SP-3632

The Folding Quefrency of the Cepstrum

H. B. Ritea

5 January 1972

R

15

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* System Development Corporation Santa Monica, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

THE FOLDING QUEFRENCY OF THE CEPSTRUM

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical report - July, 1971 - December, 1971

5. AUTHOR(S) *(First name, middle initial, last name)*

H. B. Ritea

6. REPORT DATE 5 January 1972	7a. TOTAL NO. OF PAGES 13	7b. NO. OF REFS 2

8a. CONTRACT OR GRANT NO. DAHC15-67-C-0149	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. ARPA Order No. 1327, Amendment No. 4	
c. Program Code No. 2D30 and 2P10	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

It is shown that the cepstrum is symmetric about a quefrency, termed the folding quefrency. Various allowable combinations of number of digitized samples and sampling times are given, which yield valid cepstra up to 20 msec.

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

SP-3632

SP *a professional paper*

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-67-C-0149, ARPA Order No. 1327, Amendment No.4 , Program Code No. 2D30 and 2P10.

The Folding Quefrency of the Cepstrum

by

H. B. Ritea

January 5, 1972

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

SDC

ABSTRACT

It is shown that the cepstrum is symmetric about a quefrency, termed the folding quefrency. Various allowable combinations of number of digitized samples and sampling times are given, which yield valid cepstra up to 20 msec.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In 1966, Noll [1] presented a method for determining the pitch period of voiced speech based upon the cepstrum. Briefly, the cepstrum is the power spectrum of the logarithm of the power spectrum and has a strong peak at the pitch period of the voiced-speech segment being analyzed. It is shown here that just as the spectrum has a folding frequency about which the spectrum is symmetric, the cepstrum exhibits a similar behavior, and care must be exercised in selecting the proper number of sample points and sampling period for the calculation of valid cepstra.

2. THE CEPSTRUM

A simplified model for the production of voiced-speech sounds is shown in Figure 1.

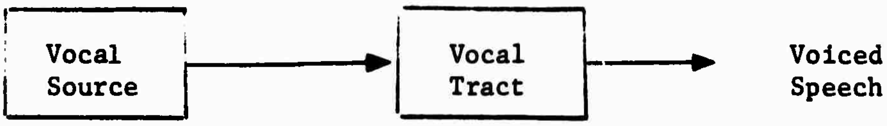

Figure 1. Model for the Production of Voiced Speech

The vocal source produces a signal g(t) consisting of periodic puffs of air that travel through the vocal cords. If h(t) is the impulse response of the vocal tract, then voiced speech, represented by f(t), is the convolution of of g and h:

$$f(t) = (g*h)(t)$$

Let F(ω), G(ω) and H(ω) denote the Fourier transforms of f(t), g(t), and h(t), respectively:

$$F(\omega) = \mathfrak{F}\left\{f(t)\right\} = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \qquad (i = \sqrt{-1})$$

$$G(\omega) = \mathfrak{F}\left\{g(t)\right\} \text{ and } H(\omega) = \mathfrak{F}\left\{h(t)\right\} .$$

Then, using the fact that the Fourier transform of a convolution is the product of the Fourier transforms, we have that

$$F(\omega) = G(\omega) \cdot H(\omega).$$

For voiced-speech sounds, g(t) is quasiperiodic; thus, f(t) is also quasiperiodic. Hence, if the period is P seconds, then the spectrum $|F(\omega)|^2$ of the speech signal consists of harmonics spaced $\frac{1}{P}$ Hz apart. This is illustrated in Figure 2, which is the spectrum of

$$f(t) = \sin 2\pi\omega t + .5\sin 4\pi\omega t + \sin 6\pi\omega t + .5\sin 10\pi\omega t,$$

where ω = 500 Hz. The "periodicity", occurring at 500 Hz intervals, is shown in the figure. Note that

$$500 \text{ Hz} = 500 \text{ cps} = \frac{1 \text{ cycle}}{2 \text{ msec.}},$$

so that P = 2 msec., as expected.

