



# On the usefulness of STFT phase spectrum in human listening tests <sup>☆</sup>

Kuldip K. Paliwal <sup>\*</sup>, Leigh D. Alsteris

*School of Microelectronic Engineering, Griffith University, Nathan campus, 4111 QLD, Australia*

Received 4 December 2003; received in revised form 6 May 2004; accepted 15 August 2004

## Abstract

The short-time Fourier transform (STFT) of a speech signal has two components: the magnitude spectrum and the phase spectrum. In this paper, the relative importance of short-time magnitude and phase spectra for speech perception is investigated. Human perception experiments are conducted to measure intelligibility of speech stimuli synthesized either from magnitude spectra or phase spectra. It is traditionally believed that the magnitude spectrum plays a dominant role for small window durations (20–40 ms); while the phase spectrum is more important for large window durations (>1 s). It is shown in this paper that even for small window durations, the phase spectrum can contribute to speech intelligibility as much as the magnitude spectrum if the analysis–modification–synthesis parameters are properly selected.

© 2004 Elsevier B.V. All rights reserved.

*Keywords:* Short-time Fourier transform; Phase spectrum; Magnitude spectrum; Speech perception; Overlap-add procedure; Automatic speech recognition

## 1. Introduction

In this paper, the usefulness of the phase spectrum <sup>1</sup> is explored in human speech perception. <sup>2</sup>

The authors have a long-term goal of utilising phase spectra in an effort to improve automatic speech recognition (ASR) performance. It is common practice in ASR to discard the phase

<sup>☆</sup> Audio files at <http://maxwell.me.gu.edu.au/spl/research/phase/project.htm>.

<sup>\*</sup> Corresponding author. Tel.: +61 7 3875 6536; fax: +61 7 3875 5384.

*E-mail addresses:* [k.paliwal@griffith.edu.au](mailto:k.paliwal@griffith.edu.au) (K.K. Paliwal), [l.alsteris@griffith.edu.au](mailto:l.alsteris@griffith.edu.au) (L.D. Alsteris).

<sup>1</sup> Throughout this paper, the modifier ‘short-time’ (i.e., finite-time) is implied when mentioning the phase spectrum and magnitude spectrum.

<sup>2</sup> There is a large amount of literature available on the topic of the perception of phase in speech, dating back to Ohm’s study (Ohm, 1843) in 1943. In our work, we only refer to a small percentage of this literature. In addition to the papers referenced throughout this text, the following selected papers may be of interest to the reader: Goldstein (1967), von Helmholtz (1912), Kim (2000), Patterson (1987), Plom and Steeneken (1969), Pobloth and Kleijn (1999), Schroeder (1959).

spectrum in favour of features that are derived purely from the magnitude spectrum<sup>3</sup> (Picone, 1993). In the ASR framework, speech is processed frame-wise using a temporal window of duration 20–40 ms. If the phase spectrum is to be of any use for ASR applications, it should provide some information about speech intelligibility using small window durations (20–40 ms) in a human perception experiment.

A few studies have been reported in the literature which discuss whether the phase spectrum provides any information which can contribute to intelligibility for human speech recognition (HSR). Schroeder (1975), and Oppenheim and Lim (1981) performed some informal perception experiments, concluding that the phase spectrum is important for intelligibility when the window duration of the short-time Fourier transform (STFT) is large ( $T_w > 1$  s), while it seems to convey negligible intelligibility at small window durations (20–40 ms).

Liu et al. (1997) have recently investigated the intelligibility of phase spectra through a more formal human speech perception study. They recorded six stop-consonants from 10 speakers in vowel–consonant–vowel context. Using these recordings, they created *magnitude-only* and *phase-only* stimuli. Magnitude-only stimuli were created by analysing the original recordings with a STFT, replacing each frame's phase spectra with random phase values, then reconstructing the speech signal using the overlap-add method. In the case of phase-only stimuli, the original phase of each frame was retained, while the magnitude of each frame was set to unity for all frequency components. The stimuli were created for various window lengths from 16 ms to 512 ms. These were played to subjects, whose task was to identify each as one of the six consonants. Their results (Fig. 1) show that intelligibility of magnitude-only stimuli decreases while the intelligibility of the phase-only stimuli increases as the window duration increases.

