PHASE IN SPEECH AND PICTURES*

Alan V. Oppenheim
Jae S. Lim
Gary Kopec**
Stephen C. Pohlig

Massachusetts Institute of Technology
Lincoln Laboratory
Lexington, Massachusetts

ABSTRACT

It is generally known that preservation of phase information in images is important as evidenced by the fact that modifying an image such that phase is preserved but the magnitude of all the spectral components is set to unity generally preserves many of the features of the original image. A similar experiment has been carried out with speech, showing that preservation of only the long-time phase retains many important features of the original speech. An interpretation of these results, and their potential application to blind deconvolution, is suggested.

I. PHASE-ONLY SPEECH AND PICTURES

It is well known that for many images, the phase of the Fourier transform is more important than the magnitude[1,2]. Specifically if

$$F(u,v) = |F(u,v)|e^{j\Theta(u,v)} \quad (1)$$

denotes the two-dimensional Fourier transform of an image f(x,y) then the inverse Fourier transform of $e^{j\Theta(u,v)}$ has many recognizable features in common with the original, whereas the inverse Fourier transform of $|F(u,v)|$ generally bears no resemblance to the original. This is illustrated in Figure 1 where Figure 1a is the original image and Figure 1b is the phase-only image, i.e., the inverse Fourier transform of $e^{j\Theta(u,v)}$.*** Clearly, the phase-only image retains many of the features of the original. By contrast the magnitude-only image, i.e., the inverse Fourier transform of $|F(u,v)|$ of Figure 1a would have a small bright region near the origin in a dark background with no resemblance to the original image. As is evident in this example, the phase-only image often has the general appearance of a high-pass filtered version of the original with additive broadband noise.

A similar experiment has been carried out with speech with similar results as indicated in Figure 2. Specifically, the magnitude of the Fourier transform of an entire sentence was set to unity and the inverse transform was computed to obtain the phase-only equivalent speech. Figure 2a corresponds to the spectrogram of the original sentence and Figure 2b to the spectrogram of the phase-only equivalent. It is evident from Figure 2b that the basic formant structure of the original sentence has been preserved. In fact, in listening to the processed sentence as represented by Figure 2b total intelligibility is retained although the speech has the general quality associated with high-pass filtering and the introduction of additive uncorrelated noise. The magnitude-only speech has a large amplitude at the origin and some formant-like structure which provides a speech-like characteristic but with no speech intelligibility.

In the case of images the transformation of an original image to a phase-only image is a spectral whitening process which can be viewed as filtering with a frequency response $H(u,v) = 1/|F(u,v)|$. Since the spectra of images tend to fall off at high frequencies, the equivalent filter frequency response H(u,v) will have a high-pass characteristic. A similar interpretation can apply in the corresponding case of phase-only speech. Although this argument provides a general basis for an interpretation of the results, it is not precise enough for a complete explanation. The basic processing to obtain the phase-only signal is of course highly nonlinear and the simplified interpretation above assumes that it can be viewed as a linear process. Furthermore, while it is reasonable to identify $1/|F(u,v)|$ as generally emphasizing high frequencies over low frequencies it will also have specific spectral details associated with it which for some examples could certainly affect or obliterate important features in the original signal. Several simple examples serve to illustrate the point.

* This work was sponsored by the Department of the Air Force.

** Gary Kopec is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, and supported by the Fannie and John Hertz Foundation.

*** Because of the limitations of the printing procedure, the images contained in the figures may not reproduce accurately in the conference proceedings. Higher quality reproductions can be obtained from the authors.

