is set equal to a small, non-zero constant. The mean square error in the reconstructed signal will remain small if $\left|X_{vv}(\tau, \phi)\right|^2 P_e(\tau, \phi)$ is small, but MSE_0 will increase as $\left|X_{vv}(\tau, \phi)\right|^2 P_e(\tau, \phi)$ becomes larger. This concept can be exploited to heuristically predict the behavior of the estimator. In particular, we shall consider a narrowband signal with center frequency f_u and narrowband noise with center frequency f_n. After the transformation (10), the signal has center frequency $\ell n\, f_u$ and the noise has center frequency $\ell n\, f_n$.

From (23),

$$P_e(\tau, \phi) = E\left\{\left|X_{nn} + X_{un} + X_{nu}\right|^2\right\}$$

$$= E\left\{\left|X_{nn}\right|^2 + \left|X_{un}\right|^2 + \left|X_{nu}\right|^2 + 2\,\mathrm{Re}\,(X_{nn} X_{un} + X_{nn} X_{nu} + X_{un} X_{nu})\right\}. \quad \text{(E1)}$$

The half-power, constant-magnitude contour of $\left|X_{vv}(\tau, \phi)\right|^2$ is approximately an ellipse on the $(\tau\text{-}\phi)$ plane (Cook and Bernfeld, 1967) with ϕ-width equal to T_v^{-1} and τ-width equal to B_v^{-1}, and the ellipse is centered at $(\tau, \phi) = (0, 0)$. The half-power, constant-magnitude contour of $\left|X_{nn}(\tau, \phi)\right|^2$ is approximately an ellipse centered at $(\tau, \phi) = (0, 0)$ with ϕ-width equal to T_n^{-1} and τ-width equal to B_n^{-1}, where B_n and T_n are the bandwidth and duration of the narrowband noise sample n(t). The half-power, constant magnitude contour of

$|X_{un}(\tau, \phi)|^2$ is approximately an ellipse centered at $(\tau, \phi) = (0, \ell n f_u - \ell n f_n)$, with ϕ-width equal to $(T_n T_u)^{-\frac{1}{2}}$ and τ-width equal to $(B_n B_u)^{-\frac{1}{2}}$. The half-power, constant-magnitude contour of $|X_{nu}(\tau, \phi)|^2$ is approximately an ellipse centered at $(\tau, \phi) = (0, \ell n f_n - \ell n f_u)$, with ϕ-width $(T_n T_u)^{-\frac{1}{2}}$ and τ-width $(B_n B_u)^{-\frac{1}{2}}$. The product $|X_{vv}(\tau, \phi)|^2 P_e(\tau, \phi)$ becomes significantly larger when the $|X_{un}|^2$ and $|X_{nu}|^2$ ellipses begin to overlap with the $|X_{vv}|^2$ ellipse. Assuming that $T_n = T_u$, i.e., that the signal and noise sample functions have the same duration, overlap occurs when

$$\left|\ell n f_n - \ell n f_u\right| < (T_v^{-1} + T_u^{-1})/2 .$$

Assuming that $T_u \gg T_v$ and $B_v \simeq T_v^{-1}$, the product $|X_{vv}|^2 P_e$ becomes larger when

$$\max (f_u/f_n, f_n/f_u) < \exp (B_v/2) . \qquad \text{(E2)}$$

The above heuristic result suggests that the noise power in the reconstructed data (MSE_0) grows monotonically as $f_n/f_u \to 1$, provided that (E2) is satisfied. For example, the detectability of a reconstructed narrowband signal should remain constant as the center frequency f_n of the narrowband noise approaches f_u from below, until $f_n = f_u \exp(-B_v/2)$. For $f_n > f_u \exp(-B_v/2)$, the detectability should be degraded because of increasing noise power in the reconstructed process, until $f_n = f_u$. This interpretation is descriptive of a critical band effect for narrowband signals that are masked by narrowband noise.

B. Signal-to-Noise Ratio in a Spectrogram Correlator

A spectrogram correlation operation can give rise to a critical band effect when a narrowband signal is masked by narrowband noise. We shall use the following definitions:

$P_u(f)$ = power spectral density of the signal, which is a narrowband random process

$P_n(f)$ = power spectral density of noise

$S_{uv}(t, f)$ = signal spectrogram

$S_{nv}(t, f)$ = noise spectrogram

$S_{u_H v}(t, f)$ = spectrogram of a reference signal, $u_H(t)$.

Although the actual signal is modelled as a random process, we assume that the sinusoidal reference signal $u_H(t)$ is not random, and it has Fourier transform $U_H(f)$.

Cross-correlation of signal and reference spectrograms products an expected signal response

$$r_s = E\left\{ \iint S_{uv}(t, f)\, S_{u_H v}(t, f)\, dt\, df \right\}.$$

Cross-correlation of noise and reference spectrograms produces an expected noise response

$$r_n = E\left\{ \iint S_{nv}(t, f)\, S_{u_H v}(t, f)\, dt\, df \right\}.$$

Signal-to-noise ratio at the detector output is then

$$\text{SNR} = r_s/r_n = \frac{\displaystyle\iint_{-\infty}^{\infty} E\left\{S_{uv}(t,f)\right\}\ S_{u_H v}(t,f)\ dt\ df}{\displaystyle\iint_{-\infty}^{\infty} E\left\{S_{nv}(t,f)\right\}\ S_{u_H v}(t,f)\ dt\ df}, \tag{E3}$$

where

$E\left\{S_{uv}(t,f)\right\}$ = expected signal power from a filter with transfer function $V(x-f)$

$$= \int_{-\infty}^{\infty} P_u(x)\ |V(x-f)|^2\ dx \tag{E4}$$

$$E\left\{S_{nv}(t,f)\right\} = \int_{-\infty}^{\infty} P_n(x)\ |V(x-f)|^2\ dx \tag{E5}$$

and

$$S_{u_H v}(t,f) = \left| \int_{-\infty}^{\infty} U_H(x)\ V(x-f)\ e^{j2\pi xt}\ dx \right|^2 . \tag{E6}$$

It follows that

$$r_s = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} P_u(x_1) \left| V(x_1 - f) \right|^2 dx_1 \right] \left[\int_{-\infty}^{\infty} S_{u_H v}(t, f)\, dt \right] df . \quad \text{(E7)}$$

$$r_n = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} P_n(x_1) \left| V(x_1 - f) \right|^2 dx \right] \left[\int_{-\infty}^{\infty} S_{u_H v}(t, f)\, dt \right] df . \quad \text{(E8)}$$

where, from Property 5 (A6) in Appendix A,

$$\int_{-\infty}^{\infty} S_{u_H v}(t, f)\, dt = \int_{-\infty}^{\infty} \left| U_H(x_2) \right|^2 \left| V(x_2 - f) \right|^2 dx_2 . \quad \text{(E9)}$$

For narrowband signal and noise processes, we have

$$\begin{aligned} P_u(f) &\approx P_u e^f \delta(e^f - f_u) \\ P_n(f) &\approx P_n e^f \delta(e^f - f_n) \\ \left| U_H(f) \right|^2 &= P_H e^f \delta(e^f - f_H) \end{aligned} \quad \text{(E10)}$$

where the narrowband spectral representations $P_u \delta(f - f_u)$, $P_n \delta(f - f_n)$, and $P_H \delta(f - f_H)$ have been pre-distorted by the transformation in (10). Substituting (E9) and (E10) into (E7) and (E8),

$$r_s = P_u P_H \int_{-\infty}^{\infty} \left[\int_0^{\infty} \delta(x_1 - f_u) \, |V(\log x_1 - f)|^2 \, dx_1 \right]$$

$$\left[\int_0^{\infty} \delta(x_2 - f_H) \, |V(\log x_2 - f)|^2 \, dx_2 \right] \, df$$

$$= P_u P_H \int_{-\infty}^{\infty} |V(\log f_u - f)|^2 \, |V(\log f_H - f)|^2 \, df \qquad \text{(E11)}$$

$$r_n = P_n P_H \int_{-\infty}^{\infty} |V(\log f_n - f)|^2 \, |V(\log f_H - f)|^2 \, df \qquad \text{(E12)}$$

Assuming that $f_H = f_u$, i.e., that the center frequency of the signal process has been correctly hypothesized, we have

$$\text{SNR} = r_s / r_n = \frac{P_u}{P_n} \cdot \frac{\int_{-\infty}^{\infty} |V(f)|^4 \, df}{\int_{-\infty}^{\infty} |V(f)|^2 \left| V\left[f + \log(f_u / f_n)\right] \right|^2 \, df} . \qquad \text{(E13)}$$

SNR is minimized if the narrowband masking or noise process is such that $f_n = f_u$. For a filter $V(f)$ with bandwidth B_v, the noise has no effect if

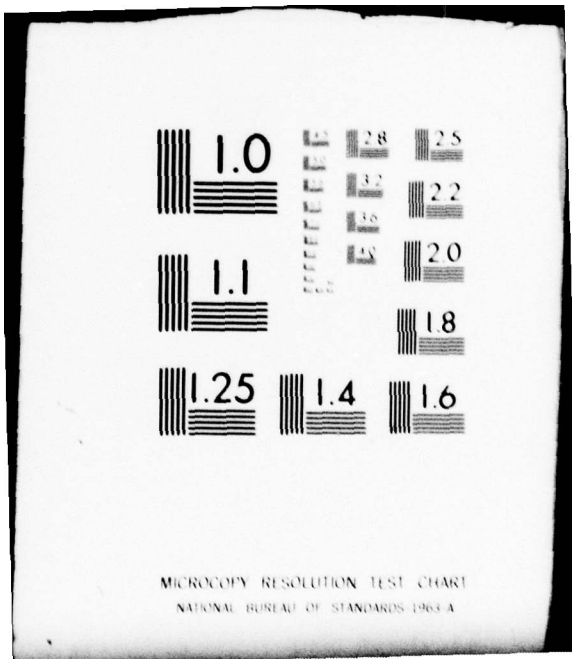
$$|\ell n (f_u / f_n)| > B_v \qquad \text{(E14)}$$

AD-A072 941 ORINCON CORP LA JOLLA CALIF F/G 17/1
FURTHER DEVELOPMENT AND NEW CONCEPTS FOR BIONIC SONAR. VOLUME 2--ETC(U)
OCT 78 R A ALTES, R W FLOYD N66001-78-C-0080
UNCLASSIFIED OC-R-78-A004-1-VOL-2 NOSC-TR-404-VOL-2 NL

2 OF 2
AD
A072941



END
DATE
FILMED
9-79
DDC


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

i.e., if

$$f_n/f_u > c \tag{E15}$$

or

$$f_n/f_u < 1/c \tag{E16}$$

where $c = \exp(B_v)$.

The SNR results for a narrowband signal in narrowband noise suggest that the filters in a spectrogram correlator can be characterized by measuring detection performance as a function of frequency difference between signal and noise. This method is well known to psychoacousticians, and (E14) – (E16) describe a critical band effect.

APPENDIX F: SLOPE OF THE P(C) VERSUS SNR CURVE FOR A 2AFC EXPERIMENT

This appendix is concerned with the interpretation of data from a two-alternative forced choice (2AFC) detection task. In particular, probability of correct response, P(C), as a function of signal energy, E_u, is to be compared for a matched filter and an energy detector. It will be shown that, for a given value of P(C), the slope of the P(C) versus E_u curve is larger for the matched filter case. The increase in slope depends upon the number of degrees of freedom, K, in the energy detector. This dependence is to be expected, since K = 2 corresponds to envelope detection at the output of a matched filter.

The reason for our interest in the slope of the P(C) versus energy curve is that a best fit to the measured psychometric function apparantly has steeper slope than the curve that is predicted on the basis of an energy detector. This disparity is clearly shown in Figure 8-3 of Green and Swets (1966). We suspect that this steeper slope is indicative of a matched filtering mechanism.