This periodicity in the spectrum can be obtained by taking the Fourier transform of the spectrum. We would then expect that the transformed spectrum would have a peak corresponding to the period. The transformed spectrum is

$$a(\tau) = \mathfrak{F}\left\{|F(\omega)|^2\right\},$$

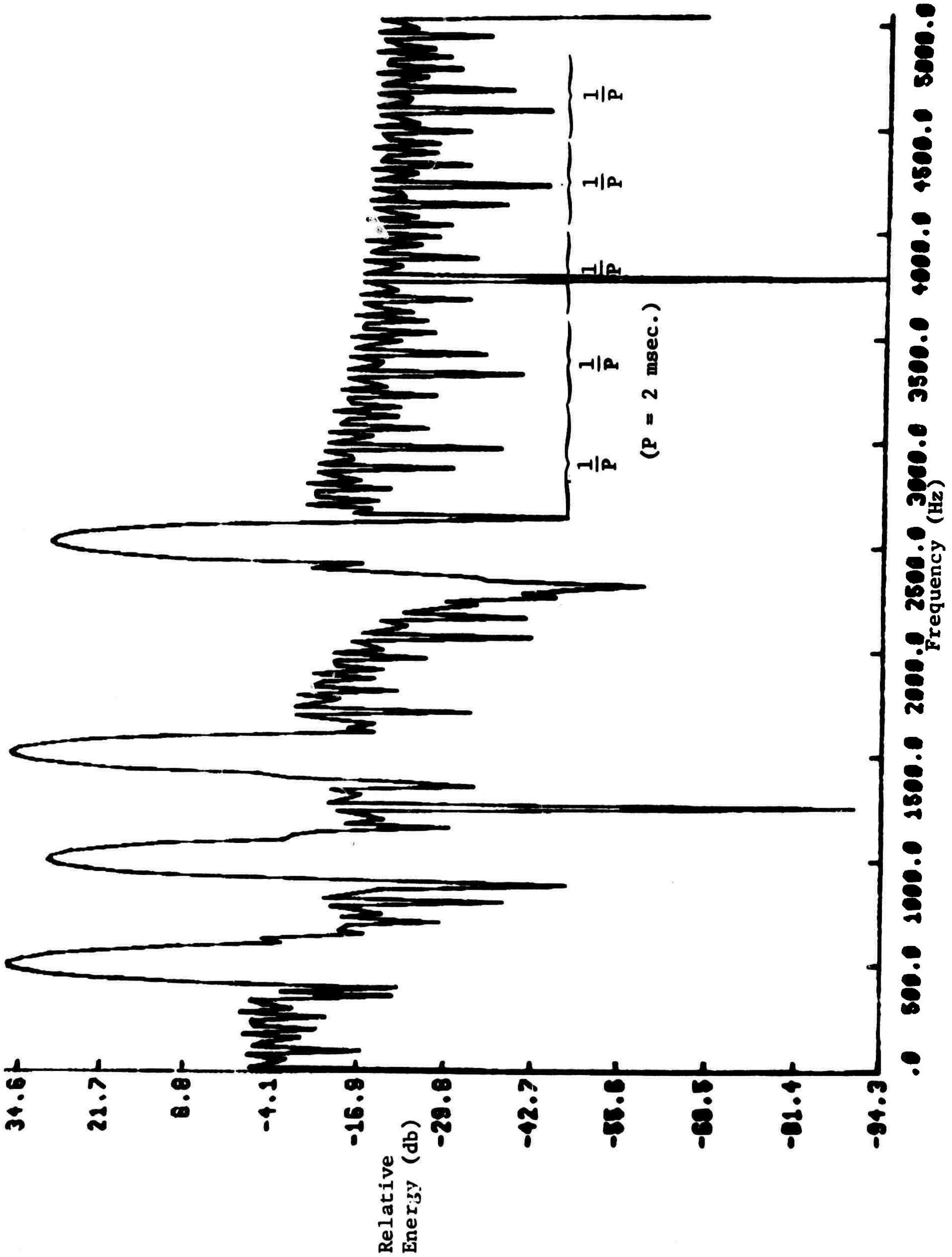

Figure 2. Spectral periodicity

which is the autocorrelation function of f(t) (see [2], pp. 241-242). Using the fact that

$$F(\omega) = G(\omega) \cdot H(\omega)$$

gives us

$$\begin{aligned} a(\tau) &= \mathfrak{F}\left\{|G(\omega)|^2 \cdot |H(\omega)|^2\right\} \\ &= \mathfrak{F}\left\{|G(\omega)|^2\right\} * \mathfrak{F}\left\{|H(\omega)|^2\right\} \\ &= a_g(\tau) * a_h(\tau), \end{aligned}$$

where * again denotes convolution and $a_g(\tau)$ and $a_h(\tau)$ are the autocorrelation functions of g(t) and h(t) respectively. Since the effects of g(t) and h(t) are convolved in the autocorrelation function $a(\tau)$, it is difficult to obtain a clear peak in $a(\tau)$ resulting from the periodicity of g(t), as desired. This problem is circumvented by taking the logarithm of the spectrum:

$$\begin{aligned} \log|F(\omega)|^2 &= \log[\,|G(\omega)|^2 \cdot |H(\omega)|^2] \\ &= \log|G(\omega)|^2 + \log|H(\omega)|^2 \end{aligned}$$

and then taking the Fourier transform of the result:

$$\mathfrak{F}\left\{\log|F(\omega)|^2\right\} = \mathfrak{F}\left\{\log|G(\omega)|^2\right\} + \mathfrak{F}\left\{\log|H(\omega)|^2\right\}.$$

