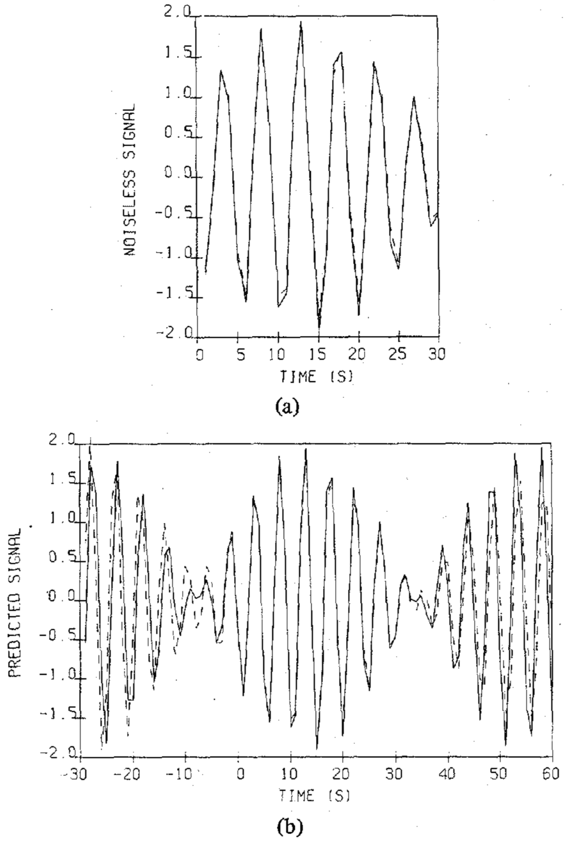

(a)

(b)

Fig. 3. (a) A comparison of the actual noiseless input signal (full curve) and the signal reconstructed from the computed Fourier coefficients (broken curve). (b) A similar comparison to that in Fig. 3(a) except that a 90-point version is shown illustrating backward and forward extrapolation. The input signal is contained between 0 and 29 s.

The noise variance was estimated to be $\sigma_n^2 = 0.19$ from the periodogram. The actual noiseless input signal and the version reconstructed from the calculated Fourier coefficients are shown in Fig. 3(a). Fig. 3(b) is a comparison of a longer 90-point noiseless input signal and the corresponding reconstructed signal. In spite of some amplitude and phase discrepancy, the correspondence is generally quite good. The reconstructed signal shown in Fig. 3(b) can now be used in the same manner as the forward and backward predicted signal in [7].

CONCLUSIONS

We have presented in this note a new approach to the determination of the power spectrum of a harmonic process. The philosophy behind this method is quite different from the one which underlies the ME method. Whereas the ME development looks for an extension of the measured or estimated autocorrelation, which is most uncommitted with respect to the data outside the known interval, the LIP approach constrains the power spectrum to have a spiky structure. It appears that this constraint allows a well-resolved spectral estimate to be made even in cases where other high resolution estimators encounter problems.

It is hard to match the ME formulation in its elegance and simplicity, and the LIP method presented here is not meant as a substitute. Where the greatest possible resolution is a prime requirement however, as, for example, in the determination of the splitting parameters associated with the free oscillation of the earth [8], the LIP spectral estimator should prove to be a very useful one.

REFERENCES

[1] S. Haykin and S. Kesler, "Prediction-error filtering and maximum-entropy spectral estimation," *Topics in Applied Physics*, vol. 34, *Nonlinear Methods of Spectral Analysis*. 1979, ch. 2, pp. 9-72.

[2] T. J. Ulrych and M. Ooe, "Autoregressive and mixed autoregressive-moving average models and spectra," *Topics in Applied Physics*, vol. 34, *Nonlinear Methods of Spectral Analysis*. 1979, ch. 3, pp. 73-125.

[3] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," *Phys. Earth Planet. Int.*, vol. 12, pp. 188-200, 1976.

[4] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661-675, 1971.

[5] S. J. Johnsen and N. Andersen, "On power estimation in maximum entropy spectral analysis," *Geophysics*, vol. 43, pp. 681-690, 1978.

[6] S. Levy and P. K. Fullagar, "Reconstruction of a sparse spike train from a portion of its spectrum, and application to high resolution deconvolution," *Geophysics*, 1981, accepted for publication.