For a sinusoidal signal with fixed duration and a fixed internal filter bandwidth, K is fixed. Signal-to-noise ratio ($2E_u/N_o$) is varied by changing the power of the signal. When the signal is absent, the energy detector output is represented by a random variable x that has a chi-square distribution with mean K and variance 2K. When the signal is present, the response of the energy detector is a random variable y that has a noncentral chi-square distribution with mean $K + 2E_u/N_o$ and variance $2K + 8E_u/N_o$, where E_u is signal energy and $N_o/2$ is noise power spectral density. If $K \geq 10$, Green and Swets (1966) contend that the chi-square and noncentral chi-square distributions are approximately normal.

In this case, the distribution of $z = y - x$ is Gaussian with mean $m_{ED} = 2E_u/N_o$ and variance $\sigma_{ED}^2 = 4K + 8E_u/N_o$. In a 2AFC procedure, the probability of a correct decision is the probability that $y - x > 0$, i.e.,

$$P_{ED}(C) = \int_0^\infty \frac{e^{-(z - m_{ED})^2/2\sigma_{ED}^2}}{\sqrt{2\pi\sigma_{ED}^2}} dz$$

$$= \int_0^\infty \frac{e^{-(z - m_{ED}/\sigma_{ED})^2/2}}{\sqrt{2\pi}} dz$$

$$= \Phi(h_{ED}), \text{ where } h_{ED} = m_{ED}/\sigma_{ED}. \qquad \text{(F1)}$$

In (F1), $\Phi(h_{ED})$ is a monotone increasing function of h_{ED}.

Now consider a 2AFC experiment with a matched filter, rather than an energy detector. In the absence of signal, the output x of a matched filter is Gaussian with mean zero and variance $(N_o/2)E_u$, where E_u is the energy of the filter impulse response, which is a time-reversed replica of the signal. When a signal is present, the maximum output y of the matched filter is Gaussian distributed with mean E_u and variance $(N_o/2)\,E_u$. The random variable $z = y - x$ is then Gaussian with mean $m_{MF} = E_u$ and variance $\sigma_{MF}^2 = N_o E_u$, and

$$P_{MF}(C) = \Phi(h_{MF}), \text{ where } h_{MF} = m_{MF}/\sigma_{MF} \qquad \text{(F2)}$$

and where $\Phi(h)$ is defined as in (F1).

For a given value of P(C), $h = \Phi^{-1} P(C)$. For example, if $P_{ED}(C) = P_{MF}(C) = 0.75$, then

$$h_{ED} = h_{MF} = 0.68 . \qquad \text{(F3)}$$

For $P(C) = 0.75$, we shall calculate the ratio r of the slopes of $P_{MF}(C)$ and $P_{ED}(C)$, as functions of E_u, i.e.,

$$r = \frac{dP_{MF}(C)/dE_u}{dP_{ED}(C)/dE_u}$$

$$= \frac{\left[d\Phi(h_{MF})/dh_{MF}\right] dh_{MF}/dE_u \Big|_{E_u = E_{MF}}}{\left[d\Phi(h_{ED})/dh_{ED}\right] dh_{ED}/dE_u \Big|_{E_u = E_{ED}}} . \qquad \text{(F4)}$$

From (F3), we have

$$r = (dh_{MF}/dE_u \Big|_{E_u = E_{MF}}) \; / \; (dh_{ED}/dE_u \Big|_{E_u = E_{ED}}) \qquad \text{(F5)}$$

where

$$h_{MF} = m_{MF}/\sigma_{MF} = \sqrt{E_u/N_o} \Big|_{E_u = E_{MF}} \equiv 0.68 \qquad \text{(F6)}$$

$$h_{ED} = m_{ED}/\sigma_{ED} = (E_u/N_o)/(K + 2E_u/N_o)^{\frac{1}{2}} \equiv 0.68 . \qquad \text{(F7)}$$

For a given value of P(C), the curves for an energy detector and a matched filter will correspond to different values of E_u. These energy values have been denoted E_{MF} and E_{ED}, and they can be determined by solving (F6) and (F7). Substituting (F6) and (F7) into (F5), we have

$$r = \frac{1}{2\sqrt{E_{MF}/N_o}} \frac{(K + 2E_{ED}/N_o)^{3/2}}{(K + E_{ED}/N_o)} . \qquad \text{(F8)}$$

It is easy to show that $r > 1$. Equating (F6) and (F7), we see that

$$\begin{aligned} E_{MF}/N_o &= (E_{ED}/N_o)^2 / (K + 2E_{ED}/N_o) \\ &< (E_{ED}/N_o)^2 / (2E_{ED}/N_o) = (1/2)(E_{ED}/N_o) \\ &< E_{ED}/N_o \end{aligned} \qquad \text{(F9)}$$

which means that the matched filter curve always lies to the left of the energy detector curve. This displacement of the P(C) versus $2E_u/N_o$ curve could ordinarily be used to differentiate between a matched filter and an energy detector, but one can argue that $2E_u/N_o$ cannot be directly measured due to internal noise. From (F9), it is apparent that

$$2\sqrt{E_{MF}/N_o} < \sqrt{K + 2E_{ED}/N_o} . \qquad \text{(F10)}$$

It is also obvious that

$$K + E_{ED}/N_o < K + 2E_{ED}/N_o \quad . \tag{F11}$$

From (F10) and (F11),

$$2\sqrt{E_{MF}/N_o}\ (K + E_{ED}/N_o) < (K + 2E_{ED}/N_o)^{3/2} \quad . \tag{F12}$$

Comparing (F12) to (F8), we see that

$$r > 1 \quad , \tag{F13}$$

which means that the slope of the P(C) versus E_u curve, for any given value of P(C), is steeper for the matched filter than for the energy detector. The actual value of r is dependent upon the specific values of K and P(C).

Equation (F8) was obtained under the assumption that $K \geq 10$, since the chi-square distributions were approximated by normal distributions. If $K >> 1$ and $K >> 2E_u/N_o$, then (F6) and (F7) become

$$E_{MF}/N_o = (0.68)^2$$

$$E_{ED}/N_o = 0.68\,K^{\frac{1}{2}}$$

and (F8) becomes

$$r \approx 0.735\,K^{\frac{1}{2}} \quad . \tag{F14}$$

Since r in (F14) increases with K, we shall choose K = 10 to obtain the least value of r that is allowed under the assumption that $K \geq 10$. For K = 10, (F6) and (F7) give

$$E_{MF}/N_o = (0.68)^2$$

$$E_{ED}/N_o = 2.66 \ .$$

Substituting these results into (F8), we obtain

$$r = 3.5 \ . \qquad \text{(F15)}$$

A best fit to the data in Figure 8-3 of Green and Swets (1966) indicates that the slope of the data curve divided by the slope of the energy detection curve at P(C) = 0.75 is

$$r' \approx 1.5 \ . \qquad \text{(F16)}$$

The slope ratio r for a matched filter relative to an energy detector with ten or more degrees of freedom is therefore greater than the actual slope ratio r'. that is observed in the data. Although a best fit to the data seems to indicate a receiver that is better than an energy detector, this receiver is not as good as a perfect matched filter, if $K \geq 10$ in the energy detector. A similar conclusion was drawn by Taylor and Forbes (1969) as the result of a monaural detection experiment with the noise-free signal presented simultaneously to the other ear. Perhaps the actual detector is an imperfect matched filter (based upon an imperfectly reconstructed signal) or an estimator-correlator.

Incidentally, the relative performance of the matched filter and energy detector for the 2AFC task is very different from the result in Figure 3, which was not based upon a forced choice experiment. In Figure 3, it was also assumed that power is fixed and E_u/N_o is varied by changing signal duration, i.e., by varying K.

For the 2AFC case with variable power and fixed K, where $2E_{ED}/N_o \ll K$, (F6) and (F7) indicate that the same performance is obtained when

$$\sqrt{E_{MF}/N_o} = E_{ED}/(N_o K^{\frac{1}{2}})$$

or

$$\alpha = SNR_{ED} / SNR_{MF} = \sqrt{2} K^{\frac{1}{2}} / SNR_{MF}^{\frac{1}{2}} . \qquad (F17)$$

For a 2AFC experiment with a fixed value of K, where $K \geq 10 \gg 2E_u/N_o$, α varies inversely with $SNR_{MF}^{\frac{1}{2}}$.

The relative performance of the matched filter is thus most impressive for small signal-to-noise ratio, as one would intuitively expect.

2.3 Parameter Estimation from Spectrograms

2.3.1 Brief Summary of Section 2.3

The concept of a locally optimum estimator is introduced and applied to spectrogram analysis. Variance bounds for maximum likelihood, locally optimum spectrogram estimates of range (or time of occurrence), and Doppler shift (or frequency) are derived. The predicted standard deviation of a spectrogram frequency estimate is compared to psychophysical data. The comparison indicates that spectrogram correlation may be a viable model for mammalian audition.

2.3.2 Introduction

Neurophysiological evidence indicates that spectrogram-like signal representations may be formed by the peripheral auditory system (Evans, 1975 and 1977; Pfeiffer and Kim, 1975). The major fault with a spectrogram model is that at low frequencies (below 6 kHz) extra timing information, indicating peak or zero crossing locations of sine-wave stimuli, is encoded by peripheral neurons (Anderson, et al., 1971). Sine wave peaks or zero crossings would not be apparent if ideal envelope detectors were used at the outputs of a bank of bandpass filters, as in a spectrogram model. Siebert (1970), however, has found that auditory frequency discrimination can best be explained if the peak or zero crossing information is disregarded by the central nervous system. Siebert's result therefore favors a conventional spectrogram representation in the central nervous system.

In this chapter, we begin with the supposition that a spectrogram is available to an ideal observer, and we derive the observer's

performance in estimating signal parameters from the spectrogram under low signal-to-noise ratio conditions. Siebert (1970) has used a similar approach to obtain estimator performance from a more detailed physiological model that includes the shape of neural tuning curves and a representation of neural firing patterns as rate-modulated Poisson processes. Here, we begin with a conventional spectrogram and "plug in" behavioral data rather than neurophysiological data. The resulting estimator performance will be seen to support the concept of a spectrogram representation in the central nervous system.

2.3.3 A Locally Optimum Maximum Likelihood Estimator

Let $\{z_{mn}\}$ denote the set of samples corresponding to a spectrogram of the data. If H_1 is true, i.e., if a signal is present, then we hypothesize that the noise-free spectrogram is given by a set of samples $\left\{s^H_{mn}\right\}$. It was shown in Section 2.2 that

$$\ln p\left[\left\{z_{mn}\right\}\middle|\left\{s^H_{mn}\right\}, H_1\right] = \ln\left[\prod_{m=1}^{M}\prod_{n=1}^{N} \sigma^{-2} e^{-\left(z_{mn}+s^H_{mn}\right)/\sigma^2} I_o\left(2\sqrt{z_{mn}\, s^H_{mn}}\,/\sigma^2\right)\right]$$

$$\approx -2MN \ln\sigma - \sum_{m,n}\left(z_{mn}+s^H_{mn}\right)/\sigma^2 + \sum_{m,n} z_{mn}\, s^H_{mn}\,/\sigma^4 \qquad (1)$$

where the approximation is valid for small signal-to-noise ratio.

It will be assumed that

$$\sum_{m,n} z_{mn} = \sum_{m,n} s^{H}_{mn} = 1, \tag{2}$$

i.e., that data and hypothetical spectrograms are normalized to have unit volume. From Property 6 in Appendix A of Section 2.2, we see that the volume normalization condition (2) implies that signal and filter functions are energy normalized. From (1) and (2), $\ell n\, p\left[\underline{z} \mid \underline{s}^{H}, H_1\right]$ is maximized when

$$\sum_{m,n} z_{mn}\, s^{H}_{mn}$$

is maximized, i.e., when the correlation of data spectrogram and hypothesized spectrogram is maximized.