The effects of g(t) and h(t) are now combined additively, rather than being convolved as in $a(\tau)$. In addition, the function c defined by

$$c(\tau) = \mathfrak{F}\left\{\log|F(\omega)|^2\right\}$$

has a strong peak corresponding to the effects of g(t) and a broader peak corresponding to h(t), since the effect of g(t) is to produce high-frequency ripples in the spectrum, while h(t) produces the low-frequency formant structure. If $c(\tau)$ is squared, $[c(\tau)]^2 = [\mathfrak{F}\left\{\log|F(\omega)|^2\right\}]^2$, then the peak in

$c(\tau)$ is made even more pronounced, and it is this last expression which is referred to as the "cepstrum". The variable τ in the cepstrum has units of cycles per hertz, or seconds. These units are commonly referred to as "quefrencies", and we shall use this terminology throughout. The cepstrum corresponding to the spectrum in Figure 2 is shown in Figure 3. Note the strong peak corresponding to the pitch period of 2 msec.

An alternate definition of the cepstrum is obtained by using the inverse Fourier transform after taking the logarithm:

$$c(\tau) = \mathfrak{F}^{-1}\left\{\log|F(\omega)|^{2}\right\}.$$

This definition is more computationally stable than the others in its discrete form, and it is the one we adopt.

3. THE CEPSTRAL FOLDING QUEFRENCY

For digital computation of the cepstrum, we need the notion of the discrete Fourier transform (DFT): if $x_0, x_1, \ldots x_{n-1}$ is a discrete time series, then the DFT of $x_0, x_1, \ldots x_{n-1}$ is

$$X_k = \sum_{n=0}^{N-1} x_n \exp(-2\pi ink/N)$$

for $k=0,1, \ldots, N-1$, where $i = \sqrt{-1}$. The inverse discrete Fourier transform (IDFT) of $X_0, X_1, \ldots, X_{N-1}$ is