CH1379-7/79/0000-0632$00.75

Consider, for example, any zero phase or linear phase image. Clearly the phase-only counterpart will consist only of a single impulse with a position dictated by the slope of the linear phase. A similar example in the case of speech would lead to a similar result. Consider, for example, the sentence illustrated in Figure 2. While the phase-only equivalent was highly intelligible, we can consider constructing a zero-phase sentence by concatenating this original with itself reversed in time, with an impulse separating them to ensure that the Fourier transform is positive. In this case, the phase-only equivalent signal will contain only an impulse. Again, in the case of speech, consider a steady state vowel. While the magnitude of the long-time Fourier transform falls off at high frequencies as we would expect, it also contains the resonances associated with the formants of the steady-state vowel. Thus the signal obtained from only the phase of the long-time Fourier transform would not be expected to contain the formants of the original vowel and thus the essential features of the original signal will have been lost. This is illustrated in Figure 3 where Figure 3a is the spectrogram of the original steady state vowel and Figure 3b is the spectrogram of the phase-only reconstruction.

The above examples suggest that there are factors to consider in addition to the basic whitening property of the phase-only reconstruction. In particular, we propose that for both speech and pictures, if the long-time Fourier transform is "sufficiently smooth," then intelligibility will be retained in the phase-only reconstruction. This condition can be interpreted in several ways. For speech, if the long-time transform is relatively smooth, the essential formant structure of the short-time transform will remain intact in the whitening process. For the case of images, if the long-time transform magnitude is smooth and falls off at high frequencies then the principal effect of the whitening process is to emphasize the high frequencies and therefore the edges in the image, thereby retaining many of the recognizable features. As developed in Appendix A, smoothness of the long-time transform magnitude equivalently implies a short impulse response for the whitening filter, and a short correlation function for the original signal. Although shortness of the correlation function does not by itself guarantee a smooth transform magnitude we believe that it is an important factor.

Importance of the long-time phase does not of course imply importance of short-time phase. This is certainly evident in the case of speech where it is well known that on a short-time basis, phase is relatively unimportant whereas, as evidenced by Figure 2, on a long-time basis the signal constructed from the phase only is intelligible and retains many of the important features of the original. One experiment that has been conducted which supports the notion that short-time phase is relatively unimportant in speech is the synthesis of speech in a homomorphic vocoder. While the speech synthesized using a vocal-tract impulse response with correct spectral magnitude but zero phase is highly intelligible, if the correct phase is used with a spectral magnitude of unity, intelligibility is almost completely lost.

The difference in importance of phase on a long-time versus short-time basis has consequences in a variety of practical areas including filtering and transform coding. In filtering, the importance of the phase of the filters is, of course, associated with the importance of phase on the time (or space) scale commensurate with the length of the filter impulse response. Thus, for example, the fact that for many pictures the phase of the overall image is important does not by itself imply that in image filtering, particular attention must be paid to the phase characteristics of the filter. In transform coding, the difference in importance of magnitude and phase as the size of the transform increases clearly indicates that the relative importance of accurate coding of each of these components changes.

II. BLIND DECONVOLUTION USING PHASE-ONLY RECONSTRUCTION

Since, on a long-time basis, phase alone contains a considerable amount of information about the signal there are a variety of potentially practical situations in which reconstruction of a signal from information only about the phase would be useful. For example, if a signal has been filtered with a zero phase filter whose spectral characteristics are otherwise unknown, reconstruction from the phase only may be preferable to the degraded signal. As a simple illustration consider the original image in Figure 1a. In Figure 4 is shown the result after blurring with a zero phase frequency response. Restoration of the blurred picture from the phase only with the magnitude set to unity results in the image in Figure 1b.

A simple and probably more preferable strategy to assigning a magnitude of unity is to apply a magnitude characteristic which is generally representative of the class of signals. For example, in the case of images, a magnitude characteristic obtained by averaging a large number of images could be used. As an illustration, for the example in Figure 1, the image obtained by applying an average magnitude characteristic to the correct phase is shown in Figure 5. Another example which demonstrates the usefulness of reconstructing a blurred image based on the phase-only reconstruction is shown in Figure 6. Figure 6a is the original, 6b is the blurred image, and 6c and 6d correspond to the phase-only reconstruction with unity magnitude and an average magnitude characteristic respectively.