For small SNR, a maximum likelihood (ML) estimate of any signal parameter ρ is obtained by computing

$$\sum_{m,n} z_{mn}\, s_{mn}(\rho_H) \tag{3}$$

for all ρ_H. The ML estimate $\hat{\rho}$ is the value of ρ_H that maximizes the above summation. The sum in (3) is a spectrogram correlation process, i.e., it is a correlation of the data spectrogram with a sequence of hypothesized noise-free signal spectrograms.

2.3.4 A Variance Bound for the Locally Optimum ML Estimate

For a large number of observations, the variance of a ML estimate approaches the Cramér-Rao (CR) lower bound. An ML

estimate for low SNR data is implemented by spectrogram correlation. The asymptotic accuracy of parameter estimation by means of a spectrogram correlator is therefore given by a CR bound that is computed for low SNR data.

For an estimate $\hat{\rho}$ of a parameter ρ, the CR bound is

$$\mathrm{Var}\,(\rho - \hat{\rho}) \geq \left\{ -E\left[(\partial^2/\partial\rho^2)\, \ell n\, p\left[\underline{z} \,|\, \underline{s}(\rho),\, H_1 \right] \right] \right\}^{-1} . \quad (4)$$

For low SNR $\ell n\, p\left[\underline{z} \,|\, \underline{s}\,(\rho),\, H_1 \right]$ is given by (1), and

$$-E\left\{ (\partial^2/\partial\rho^2)\, \ell n\, p\left[\underline{z} \,|\, \underline{s}\,(\rho),\, H_1 \right] \right\}$$

$$\approx -\sum_{m,\,n} s_{mn}''(\rho)\, s_{mn}(\rho)/\sigma^4 \quad (5)$$

where $s_{mn}''(\rho) = (\partial^2/\partial\rho^2)\, s_{mn}(\rho)$.

The dependence of the bound in (5) upon signal and filter parameters can be determined by specifying $s_{mn}(\rho)$ in terms of the signal u(t) and the filter v(t) that are used to construct the spectrogram.

The frequency window (bank-of-filters) spectrogram can be written in terms of frequency-domain functions or time-domain functions. The frequency domain form is (Ackroyd, 1971)

$$s_{mn}(\rho) = \left| \int U(f, \rho)\, V(f - f_n)\, \exp(j2\pi f t_m)\, df \right|^2 \quad (6)$$

where unlabeled integration limits correspond to $(-\infty, \infty)$. In (6), $U(f,,\rho)$ is the Fourier transform of the input signal (which depends upon the parameter ρ), and V(f) is the basic filter function that is used to construct the spectrogram. The time-domain version of the spectrogram is

$$s_{mn}(\rho) = \left| \int u(t, \rho)\, v(t_m - t)\ \exp(-j2\pi f_n t)\, dt \right|^2 . \qquad (7)$$

From (6),

$$s_{mn}''(\rho) = 2 \left| \int U'(f, \rho)\, V(f - f_n)\, e^{j2\pi f t_m}\, df \right|^2 + 2\, \mathrm{Re} \left\{ \int U''(f, \rho)\, V(f - f_n)\, e^{j2\pi f t_m}\, df \int U^*(f, \rho)\, V(f - f_n)\, e^{-j2\pi f t_m}\, df \right\} \qquad (8)$$

where the primes denote differentiation with respect to ρ. From (7),

$$s_{mn}''(\rho) = 2 \left| \int u'(t, \rho)\, v(t_m - t)\, e^{-j2\pi f_n t}\, dt \right|^2 + 2\, \mathrm{Re} \left\{ \int u''(t, \rho)\, v(t_m - t)\, e^{-j2\pi f_n t}\, dt \int u^*(t, \rho)\, v^*(t_m - t)\, e^{j2\pi f_n t}\, dt \right\} . \qquad (9)$$

Equations (6) – (9) can be substituted into (5) in order to determine the effect of signal and filter functions upon variance estimates. In the following sections, the above analysis will be applied to range and Doppler estimates, and some of the results will be compared with behavioral data.

2.3.5 Delay (Range) Estimation

For delay estimation, $U(f, \rho)$ in (6) becomes

$$U(f, \tau) = U(f) \exp(-j2\pi f \tau) . \qquad (10)$$

Substituting (10) into (8), we obtain

$$s_{mn}''(\tau) = 2 \left| \int 2\pi f\, U(f)\, V(f - f_n)\, e^{j2\pi f(t_m - \tau)}\, df \right|^2$$

$$- 2\, \mathrm{Re} \left\{ \int (2\pi f)^2\, U(f)\, V(f - f_n)\, e^{j2\pi f(t_m - \tau)}\, df \right.$$

$$\left. \int U^*(f)\, V^*(f - f_n)\, e^{-j2\pi f(t_m - \tau)}\, df \right\} . \qquad (11)$$

For a signal bandwidth B_u that is much greater than the filter bandwidth B_v, we can assume that

$$U(f) \approx U(f_n) \text{ over the bandwidth of } V(f - f_n) . \qquad (12)$$

Using this assumption and changing variables in (11), we obtain,

$$- s_{mn}''(\tau) \approx 2(2\pi)^2 |U(f_n)|^2 \left[\mathrm{Re} \left\{ \int f^2 V(f)\, e^{j2\pi f(t_m - \tau)}\, df \right. \right.$$

$$\left. \int V^*(f)\, e^{-j2\pi f(t_m - \tau)}\, df \right\}$$

$$\left. - \left| \int f V(f)\, e^{j2\pi f(t_m - \tau)}\, df \right|^2 \right] . \qquad (13)$$

From (12) and (13), we have

$$- s_{mn}''(\tau) \leq 2(2\pi)^2 |U(f_n)|^2 \int f^2 |V(f)|\, df \int |V(f)|\, df \qquad (14)$$

and

$$s_{mn}(\tau) \approx |U(f_n)|^2 \left| \int V(f)\, e^{j2\pi (f+f_n)(t_m - \tau)}\, df \right|^2$$

$$\leq |U(f_n)|^2 \left| \int |V(f)|\, df \right|^2 . \qquad (15)$$

From (14) and (15)

$$- \sum_{m=1}^{M} s_{mn}''(\tau)\, s_{mn}(\tau)$$

$$\leq -2M(2\pi)^2 \left| U(f_n) \right|^4 \int f^2 \left| V(f) \right|^2 df \left[\int \left| V(f) \right| df \right]^3$$

where M is the number of time samples such that $s_{mn}(\tau)$ is not identically zero.

We shall assume that the signal u(t) is a broadband pulse with duration T_u which is much less than the duration T_v of the filter impulse response. M is then the number of time samples in the filter impulse response v(t). Recall from Section 2.2 that the spectrogram is sampled at a rate of $4B_v$ samples/sec in the time direction. It follows that

$$M = 4B_v T_v \approx 4$$

and, from (14) and (15),

$$- \sum_{n=1}^{N} \sum_{m=1}^{M} s_{mn}''(\tau)\, s_{mn}(\tau)/\sigma^4$$

$$\leq \sum_{n=1}^{N} 8(2\pi)^2 |U(f_n)|^4 \int f^2 |V(f)|\, df \left[\int |V(f)|\, df \right]^3 / \sigma^4 . \tag{16}$$

The normalized second moment

$$\int f^2 |V(f)|\, df \; / \int |V(f)|\, df \equiv B_v^2 \tag{17}$$

is a measure of the mean square filter bandwidth. If $|U(f_n)|^2$ is approximately constant over N frequency samples and is approximately zero for all other frequency samples, then

$$|U(f_n)|^4 \left[\int |V(f)|\, df \right]^4 / \sigma^4 \equiv SNR^2 \tag{18}$$

is a measure of squared signal-to-noise ratio, and (16) becomes

$$- \sum_{n=1}^{N} \sum_{m=1}^{M} s_{mn}''(\tau)\, s_{mn}(\tau)/\sigma^4 \leq 8(2\pi)^2 N B_v^2 SNR^2 , \tag{19}$$

where N is the number of non-zero frequency samples of the signal spectrogram, which is sampled at a rate of $4T_v$ samples/Hz in the frequency direction (from Section 2.2). For a signal bandwidth B_u, we have

$$N = 4T_v B_u \approx 4B_u/B_v \text{ samples} . \tag{20}$$

Substituting (20) into (19) and using (4) and (5),

$$\text{Var}\,(\tau - \hat{\tau}) \geq \left[32(2\pi)^2 B_v B_u SNR^2 \right]^{-1} . \tag{21}$$

For a broadband pulse with $T_u << T_v$ and $B_u >> B_v$, the standard deviation in the time of arrival estimate is

$$\Delta\tau \geq (c_o B_v^{-\frac{1}{2}}) B_u^{-\frac{1}{2}} SNR^{-1} \tag{22}$$

where c_o is a constant. This result is significantly different from the estimation of range with a matched filter, which has standard deviation (Cook and Bernfeld, 1967)

$$\Delta\tau_{MF} \geq B_u^{-1} SNR^{-\frac{1}{2}} \qquad \text{for large SNR}$$

$$\Delta\tau_{MF} \geq B_u^{-1} SNR^{-1} \qquad \text{for small SNR .} \tag{23}$$

It would seem that (22) and (23) provide a basis for discriminating between a matched filter model and a spectrogram correlation model in animal echolocation. Some relevant behavioral tests will be suggested further on.

2.3.6 Frequency (Doppler Shift) Estimation

For Doppler shift estimation, $u(t,\rho)$ in (7) becomes

$$u(t, f_d) = u(t) \exp\left[j2\pi (f_u - f_d)\, t\right] \tag{24}$$

where f_u is the frequency of the transmitted signal. From (9),

$$s_{mn}''(f_d) = 2 \left| \int 2\pi t\, u(t)\, v(t_m - t)\, e^{-j2\pi(f_n + f_d - f_u)t}\, dt \right|^2$$

$$- 2\,Re \left\{ \int (2\pi t)^2\, u(t)\, v(t_m - t)\, e^{-j2\pi(f_n + f_d - f_u)t}\, dt \right.$$

$$\left. \int u^*(t)\, v^*(t_m - t)\, e^{j2\pi(f_n + f_d - f_u)t}\, dt \right\}. \tag{25}$$

For a signal envelope u(t) that is relatively smooth and is much longer than T_v, we can assume that

$$u(t) \approx u(t_m) \text{ over the duration of } v(t_m - t) . \tag{26}$$

Using this assumption in (25) and changing variables, we obtain

$$-s_{mn}''(f_d) = 2(2\pi)^2 |u(t_m)|^2 \left[\mathrm{Re} \left\{ \int_t^2 v(t)\, e^{j2\pi f t} \right. \right.$$

$$\left. \left. v^*(t)\, e^{-j2\pi f t}\, dt \right\} - \left| \int_t v(t)\, e^{j2\pi f t}\, dt \right|^2 \right] \tag{27}$$

where

$$f \equiv f_n + f_d - f_u = f_n - (f_u - f_d) . \tag{28}$$

From (26) and (27),

$$-s_{mn}''(f_d) \le 2(2\pi)^2 |u(t_m)|^2 \int_t^2 |v(t)|\, dt \int |v(t)|\, dt \tag{29}$$

and

$$s_{mn}(f_d) \approx |u(t_m)|^2 \left| \int v(t)\, e^{j2\pi f t}\, dt \right|^2$$

$$\le |u(t_m)|^2 \left[\int |v(t)|\, dt \right]^2 . \tag{30}$$

Therefore,

$$- s_{mn}''(f_d)\, s_{mn}(f_d)/\sigma^4 \le 2(2\pi)^2 |u(t_m)|^4 \int t^2 |v(t)|\, dt \left[\int |v(t)|\, dt\right]^3/\sigma^4 \quad (31)$$

where

$$\int t^2 |v(t)|\, dt \Big/ \int |v(t)|\, dt \equiv T_v^2 \quad (32)$$

is a measure of mean square duration of v(t). Assuming that $|u(t_m)|$ is approximately constant over M time samples and is approximately zero elsewhere,

$$|u(t_m)|^4 \left[\int |v(t)|\, dt\right]^4/\sigma^4 \equiv SNR^2 \quad (33)$$

is a measure of squared signal-to-noise ratio. From (31) – (33), we have

$$- \sum_{m=1}^{M} s_{mn}''(f_d)\, s_{mn}(f_d)/\sigma^4 \le 2(2\pi)^2 M T_v^2 SNR^2 \quad (34)$$

where M is the number of non-zero responses of the nth filter, where $T_u \gg T_v$ for all n. The response of each filter is approximately T_u seconds long, and the response is sampled at a rate of $4B_v$ samples/sec. It follows that

$$M \approx 4B_v T_u \approx 4T_u/T_v \text{ samples} \quad (35)$$

and

$$- \sum_{n=1}^{N} \sum_{m=1}^{M} s_{mn}''(f_d)\, s_{mn}(f_d)/\sigma^4$$

$$\leq N(8)(2\pi)^2 T_v T_u SNR^2 . \qquad (36)$$

In (36), N is the number of non-zero samples in the frequency direction. For a very narrowband signal with center frequency $f_u - f_d$ Hz, the only filters that have a non-zero response have center frequencies f_n such that $(f_u - f_d) - B_v/2 < f_n < (f_u - f_d) + B_v/2$. For a non-zero response, f_n covers a band that is B_v Hz wide. Since f_n is sampled every $(4T_v)^{-1}$ Hz $\approx B_v/4$ Hz, the number of non-zero frequency samples is

$$N \approx 4B_v^{-1} \text{ samples/Hz} \times B_v \text{ Hz} = 4 \qquad (37)$$

and

$$\text{Var}(f_d - \hat{f}_d) \geq \left[32(2\pi)^2 T_v T_u SNR^2\right]^{-1} . \qquad (38)$$