[7] T. J. Ulrych, D. E. Smylie, O. G. Jensen, and G. K. C. Clarke, "Predictive filtering and smoothing of short records using maximum entropy," *J. Geophys. Res.*, vol. 78, pp. 4959-4964, 1973.

[8] B. F. Chao and F. Gilbert, "Autoregressive estimation of complex eigenfrequencies in low frequency seismic spectra," *Geophys. J. R. Astr. Soc.*, vol. 63, pp. 641-657, 1980.

The Unimportance of Phase in Speech Enhancement

DAVID L. WANG AND JAE S. LIM

Abstract—The importance of Fourier transform phase in speech enhancement is considered. Results indicate that a more accurate estimation of phase is unwarranted in speech enhancement at the *S/N* ratios where the intelligibility scores of unprocessed speech range from 5 to 95 percent, if the phase estimate is used to reconstruct speech by combining it with an independently estimated magnitude or to reconstruct speech using the phase-only signal reconstruction algorithm.

I. INTRODUCTION

There are many situations in which enhancement of speech degraded by additive noise is desirable. Most of the systems proposed so far are based on the estimation of the short-time spectral magnitude from the degraded speech and its recombination with the phase of the degraded speech. This includes spectral subtraction [1] and parametric Wiener filtering [1]. The common assumption in these systems is that short-time phase is relatively unimportant. Hence, these systems focus on estimating the short-term spectral magnitudes of the speech signal more accurately and tend to ignore the issue of accurate phase estimation

It is well established that the ear does not have any preference among changes in the phase of sinusoidal signals or changes in the relative phase in the sinusoidal components of a signal. However, the work by Weiss *et al.* [2] indicates that rapid fluctuations in the relative phases in the sinusoidal components of a speech signal lead to significant degradation in speech

Manuscript received May 22, 1981; revised November 18, 1981.

The authors are with the Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

0096-3518/82/0800-0679$00.75 © 1982 IEEE

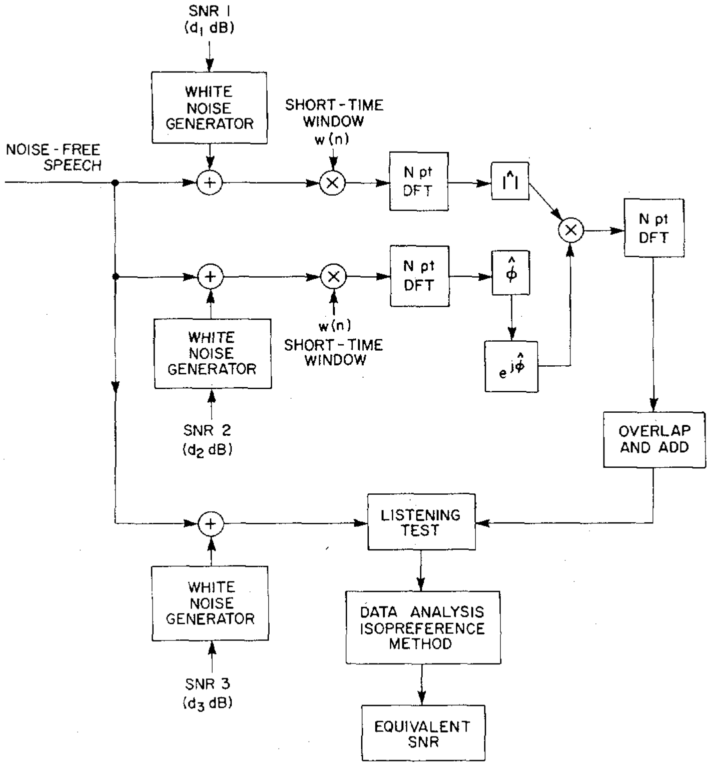

Fig. 1. Block diagram of experimental paradigm.

quality. In addition, recent work by Oppenheim and Lim [3] shows that phase contains a great deal of information in a signal.

To understand if a more accurate estimation of phase would be helpful in the context of speech enhancement, a series of experiments have been performed. The purpose of this correspondence is to report the results of these experiments.