$$x_n = \frac{1}{N} \sum_{k=0}^{n-1} X_k \exp(2\pi ink/N)$$

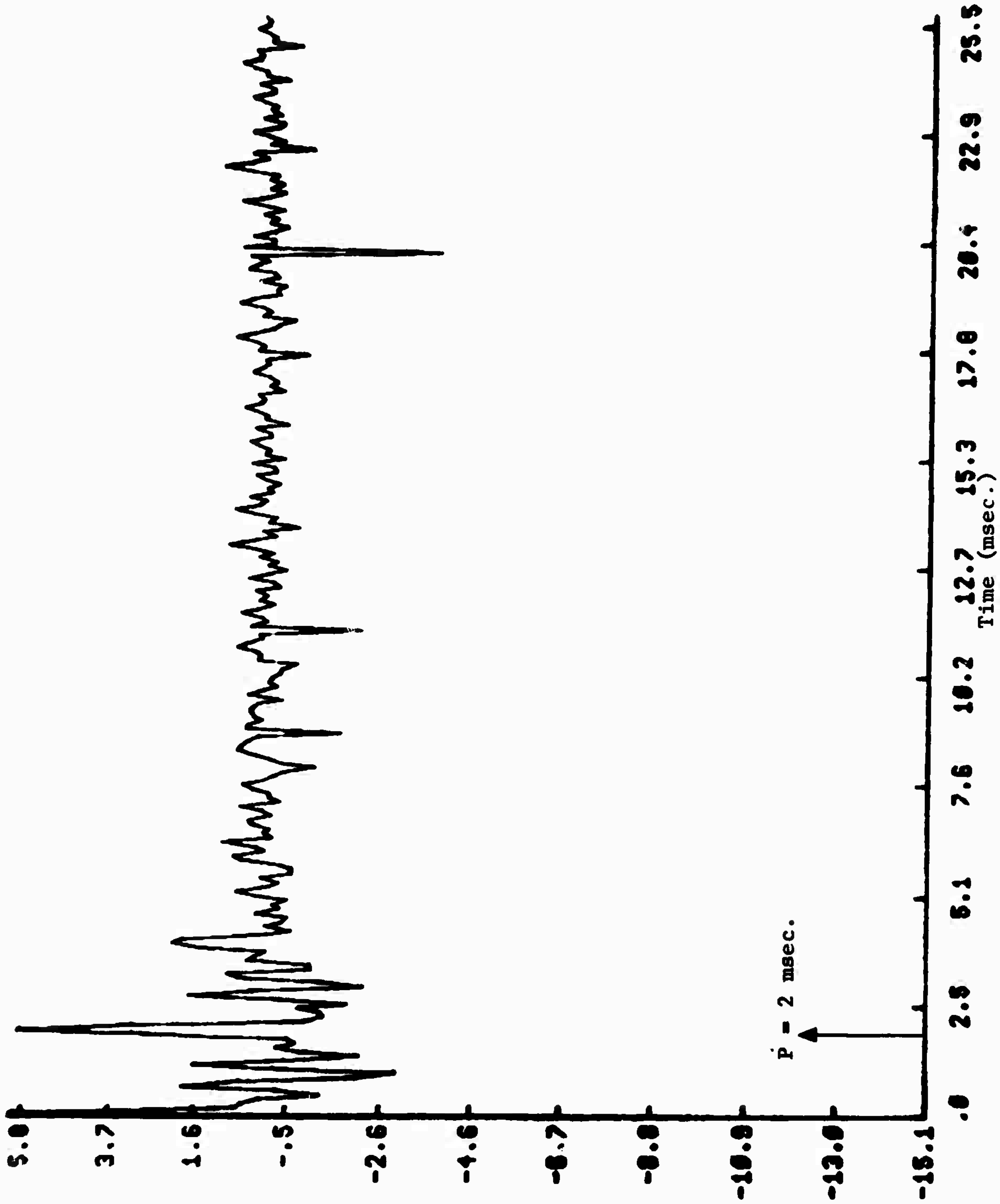

Figure 3. Short-time cepstrum

for n=0,1, ..., N-1. Thus,

$$x_n = \text{IDFT}\left\{\text{DFT}(x_n)\right\} \quad (n=0,1, \ldots,N-1).$$

Now suppose that a continuous voiced-speech signal is sampled every T seconds; e.g., for a sampling rate of 10,000 samples/sec., T = .0001. Assume also that N samples are obtained:

$$x_n = f(nT) \quad (n=0,1, \ldots,N-1)$$

Note that the Nyquist folding frequency is $\omega_f = \frac{1}{2T}$.

The spectrum of x_0, x_1, ..., x_{N-1} is

$$S_k = 20 \log_{10} |X_k| \quad (k=0,1, \ldots, N-1)$$

where

$$X_k = \sum_{n=0}^{N-1} x_n \exp(-2\pi ink/N)$$

and the constants are chosen to make S_k the relative energy level in db. The discrete cepstrum is now defined to be

$$\begin{aligned} c(nT) &= \text{Re}\left\{\text{IDFT}(S_k)\right\} \\ &= \text{Re}\left\{\frac{1}{N} \sum_{k=0}^{N-1} S_k \exp(2\pi ink/N\right\} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos(2\pi nk/N) \end{aligned}$$

for n=0,1, ... N-1. Thus, c(nT) is defined at the N discrete quefrencies 0, T, 2T, ..., (N-1)T. By the folding quefrency we mean the quefrency about which c(nT) is symmetric.

Theorem. The folding quefrency of the cepstrum c(nT) is $\frac{NT}{2}$.

Proof. By definition,

$$c(nT) = \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos(2\pi nk/N).$$

Thus,

$$c(\frac{NT}{2} + mT) = \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos[2\pi(\frac{N}{2} + m)k/N]$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos[\pi k + 2\pi mk/N]$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} S_k (-1)^k \cos(2\pi mk/N).$$

Also,

$$c(\frac{NT}{2} - mT) = \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos[2\pi(\frac{N}{2} - m)k/N]$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} S_k \cos[\pi k + 2\pi mk/n]$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} S_k (-1)^k \cos(2\pi mk/N).$$

$$= c(\frac{NT}{2} + mT),$$

and the result is proved.

In order to obtain usable cepstra, N and T must be selected so that the time interval from t = 0 to t = $\frac{NT}{2}$ is sufficiently large to encompass the pitch periods of most speakers. An interval from 0 to 20 msec is usually adequate, and Table 1 gives some appropriate combinations of N and T which meet this specification:

Table 1. Combinations of N & T for valid cepstra

N	Maximum number of Samples/sec
128	3200
256	6400
512	12,800
1024	25,600
2048	51,200

REFERENCES

1. A. M. Noll, "Cepstrum Pitch Determination," J. Acoust, Soc. Am., Vol. 41 (1967), pp. 293-309.

2. A. Papoulis, The Fourier Integral and its Applications, McGraw-Hill, New York (1962).