For a narrowband pulse with $T_u \gg T_v$ and $B_u \ll B_v$, the standard deviation of a Doppler or signal frequency estimate is

$$\Delta f \geq (c_o T_v^{-\frac{1}{2}}) T_u^{-\frac{1}{2}} SNR^{-1} . \qquad (39)$$

If n_H harmonics are used in addition to a fundamental component, then N in (36) becomes $N(1+n_H)$, and

$$\Delta f \geq (c_o T_v^{-\frac{1}{2}}) (1 + n_H)^{-\frac{1}{2}} T_u^{-\frac{1}{2}} SNR^{-1} . \qquad (40)$$

If a frequency estimate is obtained with a bank of matched filters, then (Cook and Bernfeld, 1967)

$$\Delta f \geq T_u^{-1} SNR^{-\frac{1}{2}} \qquad \text{for large SNR}$$

$$\Delta f \geq T_u^{-1} SNR^{-1} \qquad \text{for small SNR} . \qquad (41)$$

The standard deviation in (39) can be compared with behavioral data. Such a comparison is described below. Comparison of behavioral data with (39) and (41) should indicate whether spectrogram correlation or matched filtering is closest to reality.

2.3.7 Recommended Experiments and Comparison of Results with Existing Data

Error bounds have been obtained for estimates that use matched filters and for estimates that are derived from spectrograms. These bounds can be compared with behavioral data in order to determine whether spectrogram correlation or matched filter models are more descriptive of auditory signal processing in mammals.

It has been shown that the standard deviation of a delay measurement using spectrogram correlation is asymptotically proportional to $(SNR)^{-1} B_u^{-\frac{1}{2}}$ where B_u is signal bandwidth and SNR is a measure of signal-to-noise ratio. Delay measurement with a matched filter, on the other hand, has a standard deviation that varies as $(SNR)^{-1} B_u^{-1}$ under similar conditions.

An animal or human can be trained to indicate which of two pulse-pairs has the shorter time between pulses, or to distinguish between a single pulse and a pulse pair until the interval between

pulses in the pair becomes too small. The bandwidth of the pulses can then be varied for a fixed SNR. The accuracy of an interpulse delay measurement or the ability to resolve two closely spaced pulses will vary as B_u^{-1} if a matched filter model is correct and as $B_u^{-\frac{1}{2}}$ if a spectrogram correlator model is more descriptive of auditory processing.

Experimental data for frequency discrimination have already been published by Licklider (1951), Chih-an and Chistovich (1961), and Oetinger (1959). Siebert (1970) has fit this data with the equation

$$\Delta f_{data} = c_1 f T_u^{-\frac{1}{2}} , \tag{42}$$

where c_1 is a constant.

The dependence of Δf_{data} upon $T_u^{-\frac{1}{2}}$ is consistent with the bound in (39), which also varies with $T_u^{-\frac{1}{2}}$. In fact, (42) and (39) will be identical if

$$SNR^{-1} T_v^{-\frac{1}{2}} \propto f . \tag{43}$$

The duration T_v of a filter impulse response can be experimentally determined by exciting the filter with a time-gated sine wave. As the duration of the sine wave becomes longer, the filter output power increases, until the duration of the sine wave equals T_v. For longer sine wave durations, the filter output power stays constant.

Threshold as a function of sine wave duration has been measured for human subjects (Plomp and Bouman, 1959). Threshold decreases with increased stimulus duration up to a duration τ.

Figure 1 shows τ as a function of f, as measured by Plomp and Bouman (1959). The solid line in Figure 1 connects the mean values of τ for two different observers. This line is replotted in Figure 2 on coordinates of $\tau^{-\frac{1}{2}}$ versus frequency. From Figure 2, we see that the data can be approximated by a straight line, i.e.,

$$\tau^{-\frac{1}{2}} \propto f \tag{44}$$

between 0.5 kHz and 8 kHz. Assuming that $\tau \propto T_v$ and that SNR is constant, an auditory spectrogram correlator should have frequency uncertainty

$$\Delta f \approx (\text{constant})\ f\ T_u^{-\frac{1}{2}} \tag{45}$$

which is the same as Δf_{data} in (42).

Even if $T_v^{-\frac{1}{2}}$ were not proportional to f, the fact that Δf is proportional to $T_u^{-\frac{1}{2}}$ rather than to T_u^{-1} would seem to indicate that a spectrogram correlator model is better than a matched filter model for auditory frequency estimation.

2.3.8 Conclusion

For low signal-to-noise ratio, a maximum likelihood parameter estimate is obtained from spectrogram data by means of a spectrogram correlation operation. The variance of this estimate is asymptotically equal to the Cramér-Rao lower bound. This bound has been calculated for an estimate of delay (using a

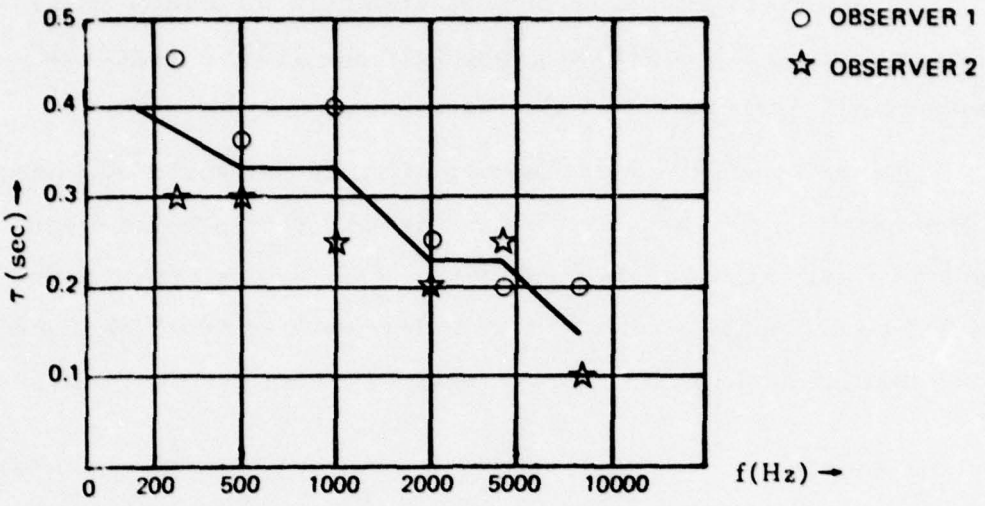


Figure 1. Duration parameter τ as a function of frequency (from Plomp and Bouman, 1959).

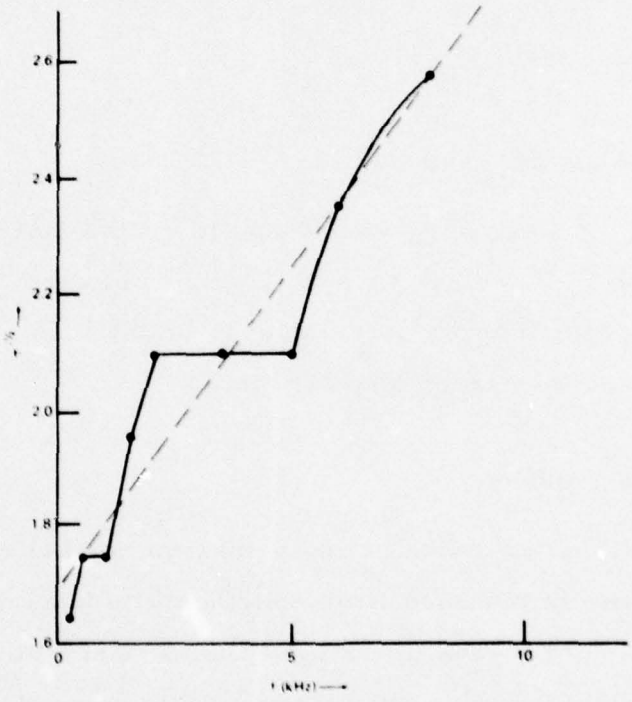


Figure 2. The solid line in Figure 1, replotted on coordinates of τ-1/2 versus frequency.

broadband pulse that is narrow in time), and for an estimate of frequency or Doppler shift (using a long-duration, narrowband signal).

The variances of the spectrogram correlator delay and frequency estimates are different from the variances that are associated with a matched filter (for delay) or a bank of matched filters (for Doppler). These differences can be used to construct behavioral tests to distinguish between matched filter and spectrogram correlator models of animal echolocation.

Frequency estimation data has already been obtained from human subjects. This data indicates that the standard deviation of a frequency estimate is inversely proportional to the square root of tone duration. The spectrogram correlator frequency estimate is also such that $\Delta f \propto T_u^{-\frac{1}{2}}$, where T_u is signal duration.

The experimentally observed standard deviation is proportional to the frequency of a sinusoidal signal, as well as to T_u^{-1}. The same frequency dependence occurs in the spectrogram correlator estimate provided that T_v, the impulse response duration of a filter that is used to construct the spectrogram, is such that $T_v^{-\frac{1}{2}} \propto f$. An estimate of T_v, obtained from auditory threshold as a function of tone duration, indicates that $T_v^{-\frac{1}{2}}$ may indeed be proportional to frequency.

The standard deviation of the spectrogram correlator frequency estimate is therefore very similar to behavioral data. This similarity suggests that spectrogram correlation may be utilized in mammalian audition. An additional test of this hypothesis is to measure the just noticeable delay difference between two broadband pulses as a function of pulse bandwidth and signal-to-noise ratio,

and to compare the data with $\Delta\tau$ in (22), the calculated standard deviation of the corresponding spectrogram correlator estimate.

2.3.9 References for Section 2.3

Ackroyd, M. H. (1971). "Short-time spectra and time-frequency energy distributions," J. Acous. Soc. Am. 50, 1229-1231.

Anderson, D. J.; Rose, J. E.; Hind, J. E.; and Brugge, J. F. (1971). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine wave stimulus: frequency and intensity effects," J. Acous. Soc. Am. 49, 1131-1139.

Chih-an, L., and Chistovich, L. A. (1961). "Frequency difference linears as a function of tonal duration," Soviet Phys. Acoust. 6, 75-80.

Cook, C. E., and Bernfeld, M. (1967). Radar Signals. New York: Academic.

Evans, E. F. (1977). "Peripheral processing of complex sounds," in Recognition of Complex Acoustic Signals, T. H. Bullock, ed. Berlin: Abakon Verlagsgesellschaft, 145-159.