II. Experimental Procedures and Results

Noisy speech is generated by adding white Gaussian noise to noise-free speech. Specifically, the original noise-free speech was low-pass filtered at 4.7 kHz and sampled at a rate of 10 kHz. Then digitally generated white Gaussian noise was added to the noise-free speech. Each noise sample was statistically independent of all other samples. The amplitude of each noise sample was scaled to obtain the specified signal-to-noise ratio. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR(dB)} = 10 \log \frac{\sum_n s^2(n)}{\sum_n d^2(n)} \qquad (1)$$

where $s(n)$ is the original speech waveform, $d(n)$ is the white Gaussian noise sequence, and the summation is over the entire length of the speech sentence.

Each test sentence is processed segment by segment where the length of each data segment N is fixed throughout the processing and is varied as an experimental parameter. Each data segment is Hanning windowed before processing and a new data segment is obtained by shifting the window $N/2$ points after processing. Thus, there is a 50 percent overlap between two consecutive data segments and the processed segments are overlapped and added to obtain a test sentence.

For each data segment, the processing consists of obtaining the Fourier transform magnitude of noisy speech $y_1(n)$ and the phase of noisy speech $y_2(n)$, and then reconstructing a test sentence $x(n)$ by combining the magnitude and phase. Specifically, a test sentence $x(n)$ is obtained by

$$x(n) = FT^{-1} [|Y_1(\omega)| e^{j\vartheta_{y_2}(\omega)}] \qquad (2)$$

where $y_1(n)$ and $y_2(n)$ are noisy speech at SNR's of d_1 and d_2 dB, respectively. FT^{-1} represents the inverse discrete time Fourier transform [4], $|Y_1(\omega)|$ represents the Fourier transform magnitude of $y_1(n)$, and $\vartheta_{y_2}(\omega)$ represents the Fourier transform phase of $y_2(n)$.

In each trial of the experiment, unprocessed noisy speech of known SNR and a test sentence $x(n)$ reconstructed in the manner discussed above were presented in a soundproofed room through earphones. Listeners were asked to choose the sentence which sounds "better in its quality." By changing the SNR of unprocessed noisy speech in different trials, it is possible to compute the equivalent SNR at which the reconstructed speech is selected 50 percent and the unprocessed noisy speech is selected 50 percent of the time. The equivalent SNR thus obtained will quantify the relative importance of phase and magnitude components in the context of speech enhancement. For example, by keeping d_1 [SNR of $y_1(n)$] fixed and only varying d_2 [SNR of $y_2(n)$], the importance of phase in the context of speech enhancement can be examined. The generation of test materials and experimental procedures are graphically illustrated in Fig. 1.

The experiment consisted of six sessions, each of which consisted of 30 trials. A total of six different noise-free sentences were used and nine listeners participated in the experiments. From the results of the experiment, the equivalent SNR was obtained for various different choices of N (length of Hanning window). Tables I, II, and III show the equivalent SNR's for $N = 64$, 512, and 4096, respectively. Each equivalent SNR in the tables, except for the diagonal entries, is based on the results of 96–108 trials. The equivalent SNR's in the diagonal entries are the same as d_1 or d_2.

TABLE I
EQUIVALENT SNR BASED ON 64-POINT DFT AND HANNING WINDOW

	EQUIVALENT SNR (dB)		
MAG (dB) PHASE (dB)	-5.0	5.0	15.0
-5.0	-5.0	—	—
5.0	-3.6	5.0	—
15.0	-4.2	5.7	15.0

TABLE II
EQUIVALENT SNR BASED ON 512-POINT DFT AND HANNING WINDOW

	EQUIVALENT SNR (dB)				
MAG (dB) PHASE (dB)	-25.0	-5.0	5.0	15.0	25.0
-25.0	-25.0	—	—	—	—
-5.0	-25.0	-5.0	—	—	—
5.0	-25.0	-3.9	5.0	—	—
15.0	-25.0	-3.9	4.9	15.0	—
25.0	-25.0	-4.0	6.0	16.0	25.0