The above procedure for blind deconvolution can in certain cases be related to estimating the blurring function by segmenting the signal and averaging the spectral magnitudes of the segments. Specifically, consider a one-dimensional signal f(t) with Fourier transform

$$F(\omega) = |F(\omega)| e^{j\Theta(\omega)} \quad . \tag{2}$$

Define the rectangular window r(t) as

$$r(t) = 1 \qquad 0 \leq t < T$$
$$= 0 \qquad \text{otherwise}$$

so that $f(t) = \sum_k f_k(t)$

where $f_k(t) = f(t)\, r(t - kT)$

Assuming that the cross-correlation between $f_k(t)$ and $f_j(t)$ for $k \neq j$ is negligible, the autocorrelation $\phi_{ff}(\tau)$ of f(t) is approximately

$$\phi_{ff}(\tau) \cong \sum_k \phi_{kk}(\tau) \qquad |\tau| < T$$
$$\cong 0 \qquad |\tau| > T$$

and in the frequency domain

$$|F(\omega)|^2 \cong \sum_k |F_k(\omega)|^2 \tag{3}$$

where $F_k(\omega)$ is the Fourier transform of $f_k(t)$.

Similarly if $f_b(t)$ represents the blurred signal with Fourier transform $F_b(\omega) = B(\omega)|F(\omega)|e^{j\Theta(\omega)}$ then

$$|F_b(\omega)|^2 \cong \sum_k |F_{bk}(\omega)|^2 \tag{4}$$

where $F_{bk}(\omega)$ is the Fourier transform of the k^{th} segment of the blurred signal. Thus the phase-only reconstructed signal is approximately obtained by dividing $F_b(\omega)$ by an estimate $\hat{B}(\omega)$ of the blurring function given by

$$\hat{B}(\omega) = \left[\sum_k |F_{bk}(\omega)|^2\right]^{1/2} \tag{5}$$

In the case when an average magnitude characteristic $A(\omega)$ is combined with the phase, the equivalent approximation to the blurring function is

$$\hat{B}(\omega) = \frac{1}{A(\omega)} \left[\sum_k |F_{bk}(\omega)|^2\right]^{1/2} \tag{6}$$

III. CONCLUDING REMARKS

In this paper, we have considered the importance of phase in speech and pictures. We have illustrated, by way of examples, that spectral whitening does not completely explain the high intelligibility of phase-only images and long-time speech. As a first step towards a more detailed interpretation, we have introduced the notion of "smoothness" of the long-time transform magnitude. With a particular measure of smoothness, we have shown that a smooth long-time transform magnitude implies a short impulse response of the whitening filter and a short correlation of the original signal. Even though we have not shown that short correlation implies a smooth long-time transform magnitude, we speculate that shortness of the correlation is an important factor for the long-time transform magnitude to be smooth. Qualitatively, if the correlation of a signal is short, the signal characteristics vary rapidly from one short-time segment to other short-time segments. Since the long-time transform magnitude squared under some conditions is approximately the average of uncorrelated short-time spectral magnitudes squared, a rapidly varying signal generally implies smooth long-time transform magnitude due to the averaging process. For a smooth long-time transform magnitude and therefore a smooth whitening filter, the spectral details of a signal will remain intact and the intelligibility will be preserved. Finally, we have considered the potential areas of application such as signal coding and restoration that could benefit from the importance of phase in a signal. As an example, we have related blind deconvolution to the restoration of an image from its phase when *the image is degraded by an unknown zero phase* blurring function.

In summary, we feel that the importance of phase in speech and pictures has important practical implications and there is considerable room for further refinement in the interpretation of the phase-only reconstruction experiments for pictures and speech. It is our hope that through this presentation, further discussion will be stimulated.