<sup>3</sup> There are other speech processing applications where spectral phase information is overlooked. For example, in speech enhancement it is common practice to modify the magnitude spectrum and keep the corrupt phase spectrum (Lim and Oppenheim, 1979; Wang and Lim, 1982).

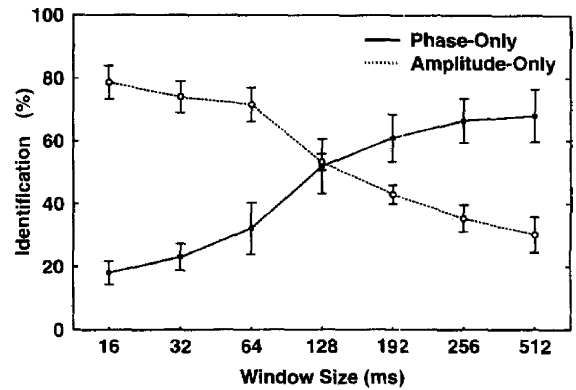


Fig. 1. Average identification performance and standard deviation as a function of window size for phase-only and magnitude-only stimuli, from the paper by Liu et al. (after Liu et al. (1997)).

For small window durations ( $T_w < 128$  ms), magnitude-only stimuli are significantly more intelligible than phase-only stimuli (while the opposite is true for larger window lengths). This implies that for small window durations (which are of relevance for ASR applications), the magnitude spectrum contributes much more towards intelligibility than the phase spectrum.

The authors of this paper initially set out to reproduce Liu's results; in doing so, made a number of modifications in Liu's analysis–modification–synthesis procedure (see Fig. 2). The modifications produce results which are different from Liu's results and more interesting from an ASR application's viewpoint. The first suggested modification is that of the analysis window type. Liu and his collaborators employed a Hamming window for construction of both the magnitude-only and phase-only stimuli. In our experiments, we find that the intelligibility of phase-only stimuli is improved significantly and becomes comparable to that of magnitude-only stimuli when a rectangular window is used. The second suggested modification is the choice of analysis frame shift; Liu et al. used a frame shift of  $T_w/2$ . As shown by Allen and Rabiner (1977), in order to avoid aliasing errors during reconstruction, the STFT sampling period (or frame shift) must be at most  $T_w/4$  for a Hamming window. In this paper, to be on the safer side, we use a frame shift of  $T_w/8$ . Our study also differs from Liu's study with

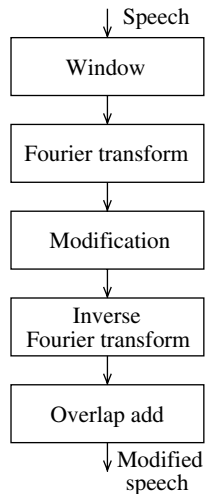


Fig. 2. Speech analysis–modification–synthesis system.

respect to the number of consonants used (16 for this study compared to 6 for Liu et al.). The design parameters are discussed in further detail later in this paper. Our results indicate that even for small window durations ( $T_w < 128$  ms), the phase spectrum can contribute to speech intelligibility as much as the magnitude spectrum if the analysis–modification–synthesis parameters are properly selected.<sup>4</sup>

The paper outline is as follows: In Section 2, we detail the analysis–modification–synthesis technique used to create the phase-only and magnitude-only stimuli. In Section 3, we describe a number of experiments which evaluate the importance of short-time phase spectra and short-time magnitude spectra in human speech perception. In the first experiment, we demonstrate that intelligibility of phase-only stimuli is improved significantly when a rectangular window is used, and it becomes comparable with that of magnitude-only stimuli even for small window durations. In Experiment 2, we construct magnitude-only and phase-only stimuli for window sizes ranging from 16 ms to 2048 ms, using both Liu’s parameter settings and our parameter settings (discussed in Experi-

ment 1) in order to compare their intelligibility. In the third experiment, we ascertain the contribution that each analysis–modification–synthesis parameter provides towards the intelligibility of signals reconstructed from phase spectra. In the aforementioned experiments, magnitude-only stimuli are created by randomising each frame’s phase spectra. It is also possible to create magnitude-only stimuli by setting all phase values for each frame to zero. Thus, in Experiment 4, we address the issue of using random-phase or zero-phase and determine if a significant difference exists between magnitude-only stimuli constructed with one or the other.