Evans, E. F. (1975). "Cochlear nerve and cochlear nucleus," in Handbook of Sensory Physiology, vol. 5, part II (W. D. Keidel and W. D. Neff, eds.). Berlin: Springer-Verlag, 1-108.

Licklider, J. C. R. (1951). "Basic correlates of the auditory stimulus," in Handbook of Experimental Psychology, S. S. Stevens, ed. New York: Wiley, 985-1039.

Oetinger, R. (1959). "Die grenzen dur horbarkeit von frequenz- und tonzahlanderungen bei tonimpulsen," Acoustica 9, 430-434.

Pfeiffer, R. R., and Kim, D. O. (1975). "Cochlear nerve fiber responses: Distribution along the cochlear partition," J. Acous. Soc. Am. 58, 867-869.

Plomp, R., and Bouman, M. A. (1959). "Relation between hearing threshold and duration for tone pulses," J. Acous. Soc. Am. 31, 749-758.

Siebert, W. M. (1970). "Frequency discrimination in the auditory system: place or periodicity mechanisms?" Proc. IEEE 58, 723-730.

2.4 Spectrogram Correlation as an Ideal Detection Process for Signals that Have Been Passed Through a Time-Varying, Random Channel

2.4.1 Brief Summary of Section 2.4

When a random or deterministic signal is passed through a time-varying, random channel, the expected spectrogram of the channel output is the two-dimensional convolution of the channel scattering function and the expected spectrogram of the input signal. In Gaussian noise, a locally optimum detector correlates samples of the data spectrogram with corresponding samples of the noise-free spectrogram of the channel output. In non-Gaussian noise, a recent result of Poor and Thomas indicates that a modified data spectrogram correlator should be used, where the usual square-law envelope detector is replaced by a detector with a different power law. These results are compared to a Karhounen-Loève (K-L) detector implementation. The comparison indicates that under low SNR conditions, the spectrogram correlator is equivalent to the K-L detector, except that the spectrogram generally has more degrees of freedom. The spectrogram correlator may be preferable because:

(1) It does not require the solution of an eigenfunction equation;

(2) it can be configured as an iterative estimation device that adjusts to changes in the channel scattering function;

(3) it can be implemented with standard FFT techniques;

(4) the number of degrees of freedom can be reduced by transmitting a periodic signal; and

(5) more degrees of freedom allow for the implementation of a more sensitive mismatched correlator for maximization of signal-to-clutter ratio.

2.4.2 Introduction

In Section 2.2, it was shown that, if we are given the spectrogram of an observed time function, and if we know the noise free spectrogram that would be observed in the presence of a signal, then a locally optimum detector can be implemented by correlating samples of the data spectrogram with corresponding samples of the noise-free signal spectrogram. Since we were concerned with models of hearing, it was reasonable to assume that the data was indeed presented to the detector in the form of a spectrogram.

In this section, we are concerned with the class of detection problems for which it is optimum to describe the data in terms of its spectrogram. In other words, we no longer assume that we are given the data in the form of a spectrogram. In this section, we are given the data waveform itself and we are free to build a matched filter, a spectrogram correlator, or any other detection device. If the noise is additive and Gaussian, and if the noise-free signal is known except for time of arrival, the optimum detector is a whiten-and-match filter followed by a threshold, and spectrogram correlation is irrelevant. On the other hand, there may exist an important class of detection problems for which a spectrogram correlator is better than a matched filter.

The main result of this section is that there is indeed an important detection problem with a solution that can be structured in terms of a spectrogram correlation operation. This problem is to detect a deterministic or random signal that has been passed through a randomly time-varying channel, under low energy coherence (LEC) conditions. Low energy coherence means that the energy of the channel output is spread out over time and/or frequency, with relatively small SNR over any one time-frequency interval. This signal description is typical of sonar echo data from range and/or Doppler distributed targets, from clutter or reverberation, and from propagation through a time-varying, multipath channel.

The analysis will begin by obtaining an intuitively pleasing result, viz., the expected spectrogram of the channel output is the convolution, in time and frequency, of the channel scattering function[1] with the expected spectrogram of the channel input. The input spectrogram is thus "smeared" by the scattering function, which describes the expected delay and Doppler variation that is introduced by the channel. If a channel can be described in terms of its scattering function, it follows that the expected noise-free spectrogram of the channel output can be determined from knowledge of the scattering function and the expected spectrogram of the input signal. If the scattering function is itself unknown or slowly time varying, then the scattering function at any given time can be estimated from a sequence of previous spectrogram measurements.

A spectrogram representation therefore appears to be a convenient and natural way to describe the response of a time-varying, random channel. We have already shown in Section 2.2 that, under LEC conditions, the locally optimum detector for a signal with known spectrogram in white, Gaussian noise is a spectrogram correlator.

Poor and Thomas[2] have recently generalized this result to include non-Gaussian noise. If the noise that accompanies the data is not Gaussian, a different power law is used for envelope detection, rather than the usual square-law device, but the spectrogram correlator configuration remains valid.

Another way to describe the channel output is in terms of a Karhounen-Loève (K-L) expansion, i.e., in terms of the eigenfunctions of the signal covariance matrix. The locally optimum detector for the K-L representation will be compared to the spectrogram correlator, and some advantages of the spectrogram representation will become apparent.

Signal and filter design will also be considered from the viewpoint of spectrogram correlation. Price's criterion[3] for LEC signal design can easily be obtained from the spectrogram formulation. The use of a spectrogram allows the receiver to manipulate the filter or time window that is used to construct the spectrogram, as an alternative to adjusting the transmitted signal.

Finally, results will be described in terms of the usual bank of matched filters that is used to detect a narrowband signal with unknown Doppler shift. Such a filter bank can be used to construct a special kind of spectrogram, and spectrogram correlation can be applied to the envelope detected filter responses. A receiver that is designed for detection (in white, Gaussian noise) of a signal that is exactly known except for time of arrival and frequency can easily be reconfigured to detect the random output of a time-varying channel in non-Gaussian noise.

2.4.3 Scattering Functions and Spectrograms

Bello[1] has defined a symmetrized expected ambiguity function $\psi_{uu}(\tau,\varphi)$ as follows:

$$\psi_{uu}(\tau,\phi) = \int_{-\infty}^{\infty} E[u^*(t-\tau/2)\, u(t+\tau/2)] \exp(-j2\pi\phi t)dt. \tag{1}$$

From Ackroyd[4] the frequency-window spectrogram of a signal u(t) is

$$S_{uv}(t_1, f_1) = \left| \int_{-\infty}^{\infty} U(f)\, V(f-f_1) \exp(j2\pi f t_1)\, df \right|^2$$

$$= \left| \int_{-\infty}^{\infty} u(t)\, v(t_1-t) \exp(-j2\pi f_1 t)\, dt \right|^2. \tag{2}$$

It is shown in the Appendix (Section 2.4.9) that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[S_{uv}(t,f)] e^{-j2\pi(\phi t-\tau f)} dtdf = \psi_{uu}(\tau,\phi)\, \psi_{vv}(-\tau,\phi). \tag{3}$$

In other words, the Fourier transform of a spectrogram is the product of the signal ambiguity function and a filter ambiguity function, or

$$F_1\left\{E[S_{uv}(t,f)];\tau,\phi\right\} = \psi_{uu}(\tau,\phi)\, \psi_{vv}(-\tau,\phi) \tag{4}$$

and

$$F_1\left\{E[S_{zv}(t,f)];\tau,\phi\right\} = \psi_{zz}(\tau,\phi)\, \psi_{vv}(-\tau,\phi) \tag{5}$$

where $F_1\{\circ\}$ is a two dimensional Fourier transform as in (3) and $\psi_{zz}(\tau,\phi)$ is the ambiguity function of a channel output signal, z(t), in the absence of noise.

A fundamental input-output relation for time varying random channels is[1,9]

$$\psi_{zz}(\tau,\phi) = R_H(\tau,\phi)\, \psi_{uu}(\tau,\phi), \tag{6}$$

where $R_H(\tau, \phi)$ is the time frequency channel autocorrelation function

$$R_H(\tau, \phi) = E\left\{H^*(t, f)\, H(t+\tau, f+\phi)\right\} \tag{7}$$

and where $H(t, f)$ is the time-varying transfer function of the channel. Bello[1] has defined the scattering function as

$$\begin{aligned}\bar{S}(t, f) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_H(\tau, \phi) e^{j2\pi(t\phi - f\tau)} d\tau d\phi \\ &= F_1^{-1}\left\{R_H(\tau, \phi); t, f\right\}. \end{aligned} \tag{8}$$

From (6), we have

$$\psi_{zz}(\tau, \phi)\, \psi_{vv}(-\tau, \phi) = R_H(\tau, \phi)\, \psi_{uu}(\tau, \phi)\, \psi_{vv}(-\tau, \phi). \tag{9}$$

Taking the inverse Fourier transform of both sides of (9) we have

$$E\left\{S_{zv}(t, f)\right\} = \bar{S}(t, f) * E\left\{S_{uv}(t, f)\right\} \tag{10}$$

where $(*)$ denotes a two-dimensional convolution operation in time and frequency.

Equation (10) says that the expected output spectrogram is the two-dimensional convolution of the medium scattering function and the expected input spectrogram. The scattering function therefore determines the extent to which a given part of the signal in time and frequency coordinates is smeared by the channel. This channel-induced smear can be observed in practice by computing the frequency window spectrogram of the channel response. Alternatively, one can compute the time window spectrogram[4]

$$E\left\{S_{zw}^{T}(t_1, f_1)\right\} = E\left\{\left|\int_{-\infty}^{\infty} z(t)\, w(t-t_1) \exp(-j2\pi f_1 t) dt\right|^2\right\}. \tag{11}$$

By comparing (11) and (2), we see that

$$S^{T}_{zw}(t_1, f_1) = S_{zv}(t_1, f_1) \tag{12}$$

if

$$w(t) = v(-t). \tag{13}$$

The time window and frequency window spectrograms are identical if the time window is appropriately defined.

2.4.4 Detection Via Spectrogram Correlation

Under hypothesis H_1, we have a random signal with known expected spectrogram, added to white, Gaussian noise. Under H_0, we have noise alone. In Section 2.2, it was shown that statistically uncorrelated samples of the spectrogram are spaced $(4B_v)^{-1}$ seconds apart in the time direction and $(4T_v)^{-1}$ Hz apart in the frequency direction, where the filter V(f) has bandwidth B_v and impulse response duration T_v. For Gaussian input noise, the sampled spectrogram

$$z_{ij} = S_{rv}(t_i, f_j) \; ; \quad r(t) = z(t) + \text{noise} \tag{14}$$

has probability density function

$$p(z_{ij} | s_{ij}, H_1) = \sigma^{-2} \exp[-(z_{ij} + s_{ij})/\sigma^2] \, I_0(2\sqrt{z_{ij} s_{ij}}/\sigma^2) \tag{15}$$

when H_1 is true, where σ^2 is the average noise power and

$$s_{ij} = E\left\{S_{zv}(t_i, f_j)\right\}; \quad z(t) = \text{noise-free signal.} \tag{16}$$

If the signal is multiplied by a parameter A, then s_{ij} is multiplied by A^2. Replacing s_{ij} by $A^2 s_{ij}$ in (15), we have

$$p(z_{ij} | s_{ij}, H_1) = \sigma^{-2} \exp[-(z_{ij}/\sigma^2 + \theta s_{ij})] \, I_0(2\sqrt{z_{ij} \theta s_{ij}}/\sigma) \tag{17}$$

where

$$\theta = A^2/\sigma^2 \tag{18}$$

is a signal-to-noise ratio parameter.