APPENDIX A

In this Appendix we relate the duration of the *equivalent whitening filter, and the autocorrelation* function to the smoothness of the Fourier transform magnitude. In the discussion we consider only a one-dimensional signal. The extension to two-dimensional signals is straightforward with identical results and conclusions.

Let f(t) denote a one-dimensional signal with Fourier transform $F(\omega) = |F(\omega)|e^{j\Theta(\omega)}$. We define the "smoothness" S of $|F(\omega)|$ as

$$S^2 = \int_{-\infty}^{+\infty} \left[\frac{1}{|F(\omega)|} \frac{d}{d\omega} |F(\omega)|\right]^2 d\omega \tag{A.1}$$

or, equivalently,

$$S^2 = \int_{-\infty}^{+\infty} \left[\frac{d}{d\omega} \log |F(\omega)|\right]^2 d\omega \quad . \tag{A.2}$$

With this particular choice of smoothness measure and a particular definition of signal duration, S provides an upper bound on the duration of the inverse transform of $|F(\omega)|$, on the duration of the autocorrelation function of f(t), and on the duration of the inverse transform of $1/|F(\omega)|$. More specifically, let g(t) denote the inverse Fourier

transform of $|F(\omega)|$ and define the duration Δg of $g(t)$ as

$$(\Delta g)^2 = \frac{\int_{-\infty}^{+\infty} t^2 g^2(t)\, dt}{\int_{-\infty}^{+\infty} g^2(t)\, dt} \qquad (A.3)$$

or, equivalently,

$$(\Delta g)^2 = \frac{\int_{-\infty}^{+\infty} [\frac{d}{d\omega}|F(\omega)|]^2\, d\omega}{\int_{-\infty}^{+\infty} [|F(\omega)|]^2\, d\omega} \qquad (A.4)$$

The numerator in (A.4) can be rewritten as

$$\int_{-\infty}^{+\infty} [\frac{d}{d\omega}|F(\omega)|]^2\, d\omega = \int_{-\infty}^{+\infty} |F(\omega)|^2 \times [\frac{1}{|F(\omega)|} \frac{d}{d\omega} |F(\omega)|]^2\, d\omega$$

and, using the Schwarz inequality,

$$\int_{-\infty}^{+\infty} [\frac{d}{d\omega}|F(\omega)|]^2\, d\omega \leqslant \int_{-\infty}^{+\infty} |F(\omega)|^2\, d\omega \times \int_{-\infty}^{+\infty} [\frac{1}{|F(\omega)|} \frac{d}{d\omega} |F(\omega)|]^2\, d\omega \qquad (A.5)$$

Substituting inequality (A.5) into (A.4),

$$\Delta g \leqslant S \qquad (A.6)$$

where S is defined in eq.(A.1) and (A.2). In a similar manner, we can show that the duration $\Delta\phi_f$ of the autocorrelation function $\phi_{ff}(\tau)$ of f(t) satisfies the inequality

$$\Delta\phi_f \leqslant 2S \quad , \qquad (A.7)$$

and that the duration Δh of h(t), the inverse Fourier transform of $1/|F(\omega)|$, satisfies the inequality

$$\Delta h \leqslant S \quad . \qquad (A.8)$$

ACKNOWLEDGMENT

We would like to express our appreciation to Mr. Alfred Gschwendtner, Dr. Herbert Kleiman and members of their research group at Lincoln Laboratory for making their image processing facilities available to us.

REFERENCES

(1) T. S. Huang, J. W. Burnett, A. G. Deczky, "The Importance of Phase in Image Processing Filters," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 23, No. 6, December 1975.

(2) W. K. Pratt, Digital Image Processing, John Wiley and Sons, 1978. See Figure 12.5-1 on page 328.