## 2. STFT analysis–modification–synthesis technique

Although speech is a non-stationary signal, it is generally assumed to be quasi-stationary and, therefore, can be processed through a short-time Fourier analysis (Allen, 1977; Allen and Rabiner, 1977; Crochiere, 1980; Flanagan and Golden, 1966; Griffin and Lim, 1984; Mathes and Miller, 1947; Portnoff, 1976, 1979, 1980, 1981; Quatieri, 2002; Rabiner and Schafer, 1978; Schafer and Rabiner, 1973). Note that the modifier ‘short-time’ implies a finite-time window over which the properties of speech may be assumed stationary; it does not refer to the actual duration of the window.<sup>5</sup> The STFT of a speech signal  $s(t)$  is given by

$$S(f, t) = \int_{-\infty}^{\infty} s(\tau)w(t - \tau)e^{-j2\pi f\tau}d\tau, \quad (1)$$

where  $w(t)$  is a window function of duration  $T_w$ . In speech processing, the Hamming window function is typically used and its width  $T_w$  is normally 20–40 ms.

We can decompose  $S(f, t)$  as follows:

$$S(f, t) = |S(f, t)|e^{j\psi(f, t)}, \quad (2)$$

where  $|S(f, t)|$  is the short-time magnitude spectrum and  $\psi(f, t) = \angle S(f, t)$  is the short-time phase spectrum. The signal  $s(t)$  is completely characterized by its short-time magnitude and phase spectra.

<sup>4</sup> Some of our preliminary results have been presented earlier in conferences (Paliwal, 2003; Paliwal and Alsteris, 2003; Alsteris and Paliwal, 2004).

<sup>5</sup> We use the qualitative terms ‘small’ and ‘large’ to make reference to the duration.

The aim of the experiments in Section 3 is to determine the contribution that the phase and magnitude spectra provide towards speech intelligibility. Accordingly, stimuli are created either from phase or magnitude spectrum (see Fig. 2). In order to construct, for example, an utterance with only phase spectra, the signal is processed through the STFT analysis using Eq. (1) and the magnitude spectrum is made unity in the modified STFT  $\hat{S}(f, t)$ ; that is

$$\hat{S}(f, t) = e^{j\phi(f, t)}. \quad (3)$$

This modified STFT is then used to synthesize the signal  $\hat{s}(t)$  using the overlap-add method<sup>6</sup> (Allen and Rabiner, 1977). The synthesized signal  $\hat{s}(t)$  contains all of the information about the short-time phase spectra contained in the original signal  $s(t)$ , but will have no information about the short-time magnitude spectra. We refer to this procedure as the STFT *phase-only synthesis* and the utterances synthesized by this procedure as the *phase-only* utterances. Similarly, for generating *magnitude-only* utterances, we retain each frame's magnitude spectrum and randomise each frame's phase spectrum; that is, the modified STFT is computed as follows:

$$\hat{S}(f, t) = |S(f, t)|e^{j\phi}, \quad (4)$$

where  $\phi$  is a random variable uniformly distributed between 0 and  $2\pi$ . It may also seem plausible to set  $\phi$  to zero for all values of  $f$  and  $t$ . In Experiment 4, reported later in Section 3, we test the intelligibility of magnitude-only stimuli constructed with zero-phase.

In the STFT-based speech analysis–modification–synthesis system (shown in Fig. 1), there are 4 design issues that must be addressed.