When H_0 is true, we have

$$p(z_{ij}|H_0) = \sigma^{-2}\exp(-z_{ij}/\sigma^2). \tag{19}$$

The likelihood ratio is

$$L_\theta(z) = \prod_{i=1}^{N}\prod_{j=1}^{M} \exp(-\theta s_{ij}) I_o(2\theta^{1/2}\sqrt{z_{ij}s_{ij}}/\sigma). \tag{20}$$

The locally optimum test[2, 5, 6] for H_1 vs. H_0 compares $dL_\theta(z)/d\theta$ for $\theta = 0$ with a threshold. Substituting the standard series representation for the modified Bessel function[7] into (20), we have

$$\left.\frac{dL_\theta(\underline{z})}{d\theta}\right|_{\theta=0} = \sum_{i=1}^{N}\sum_{j=1}^{M}\left\{-s_{ij} + \frac{d}{d\theta}\left[\sum_{k=0}^{\infty}\frac{[\theta(z_{ij}s_{ij})/\sigma^2]^k}{(k!)^2}\right]\right|_{\theta=0}\Bigg\}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{M}(-s_{ij} + z_{ij}s_{ij}/\sigma^2). \tag{21}$$

The first term in the sum is independent of the data, z_{ij}. The data-dependent operation is the computation of

$$\ell_s(\underline{z}) = \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij}s_{ij} \tag{22}$$

which is a spectrogram correlation operation.

Poor and Thomas[2] have generalized the above result for an arbitrary noise p.d.f. f(x). Their result is equivalent to the statistic

$$\ell_s(\underline{z}) = \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij}\, g(x_{ij}) \qquad (23)$$

where x_{ij} is the output of the jth filter at time i and

$$g(x) \equiv f''(x) / f(x). \qquad (24)$$

In non-Gaussian noise, samples of a modified data spectrogram with a power law determined as in (24) are correlated with samples of an ordinary signal spectrogram defined by (16).

The locally optimum detector in (22-24) is ideal when each sample of the data spectrogram has small signal-to-noise ratio. If the samples s_{ij} in (22) are all nonzero, it would appear that the energy of the channel output has been "smeared out" over time and/or frequency. This situation is called a low energy coherence (LEC) condition.[3]

The locally optimum detector is not ideal for larger SNR, but for large SNR it is usually not critical to use the best possible detection procedure.[5]

2.4.5 K-L Representations Versus Spectrograms

The standard detector[8] for the output of a time-varying, random channel in the presence of Gaussian noise is obtained by using a Karkounen-Loève (K-L) signal representation. The K-L representation uses eigenfunctions of the signal covariance function $K_z(x, y)$ as basis functions for decomposition of the data. In terms of signal and channel representations,

$$K_z(x, y) = K_z(t+\tau/2,\ t-\tau/2) \text{ for } t = \frac{x+y}{2};\ \tau = x-y,$$

where

$$K_z(t+\tau/2,\ t-\tau/2) = \int_{-\infty}^{\infty} \psi_{zz}(\tau, \phi) \exp(j2\pi t\phi) d\phi$$

$$= \int_{-\infty}^{\infty} R_H(\tau, \phi)\, \psi_{uu}(\tau, \phi) \exp(j2\pi t\phi) d\phi . \qquad (25)$$

The eigenfunctions $\varphi_i(n)$, $i=1, 2, \ldots, N$ are such that

$$\int_0^T K_z(x, y)\, \varphi_i(y) dy = \lambda_i \varphi_i(x), \quad 0 \le x \le T, \qquad (26)$$

and the detection statistic is

$$\ell_{KL}(\underline{z}) = \sum_{i=1}^{N} \left(\frac{\lambda_i}{\lambda_i + N_o/2} \right) |x_i|^2$$

where $N_o/2$ is noise power spectral density and

$$|x_i|^2 = \left| \int_0^T r(t)\, \varphi_i(t) dt \right|^2 . \qquad (27)$$

In (27), $r(t)$ is the data signal, which is $z(t)$ + Gaussian noise under H_1, and the number $|x_i|^2$ is the envelope detected response of a filter with impulse response $\varphi_i(-t)$.

For the small SNR, LEC case,

$$\lambda_i << N_o/2 , \quad i=1, 2, \ldots, N,$$

and if the threshold is multiplied by $N_o/2$, we have

$$\ell_{KL}^{LEC}(\underline{z}) = \sum_{i=1}^{N} \lambda_i |x_i|^2 . \qquad (28)$$

In (28),

$$\lambda_i = \int_0^T \int_0^T \varphi_i^*(x) K_z(x, y) \varphi_i(y)\, dx\, dy$$

$$= E\left\{ \int_0^T \varphi_i^*(x)\, z^*(x)\, dx \int_0^T \varphi_i(y)\, z(y)\, dy \right\}$$

$$= E\left\{|x_i|^2\right\} \text{ in the absence of noise.} \tag{29}$$

The K-L detector for an LEC condition therefore correlates a sequence of envelope detected filter responses $|x_i|^2$ with the expected values of these responses for a noise-free channel output. The eigenvalues λ_i in (28) are thus equivalent to the noise-free, expected spectrogram samples s_{ij} in (22). The spectrogram analog to (27) is

$$\begin{aligned} z_{ij} &= S_{rv}(t_i, f_j) \\ &= \left|\int_{-\infty}^{\infty} r(t)\, v(t_i - t) \exp(-j2\pi f_j t)\, dt\right|^2. \end{aligned} \tag{30}$$

Equation (29) is identical to (26) if $v(t_i - t)\exp(-j2\pi f_j t)$ is an eigenfunction $\varphi_j(t)$ of $K_z(x, t)$ and if $v(t_i - t)$ has duration equal to T, the integration time in (26). Consider, for example, the special case where

$$K_z(x, y) = K_z(x - y) = K_z(\tau)$$

is periodic in τ with period T, i.e.,

$$\begin{aligned} K_z(\tau) &= \sum_{n=0}^{N} a_n \exp(-j2\pi n\tau/T) \\ &= \sum_{n=0}^{N} a_n \exp(-j2\pi nx/T) \exp(j2\pi ny/T). \end{aligned}$$

Defining

$$\varphi_n(t) = \begin{cases} T^{-1/2} \exp(-j2\pi n t/T), & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases}$$

we have

$$K_z(x, y) = T \sum_{n=0}^{N} a_n \varphi_n(x) \varphi_n^*(y) \qquad \text{for } 0 \le x \le T \text{ and } 0 \le y \le T.$$

Letting

$$v(-t) = \begin{cases} T^{-1/2} & , \quad 0 \le t \le T \\ 0 & , \quad \text{otherwise} \end{cases}$$

we have

$$\begin{aligned} z_{in} &= \left| T^{-1/2} \int_{-t_i}^{T-t_i} r(t) \exp(-j2\pi f_n t)\, dt \right|^2 \\ &= \left| T^{-1/2} \int_{0}^{T} r(t+t_i) \exp(-j2\pi f_n t)\, dt \right|^2 \\ &= \left| \int_{-\infty}^{\infty} r(t+t_i)\, \varphi_n(t)\, dt \right|^2 \end{aligned}$$

where

$$f_n \equiv n/T.$$

If $r(t)$ is periodic with period T, then $E\{z_{in}\}$ is independent of t_i and

$$s_{ij} = s_j$$

$$z_{ij} = z_j.$$

In this case

$$\begin{aligned} \ell_s(\underline{z}) &= \sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij} s_{ij} \\ &\propto \sum_{j=1}^{M} z_j s_j \\ &= \ell_{KL}^{LEC}(\underline{z}). \end{aligned} \tag{31}$$

A spectrogram correlator is thus identical to the usual K-L receiver under LEC conditions, if the basis functions that are used for spectrogram construction happen to be eigenfunctions of the signal covariance function $K_z(x, y)$. This situation occurs when the covariance function is a periodic function of (x-y). The periodicity causes the noise-free reference array s_{ij} to be independent of the timing parameter i, and the two dimensional time-frequency correlation in the

spectrogram detector becomes a one-dimensional time-independent correlation as in the K-L detector.

If the signal u(t) is periodic with period T, then $K_z(\tau)$ is also periodic with period T. This statement follows from the fact that $\psi_{uu}(\tau,\phi)$ is a "bed of nails" that are separated by T seconds in the τ-direction and by T^{-1}Hz in the ϕ-direction.[9] Therefore $\psi_{zz}(\tau,\phi)$ in (6) has the same form. The Fourier transform of $\psi_{zz}(\tau,\phi)$ with respect to the ϕ-variable in (25) is then a "bed of nails" that are separated by T seconds in both coordinates.

The special case in (31) should therefore be applicable to many practical communication systems. For a sonar with unchanging pulse repetition frequency (PRF), the gated data from a particular range interval will also be periodic, if the scattering function of the reflectors within that interval is stationary. If a range-extended target covers a sequence of such gated intervals, and if the target scattering function is range-dependent, then one must use a sequence of K-L detectors, one for each range bin. For a given, steady PRF = T^{-1}, each K-L detector is a bank of band-pass filters, and the sequence of K-L detectors for different gate positions on a range-extended target is a spectrogram correlator.

Implementations of a K-L, LEC detector and a spectrogram correlator are compared in Figs. 1 and 2. Both receivers consist of filter banks followed by envelope detectors, where the envelopes can be generalized for non-Gaussian noise as in (23) - (24). The most obvious difference between the two systems is that the K-L detector uses a single weighting parameter for each envelope detected filter response, while the spectrogram detector uses a sequence of weights that can be synthesized as a transversal filter.

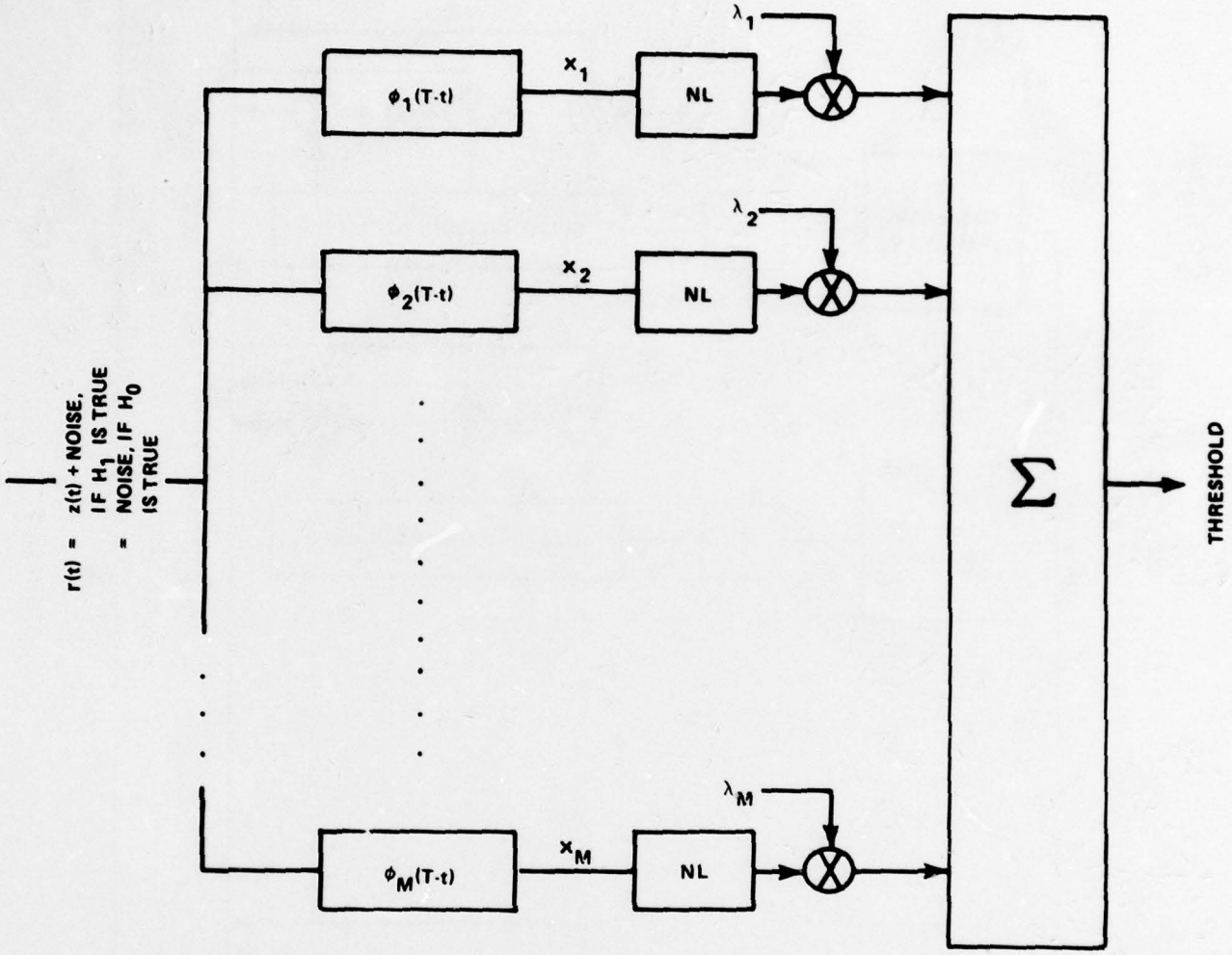


Figure 1. K-L detector for low SNR condition; $\lambda_m = E\{|x_m|^2\}$ when $r(t) = z(t)$, i.e., for a noise-free signal. NL = nonlinear device = square law for Gaussian noise.