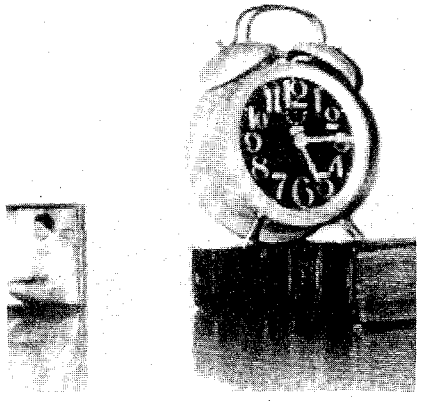

Fig. 1a. An original image.

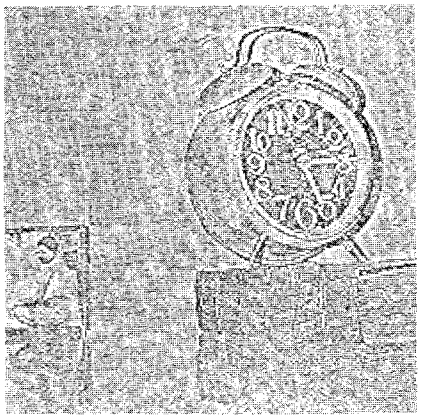

Fig. 1b. Phase only image of Fig. 1a. with the magnitude set to unity.

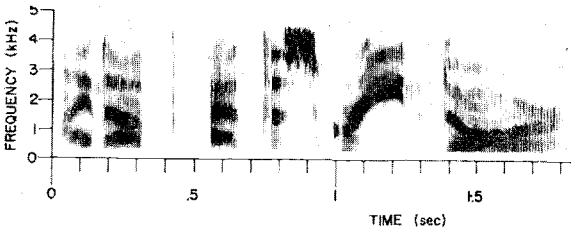

Fig. 2a. Spectrogram of a sentence "Line up at the screen door" spoken by a male speaker.

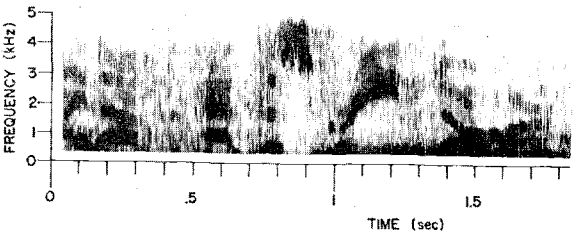

Fig. 2b. Spectrogram of phase only speech of Fig. 2a. with the magnitude set to unity.

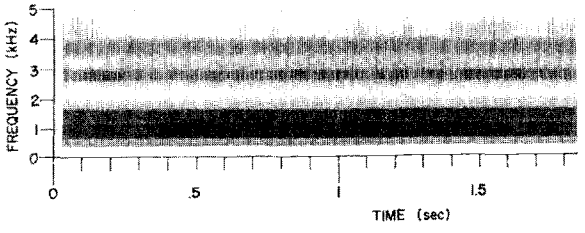

Fig. 3a. Spectrogram of a steady state vowel.

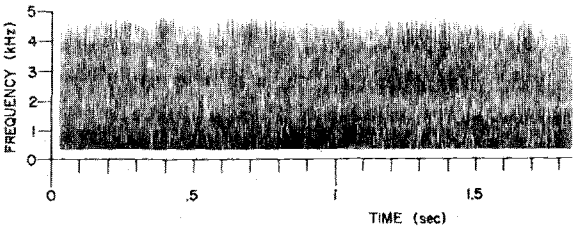

Fig. 3b. Spectrogram of phase only speech of Fig. 3a. with the magnitude set to unity.

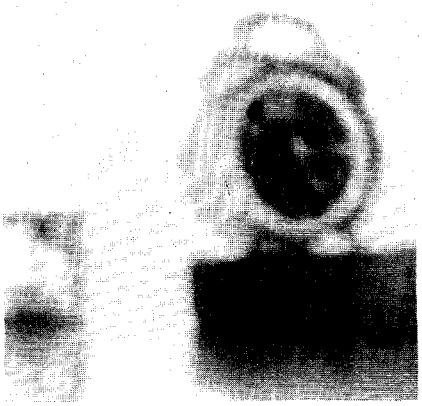

Fig. 4. The image in Fig. 1a. blurred by a blurring function with zero phase frequency response.