- (1) *Analysis window type*. This refers to the type of window function  $w(t)$  used for computing the STFT (Eq. (1)). A tapered window function (such as Hanning, Hamming or triangular) has been used in earlier studies (Liu et al.,

1997). Considering these studies have found the phase spectrum to be unimportant at small window durations, a rectangular (non-tapered) window function is investigated in this study in addition to a Hamming window function.

- (2) *Analysis window duration*. Over the course of the experiments, we investigate eight window durations (16, 32, 64, 128, 256, 1024, and 2048 ms).
- (3) *STFT sampling period (frame shift)*. In order to avoid aliasing during reconstruction, the STFT must be adequately sampled across the time axis. The STFT sampling period is decided by the window function  $w(t)$  used in the analysis. For example, for a Hamming window, the sampling period should be at most  $T_w/4$  (Allen and Rabiner, 1977). To be on the safer side, we have used a sampling period of  $T_w/8$ . Although the rectangular window can be used with a larger sampling period, we use the same sampling period (i.e.,  $T_w/8$ ) to maintain consistency. In this paper, we also refer to the STFT sampling period as the frame shift.
- (4) *Zero-padding*. For a windowed frame of length  $N$  (where  $N$  is a power of 2), the Fourier transform is computed using the fast Fourier transform (FFT) algorithm with a FFT size of  $2N$  points. This is equivalent to appending  $N$  zeros to the end of the  $N$ -length frame prior to performing the FFT. The resulting STFT is modified, then inverse Fourier transformed to get a reconstructed signal of length  $2N$ . Only the first  $N$  points are retained, while the last  $N$  points are discarded. This is done in order to minimise aliasing effects. Zero-padding is used in the construction of all stimuli in this study, unless otherwise stated.

### 3. Human perception experiments

#### 3.1. Experiment 1

In this experiment we compare the intelligibility of magnitude-only and phase-only stimuli using two window types: (1) a rectangular window, and

<sup>6</sup> In the following experiments, we use Allen and Rabiner's reconstruction method (Allen and Rabiner, 1977) rather than Griffin and Lim's method (Griffin and Lim, 1984). We have performed some experiments which indicate that there is no significant difference in intelligibility between stimuli constructed from either method.

Table 1  
Consonants used in all perception testing

a-Consonant-a	As in
aba	<u>b</u> at
ada	<u>d</u> eep
afa	<u>f</u> ive
aga	<u>g</u> o
aka	<u>k</u> ick
ama	<u>m</u> um
ana	<u>n</u> oon
apa	<u>p</u> ea
asa	<u>s</u> o
ata	<u>t</u> ea
ava	<u>v</u> ice
aza	<u>z</u> ebra
adha	<u>th</u> en
asha	<u>sh</u> ow
atha	<u>th</u> ing
azha	<u>m</u> easure

(2) a Hamming window.<sup>7</sup> This comparison is done at a small window duration of 32ms as well as a large window duration of 1024ms.

### 3.1.1. Recordings

We record 16 commonly occurring consonants in Australian English in aCa context (Table 1) spoken in a carrier sentence “Hear aCa now”. For example, for the consonant /d/, the recorded utterance is “Hear ada now”. These 16 consonants in the carrier sentence are recorded for four speakers: 2 males and 2 females, providing a total of 64 utterances. The recordings are made in a silent room with a SONY ECM-MS907 microphone (90° position). The signals are sampled at 16kHz with 16-bit precision. The duration of each recorded signal is approximately 3s.<sup>8</sup>

### 3.1.2. Stimuli

Each of the recordings are processed through the STFT-based speech analysis–modification–synthesis system to retain either only phase information or only magnitude information. In this

<sup>7</sup> A triangular window function is also investigated. Results are similar to those provided by the Hamming window in all test conditions. For the sake of clarity, we do not report these results in this paper.

<sup>8</sup> This time is inclusive of leading and trailing silence periods.