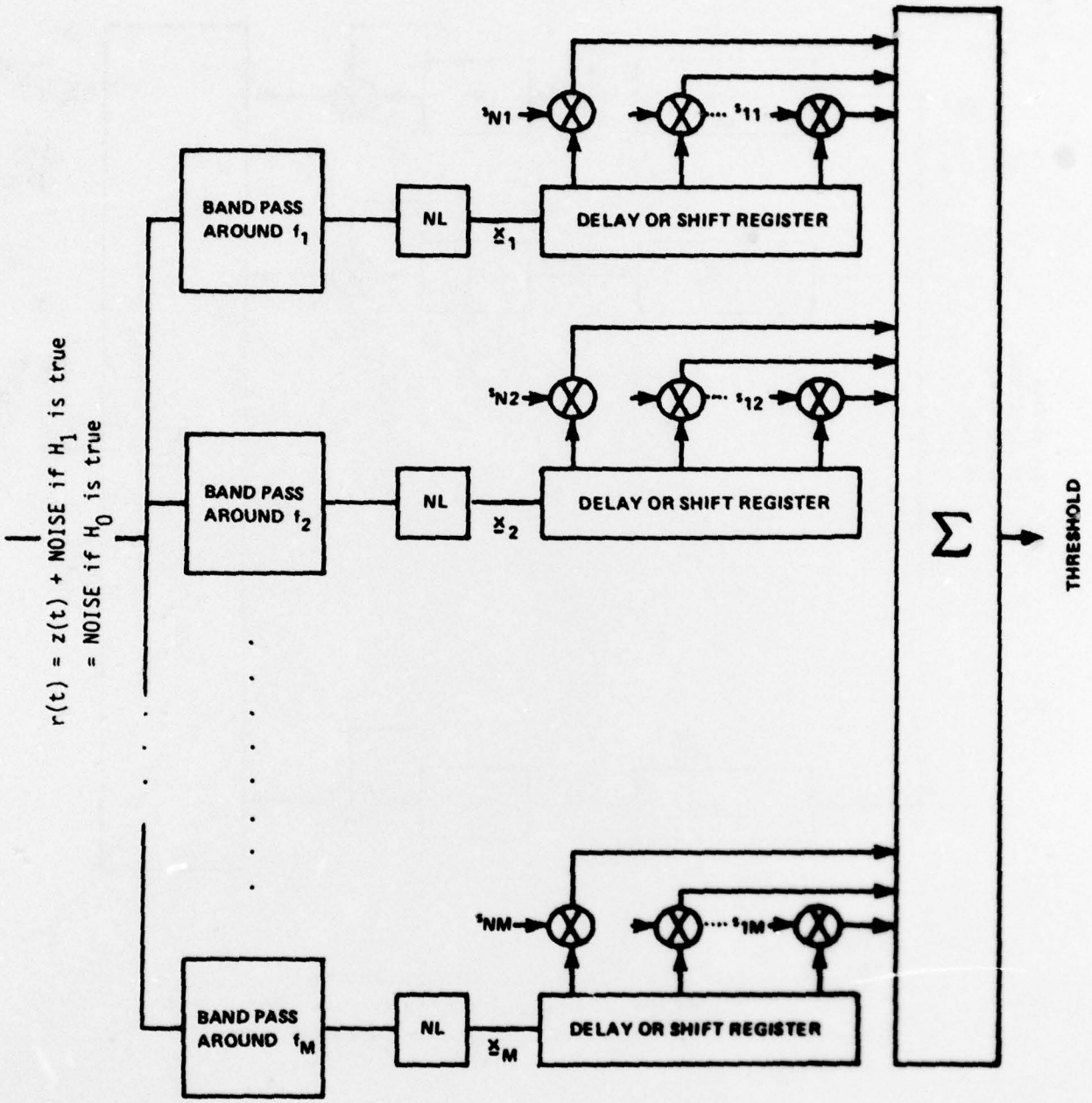


Figure 2. Spectrogram correlator; $s_{nm} = E\{|x_{nm}|^2\}$ when input = noise-free signal. NL = square law for Gaussian noise.

The K-L detector thus appears to possess the simplest implementation.

Another advantage of the K-L detector is associated with the relatively small number of weights in Fig. 1 and with the fact that a K-L expansion is equivalent to a principal component analysis. A principal component analysis provides the best (MMSE) representation of the data for any limited number of basis functions.[10] Let the first filter $\varphi_1(T-t)$ be associated with the largest eigenvalue λ_1, let $\varphi_2(T-t)$ correspond to the second largest eigenvalue, etc. If the eigenvalues happen to decrease rapidly, then much of the expected signal energy will be concentrated at the outputs of the first few filters, and comparison of Figs. 1 and 2 indicates that this energy will also be concentrated in time. The K-L system therefore tends to concentrate signal energy into a relatively small number of uncorrelated samples. Under low SNR, LEC conditions, this concentration of energy has the same intuitive appeal as a pulse compression process[9] for detection of a known signal with low instantaneous power.

A rigorous approach to the diversity vs. concentration problem has been formulated by Daly[11], who found that a noncoherent summation as in (22) and (28) gives the best detection performance if each term in the sum has SNR$\approx$2. The LEC condition corresponds to SNR$\ll$1 for each of the samples in Fig. 2. The best detection performance is then obtained by decreasing the number of uncorrelated samples and by concentrating signal energy, as in Fig. 1.

There are several reasons, however, why a spectrogram detector may be preferable:

1. The spectrogram correlator does not require a new solution of the eigenfunction equation (25) every time the scattering function changes.

2. Gradual changes in the scattering function can be estimated and "tracked" with the spectrogram configuration. For example, an unbiased estimate of s_{ij}, the expected squared response of the jth filter at time i in the absence of noise, is

$$\hat{s}_{ij}(n) = \frac{1}{K+1} \sum_{k=0}^{K} z_{ij}(n-K) - \sigma_j^2$$

$$= \frac{1}{K+1} \left\{ [z_{ij}(n) - \sigma_j^2] + K \hat{s}_{ij}(n-1) \right\} \qquad (32)$$

where $s_{ij}(n)$ is the estimate of s_{ij} for the nth observation of the signal, $z_{ij}(n)$ is the observed response of the jth filter at time i for the nth observation, and σ_j^2 is the average noise power at the filter output. Equation (32) is a simple form of recursive spectrogram estimate.

3. From a practical viewpoint, the spectrogram processor is more attractive than the K-L system because the receiver in Fig. 2 can be implemented with a sequence of fast Fourier transform operations. In particular, one can take advantage of (11) - (13), which imply that a time window spectrogram can be used as the basic element in the implementation of a spectrogram correlator.

4. If it is desirable to concentrate signal energy by limiting the number of uncorrelated samples or degrees of freedom, one can force the spectrogram correlator into a K-L configuration by transmitting a pulse train with a stable pulse repetition frequency T^{-1}, i.e., by using a periodic signal, and by using an integration time equal to T.

2.4.6 Signal Design for Maximization of Detector Signal-to-Noise Ratio and Signal-to-Interference Ratio

The expected response of the spectrogram correlator $\ell_s(\underline{z})$ in (22) to a noise-free signal is

$$\ell_s(\underline{s}) = \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij}^2 \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| E\left\{ S_{zv}(t,f) \right\} \right|^2 dt\, df \tag{33}$$

The expected response to noise, in the absence of signal, is

$$\ell_s(\text{noise}) = \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij} \propto \sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\left\{ S_{zv}(t,f) \right\} dt\, df \tag{34}$$

where σ^2 is the expected noise power at the output of each filter.

From Eq. (A8) in Section 2.2,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\left\{ S_{zv}(t,f) \right\} dt\, df = E_z E_v. \tag{35}$$

For a given signal energy E_z and filter energy E_v, SNR is maximized by making $\ell_s(\underline{s})$ as large as possible. From Parseval's theorem and (5) - (6)

$$\text{SNR} \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| R_H(\tau,\phi)\, \psi_{uu}(\tau,\phi)\, \psi_{vv}(-\tau,\phi) \right|^2 d\tau d\phi. \tag{36}$$

A spectrogram filter bank or window function is usually designed so that v(t) has minimum time-bandwidth product. The ambiguity function of v(t) is then generally broad and smooth relative to $\psi_{uu}(\tau,\phi)$, and

$$\text{SNR} \leq \max_{\tau,\phi} \left| \psi_{vv}(\tau,\phi) \right|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| R_H(\tau,\phi) \right|^2 \left| \psi_{uu}(\tau,\phi) \right|^2 d\tau\, d\phi \tag{37}$$

where the inequality is very nearly an equality. Since[9]

$$\max_{\tau,\phi} \left| \psi_{vv}(\tau,\phi) \right|^2 = \left| \psi_{vv}(0,0) \right|^2 = E_v^2 \equiv 1, \tag{38}$$

we have

$$SNR \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |R_H(\tau,\phi)|^2 \; |\psi_{uu}(\tau,\phi)|^2 \, d\tau d\phi \tag{39}$$

and SNR is maximized by maximizing the right band side of (39). This criterion for LEC signal design was first introduced by Price.[3] If spectrogram correlation is used and if one is willing to manipulate the spectrogram filter or window function, then $|\psi_{vv}(-\tau,\phi)|^2$ in (36) can be adjusted as an alternative to adjusting the transmitted signal.

One advantage of the spectrogram approach is that we can easily visualize the effects of cross-talk or interference in binary communication and the effects of clutter and reverberation with a specified scattering function.

To eliminate cross-talk, i.e., to maximize signal to interference ratio for a binary communications problem, we should have

$$\ell_s(\underline{s}^{(1)}, \underline{s}^{(2)}) = \sum_{i=1}^{N}\sum_{j=1}^{M} s_{ij}^{(1)} \, s_{ij}^{(2)} \equiv 0 \tag{40}$$

where $\left\{s_{ij}^{(1)}\right\}$ are samples of the expected noise-free spectrogram when signal $u_1(t)$ is transmitted and where $\left\{s_{ij}^{(2)}\right\}$ are samples of the noise-free spectrogram when $u_2(t)$ is transmitted. Again using Parseval's theorem, we have

$$\ell_s(\underline{s}^{(1)}, \underline{s}^{(2)}) \propto \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[S_{zv}^{(1)}(t,f)] \left\{E[S_{zv}^{(2)}(t,f)]\right\}^* dt df$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |R_H(\tau,\phi)|^2 \; |\psi_{vv}(-\tau,\phi)|^2 \; \psi_{u_1u_1}(\tau,\phi) \; \psi_{u_2u_2}^*(\tau,\phi) \, d\tau d\phi. \tag{41}$$

To minimize cross-talk, the inner product of the weighted ambiguity functions $R_H\psi_{vv}\psi_{u_1u_1}$ and $R_H\psi_{vv}\psi_{u_2u_2}$ should be minimized.