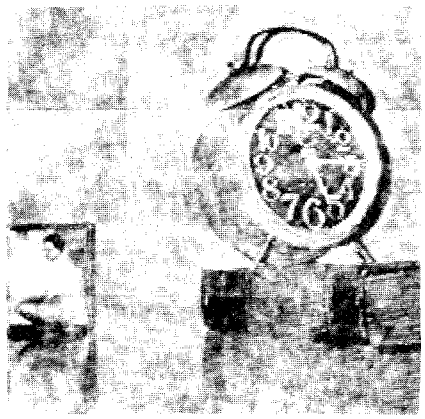

Fig. 5. Phase only image of Fig. 1a. with the magnitude set to an average magnitude of many images.

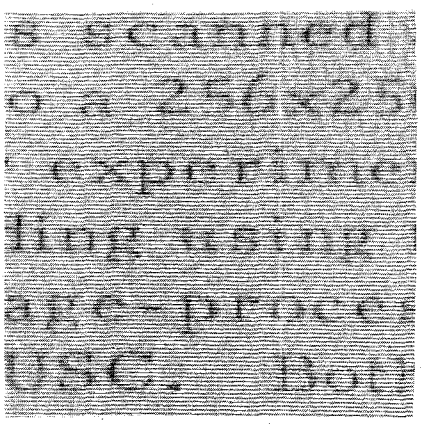

Fig. 6a. An original image.

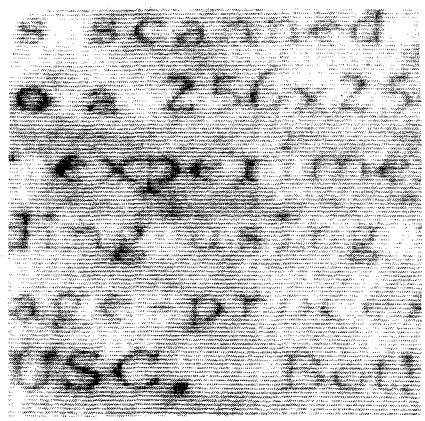

Fig. 6b. The image in Fig. 6a. blurred by a blurring function with zero phase frequency response.

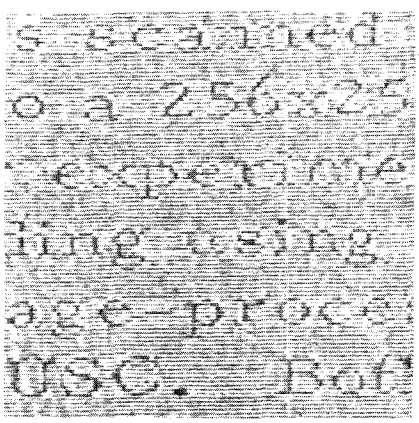

Fig. 6c. Phase only image of Fig. 6a. with the magnitude set to unity.

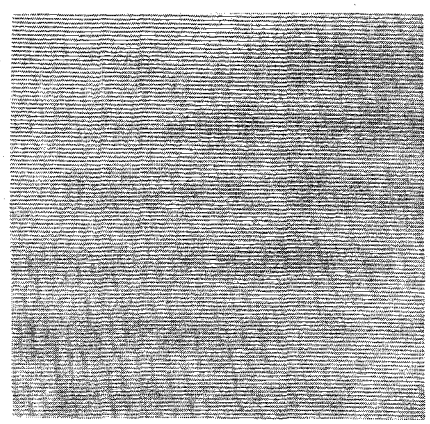

Fig. 6d. Phase only image of Fig. 6a. with the magnitude set to an average magnitude of many images.