TABLE III
EQUIVALENT SNR BASED ON 4096-POINT DFT AND HANNING WINDOW

	EQUIVALENT SNR (dB)				
MAG (dB) PHASE (dB)	-25.0	-5.0	5.0	15.0	25.0
-25.0	-25.0	—	—	—	—
-5.0	-17.0	-5.0	—	—	—
5.0	-14.1	-3.1	5.0	—	—
15.0	-13.2	-2.3	5.6	15.0	—
25.0	-11.2	-3.2	5.5	16.0	25.0

III. DISCUSSIONS AND CONCLUSIONS

The results in Tables I, II, and III show that except for the case when $N = 4096$ and very low SNR's for the magnitude components (very low d_2), the equivalent SNR's do not improve significantly when a more accurate estimate of the phase is available. For example, for $N = 512$ (Table II) and $d_1 = -5$ dB, an increase of d_2 by 30 dB results in the improvement of the equivalent SNR by only 1 dB. When $N = 4096$ (Table III) and $d_1 = -25$ dB, an increase of d_2 by 30 dB results in the increase of the equivalent SNR by 11 dB. This is consistent with the results of previous study by Oppenheim and Lim [3]. For such a low SNR, however, a more accurate estimation of phase will be an extremely difficult task.

In the experiments discussed above, the magnitude was obtained from the noisy speech at d_1 dB. In another set of experiments, we have used the magnitude estimated from the noisy speech at SNR of d_1 dB using a spectral subtraction speech enhancement system [1]. Again, an increase in d_2 dB for fixed d_1 resulted in little improvement in the equivalent SNR.

Tables I, II, and III show only the results for $d_2 \geqslant d_1$, since any attempt to estimate the phase more accurately in the context of speech enhancement makes sense only when such an attempt in fact improves the phase estimate over the degraded phase. We have carried out similar experiments, however, for $d_2 \leqslant d_1$. In this case, we have observed that a further decrease in the accuracy of the phase estimate can lead to a noticeable decrease in the equivalent SNR. This implies that a more accurate phase estimate than the degraded phase will not significantly improve the equivalent SNR, while a less accurate phase estimate than the degraded phase can lead to a noticeable decrease in the equivalent SNR.

In the above discussions, we have considered the case in which the magnitude is obtained independently from the phase. It is known [5], however, that for almost all one-dimensional finite duration signals, a signal can be reconstructed within a scale factor from only the Fourier transform phase. Thus, in the context of speech enhancement, we may consider first estimating the phase more accurately and then attempting to reconstruct the signal from the phase information. Unfortunately, the accuracy in the reconstructed signal appears to be quite sensitive [6] to the accuracy of the phase, and such an approach for speech enhancement requires the ability to estimate very accurately the Fourier transform phase of the noise-free speech from the noisy speech.

In summary, we conclude that an effort to more accurately estimate the phase from the noisy speech is unwarranted in the context of speech enhancement if the estimate is used to reconstruct a signal by combining it with an independently estimated magnitude or to reconstruct the signal using the phase-only signal reconstruction algorithm [5]. However, if a significantly different approach is used to exploit the phase information such as using the phase estimate to further improve the magnitude estimate, then a more accurate estimation of phase may be important.

REFERENCES

[1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, Dec. 1979.

[2] M. R. Weiss, A. E. Aschkenasy, and T. W. Parsons, "Study and development of the INTEL technique for improving speech intelligibility," Nicolet Scientific Corp., Final Rep. NSC-FR/4023, Dec. 1974.

[3] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, May 1981.

[4] A. V. Oppenheim and R. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[5] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust., Speech, Signal Processing*, Dec. 1980.

[6] C. Y. Espy and J. S. Lim, "Effects of noise on signal reconstruction from Fourier transform phase," in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 1982, pp. 1833–1836.

Correction to "On the Effect of Correlation Between Truncation Errors in Fixed-Point Error Analysis of Winograd Short-Length DFT Algorithms"

G. PANDA, R. N. PAL, AND B. CHATTERJEE

On page 101 of the above paper,[1] the values of d_1, d_2, and d_5 of Fig. 1 are negative.

Manuscript received April 1, 1982.

The authors are with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India.

[1] G. Panda, R. N. Pal, and B. Chatterjee, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 100–104, Feb. 1982.

0096-3518/82/0800-0681$00.75 © 1982 IEEE