Another approach to the cross-talk problem is to use mismatched filtering in the spectrogram domain. Suppose that the output spectrograms for $u_1(t)$ and $u_2(t)$ are partially overlapping. A reasonable strategy is then to deaccentuate those reference samples $s_{ij}^{(1)}$ and $s_{ij}^{(2)}$ that occur in the region of expected overlap.

For a range-distributed sonar target with transfer function $H_1(t,f)$ and for reverberation or clutter with transfer function $H_2(t,f)$, the interference that is caused by clutter echoes can be written in the same form as in (40). The effect of clutter is then minimized when

$$\ell_s(\underline{s}^{(1)}, \underline{s}^{(2)}) \propto \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{H_1}(\tau,\phi)\, R_{H_2}{}^*(\tau,\phi)\, |\psi_{vv}(-\tau,\phi)\, \psi_{uu}(\tau,\phi)|^2\, d\tau d\phi$$

is minimized. The signal design problem for sonar is then to maximize the signal-to-clutter ratio

$$\text{SCR} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |R_{H_1}(\tau,\phi)|^2\, |\psi_{vv}(-\tau,\phi)\, \psi_{uu}(\tau,\phi)|^2\, d\tau d\phi}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{H_1}(\tau,\phi)\, R_{H_2}{}^*(\tau,\phi)\, |\psi_{vv}(-\tau,\phi)\, \psi_{uu}(\tau,\phi)|^2 d\tau d\phi} \tag{42}$$

by an appropriate choice of signal and filter functions. Again, we observe that the spectrogram formulation allows the receiver to increase SCR by manipulation of the spectrogram window function or filter transfer function, $v(t)$. The reference spectrogram samples $\left\{s_{ij}^{(1)}\right\}$ for the target can also be adjusted to deaccentuate the areas in time-frequency space where the expected target and clutter echoes overlap.

If the samples $\left\{s_{ij}^{(1)}\right\}$ of the expected noise-free target spectrogram are to be adjusted to deaccentuate the clutter response of the spectrogram correlator, then it may be advantageous to spread

out the expected target and clutter echoes in time-frequency space, so that the best weighting parameter can be used at each point. This observation suggests that the K-L detector in Figure 1 may not possess a large enough number of samples to effectively discriminate against spurious echoes, especially in a changing clutter environment.

2.4.7 Detection of a Random, Time-Varying, Range-Extended Target with a Bank of Matched Filters

A standard narrow band sonar receiver uses a bank of matched filters, where each filter is matched to a different Doppler-shifted version of the transmitted signal. This system is designed for the detection of a signal that is known except for Doppler shift and delay, in white Gaussian noise. The corresponding sonar reflector is an ideal point target with unknown, constant radial velocity.

The envelope detected matched filter responses as a function of delay and Doppler shift describe a magnitude-squared ambiguity function. The response of the filter bank can also be written in terms of a spectrogram. Suppose that the filter for a frequency window spectrogram has impulse response $v(t) = u^*(-t)$. From (2),

$$S_{uv}(t_1, f_1) = \left| \int_{-\infty}^{\infty} u(t)\, u^*(t-t_1)\, \exp(-j2\pi f_1 t)\, dt \right|^2$$

$$\equiv \left| \chi_{uu}(-t_1, f_1) \right|^2 \qquad (43)$$

where $\chi_{uu}(\tau, \phi)$ is a non-symmetrized ambiguity function. Using (10), the response of the filter bank to a range-extended, random, time-varying target can be written

$$E\left\{ S_{zv}(t, f) \right\} = \overline{S}(t, f) * \left| \chi_{uu}(-t, f) \right|^2 \qquad (44)$$

where $\bar{S}(t, f)$ is the channel scattering function and $v(t) = u^*(-t)$. For a motionless point target, $\bar{S}(t, f) = \delta(t)\,\delta(f)$ and $E\left\{S_{zv}(t, f)\right\} = |\chi_{uu}(-t, f)|^2$.

Since (44) describes a special kind of spectrogram, the spectrogram correlation process in (22) and Fig. 2 can be applied. Locally optimum detection with a bank of matched filters for a random time-varying, range-extended target is then implemented by forming a weighted sum of sampled, envelope detected filter responses. The weighting parameters take account of the time-frequency "smear" that is introduced by the scattering function of the target and the propagation medium. If a target echo consists of a sequence of separate highlights, for example, then the receiver attempts to add the envelope detected matched filter responses from the various highlights in order to improve detection performance.

For clutter rejection, the formulation in (40) is still applicable, i.e., the weights s_{ij} should be adjusted so that the spectrogram correlator response to the target is large relative to the clutter response.

2.4.8 Conclusion

A spectrogram correlator is a locally optimum detector for signals that have been passed through a random, time-varying channel. The standard (Karkounen-Loève) detector for such signals forms a weighted sum of the envelope detected outputs of a customized set of filters that are determined from a channel description by means of an eigenfunction equation. The spectrogram correlator is easier to implement because it uses standard FFT

processing and there is no need to solve an eigenfunction equation. The spectrogram correlator can also be used to estimate changes in the channel, and it can easily be updated to account for these changes.

The standard K-L detector tends to concentrate signal energy into relatively few sample points, and this energy concentration appears to be advantageous under low SNR conditions. The use of a periodic signal (constant PRF in sonar) appears to have the same concentration effect upon the spectrogram detector, and in fact leads to a K-L detector implementation that is the same as a spectrogram correlator. For interference rejection, a spectrogram representation with more degrees of freedom may be preferable.

Both the standard K-L detector and the spectrogram correlator can be adapted to non-Gaussian noise environments by changing the power law of the envelope detection operation.[2]

The use of a spectrogram correlator leads to an especially clear intuitive picture of signal and receiver requirements for the elimination of cross-talk and for clutter rejection. In addition to adjusting the signal, one can adjust the time window or filter function that is used to construct the spectrogram. The reference samples $\{s_{ij}\}$ can also be adjusted to increase the output signal-to-clutter ratio.

The response of a bank-of-matched-filters sonar receiver can be viewed as a special kind of spectrogram. This interpretation provides insight into the evolution of a receiver when the detection problem changes from (1) a known signal in Gaussian noise

to (2) a signal with unknown parameters in non-Gaussian noise to (3) a random signal in non-Gaussian noise. The transition to a signal with unknown parameters leads to a bank of matched filters instead of just one, and the possibility of non-Gaussian noise results in an adjustable power-law device at the output of each filter. Finally, the transition to a random signal results in a detector that computes a weighted sum of sampled, envelope detected filter responses. In this case, the channel is modelled as an array of point scatterers with different ranges and velocities.

2.4.9 References for Section 2.4

1. P. A. Bello, "Characterization of randomly time-variant linear channels," IEEE Trans. on Communications Systems, vol. CS-11, pp. 360-393, 1963.

2. H. V. Poor and J. B. Thomas, "Locally optimum detection of discrete-time stochastic signals in non-Gaussian noise," J. Acous. Soc. Amer., vol. 63, pp. 75-80, 1978.

3. R. Price, "Detectors for radar astronomy targets," in "Signal processing on radar astronomy - communication via fluctuating multipath media," by R. Price and P. E. Green, MIT Lincoln Lab. Tech. Report No. 234, 6 Oct. 1960. DDC No. AD 246 782.

4. M. H. Ackroyd, "Short-time spectra and time-frequency energy distributions, "J. Acous. Soc. Amer., vol. 50, pp. 1229-1231, 1971.

5. J. Capon, "On the asymptotic efficiency of locally optimum detectors," IRE Trans. Inform. Theory, IT-7, pp. 67-71, 1961.

6. D. Middleton, "Canonically optimum threshold detection," IEEE Trans. on Inform. Theory, vol. IT-12, pp. 230-243, 1966.

7. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, NBS AMS 55, Washington, D. C., U. S. Government Printing Office, p. 375.

8. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part III. New York: Wiley, 1971.

9. C. E. Cook and M. Bernfeld, Radar Signals. New York: Academic, 1967.

10. Y. T. Chien and K. S. Fu, "Selection and ordering of feature observations in a pattern recognition system," Inform. and Control, Vol. 12, pp. 395-414, 1968.

11. R. F. Daly, "Signal design for efficient detection in dispersive channels," IEEE Trans. on Inform. Theory, vol. IT-16, pp. 206-213, 1970.

2.4.10 Appendix to Section 2.4

This appendix provides a proof of Equation (3), which reads

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[\,S_{uv}(t,f)\,]\, e^{-j2\pi(\phi t-\tau f)}\,dt\,df = \psi_{uu}(\tau,\phi)\,\psi_{vv}(-\tau,\phi). \quad \text{(A1)}$$

where

$$\psi_{uu}(\tau,\phi) = \int_{-\infty}^{\infty} E[u^*(t-\tau/2)\,u(t+\tau/2)\,]\exp(-j2\pi\phi t)\,dt \quad \text{(A2)}$$

$$\psi_{vv}(-\tau,\phi) = \int_{-\infty}^{\infty} v^*(t+\tau/2)\,v(t-\tau/2)\exp(-j2\pi\phi t)\,dt. \quad \text{(A3)}$$

Eq. (A1) defines a two-dimensional Fourier transform operation, and the inverse transform is

$$\iint \psi_{uu}(\tau,\phi)\,\psi_{vv}(-\tau,\phi)\, e^{j2\pi(\phi t-\tau f)}\,d\tau\,d\phi$$

$$= \iiiint e^{-j2\pi\tau f} E[u^*(x-\tau/2)\,u(x+\tau/2)\,]\,v^*(y+\tau/2)\,v(y-\tau/2)$$

$$\left\{\exp\{-j2\pi\phi[y-(t-x)]\}\,d\phi\right\}dy\,dx\,d\tau$$

$$= \iiint e^{-j2\pi\tau f} E[u^*(x-\tau/2)\, u(x+\tau/2)\, v^*(y+\tau/2)\, v(y-\tau/2)$$

$$\delta[y-(t-x)]\, dy\, dx\, d\tau$$

$$= \iint e^{-j2\pi\tau f} E[u^*(x-\tau/2)\, u(x+\tau/2)]\, v^*(t-x+\tau/2)\, v(t-x-\tau/2)\, dx\, d\tau. \tag{A4}$$

Letting

$$\begin{bmatrix} x_1 \\ \tau_1 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 \\ 1 & -1/2 \end{bmatrix} \begin{bmatrix} x \\ \tau \end{bmatrix}, \quad \begin{bmatrix} x \\ \tau \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ \tau_1 \end{bmatrix} \tag{A5}$$

(A4) becomes

$$\iint e^{-j2\pi f(x_1-\tau_1)} E[u^*(\tau_1)\, u(x_1)]\, v^*(t-\tau_1)\, v(t-x_1)\, dx_1 d\tau_1$$

$$= E\left[\left| \int_{-\infty}^{\infty} u(x_1)\, v(t-x_1)\, e^{-j2\pi f x_1}\, dx_1 \right|^2\right]$$

$$\equiv E[S_{uv}(t, f)] \tag{A6}$$

where $S_{uv}(t, f)$ is defined in (2). Eq. (A1) follows from (A4) - (A6).

2.5 Summary and Conclusions for Volume 2

From an engineering point of view, the most important result in this volume is that a spectrogram representation of a random signal contains the same information as the signal covariance function. A detector that uses a spectrogram representation is therefore equivalent to one that is based upon a description of the signal covariance function. The spectrogram detector, in fact, appears to be more robust in situations where (1) the signal covariance matrix is not known exactly, and (2) the background noise is not necessarily Gaussian.

From a biological modelling viewpoint, a spectrogram correlation process seems to explain human frequency discrimination data as well as the performance of humans in detecting sinusoids with harmonics.

Given the above observations, it is not surprising that spectrogram correlation, combined with principal component analysis, has out-performed some other processors (Volume 1). If mismatched filtering for maximization of signal-to-interference ratio had been applied, the clutter performance in Volume 1 could probably have been improved even further!