Table 2  
Stimuli for Experiment 1 (with frame shift of  $T_w/8$ )

Type of stimuli	Retained spectrum	Window type	Window duration (ms)
A1	Magnitude	Hamming	32
B1	Magnitude	Rectangular	32
C1	Phase	Hamming	32
D1	Phase	Rectangular	32
E1	Magnitude	Hamming	1024
F1	Magnitude	Rectangular	1024
G1	Phase	Hamming	1024
H1	Phase	Rectangular	1024

experiment, we investigate two window durations: (1)  $T_w = 32$ ms and (2)  $T_w = 1024$ ms.

There are eight types of stimuli for Experiment 1. The description of each type is provided in Table 2. Some extra details for stimuli construction are presented in Table 3.

### 3.1.3. Subjects

As listeners, we use 12 native Australian English speakers with normal hearing, all within the age group of 20–35 years. The subjects are different from those used for recording the speech stimuli.

### 3.1.4. Procedure

The perception tests for this experiment are conducted over two sessions. In the first session, the original speech signals and stimuli types A1, B1, C1, and D1 are presented. In the second session we present the original speech signals again, in addition to stimuli types E1, F1, G1, and H1.

The subjects are tested in isolation in a silent room. The reconstructed signals and the original signals (a total of 320 for each session) are played in random order via SONY MDR-V5000DF earphones at a comfortable listening level. The task is to identify each utterance as one of the 16

Table 3  
Detailed listing of settings for stimuli construction in Experiment 1

$T_w$ (ms)	32	1024
$T_w$ (samples)	512	16384
FFT length	1024	32768
$T_w/8$ (ms)	4	128
$T_w/8$ (samples)	64	2048

consonants. This way, we attain consonant identification accuracy (or, intelligibility) for each subject for different conditions. In both sessions, the subjects are first familiarised with the task through a short practice test. Session 1 (small window) results are provided in Table 4 and session 2 (large window) results are provided in Table 5. Results are averaged over the 12 subjects. The intelligibility of the original recordings is averaged over both sessions.

Responses are collected through software. The software displays the 16 aCa possibilities as well as an extra option for a null response. Participants are instructed to only choose the null response when they have no clue as to what the consonant may be. Responses are input via the keyboard in the form of numbers (1–17). Each audio file is presented once. No feedback is provided.

### 3.1.5. Results and discussion

The following observations can be made from Tables 4 and 5:

- (1) For the large window duration of 1024ms, the phase spectrum provides significantly more information than the magnitude spectrum for

Table 4

Experiment 1: consonant intelligibility (or, identification accuracy) of magnitude-only and phase-only stimuli for a small window duration of 32ms (with  $T_w/8$  frame shift)

Type of stimuli	Intelligibility (in %) for	
	Hamming window	Rectangular window
Original	89.9	89.9
Magn. only	84.2 (A1)	78.1 (B1)
Phase only	59.8 (C1)	79.9 (D1)

Table 5

Experiment 1: consonant intelligibility (or, identification accuracy) of magnitude-only and phase-only stimuli for a large window duration of 1024ms (with  $T_w/8$  frame shift)

Type of stimuli	Intelligibility (in %) for	
	Hamming window	Rectangular window
Original	89.9	89.9
Magn. only	14.1 (E1)	13.3 (F1)
Phase only	88.0 (G1)	89.3 (H1)

both the Hamming window function ( $F[1, 11] = 2880.57$ ,  $p < 0.01$ ) and the rectangular window function ( $F[1, 11] = 1582.38$ ,  $p < 0.01$ ). This observation is consistent with the results reported earlier in the literature (Liu et al., 1997; Oppenheim and Lim, 1981; Schroeder, 1975).

- (2) The difference in intelligibility between magnitude-only stimuli constructed with a Hamming window and magnitude-only stimuli constructed with a rectangular window at a large window duration of 1024ms is insignificant ( $F[1, 11] = 0.63$ ,  $p < 0.01$ ). The same can also be said for phase-only signals constructed with either window type at the large window duration ( $F[1, 11] = 1.18$ ,  $p < 0.01$ ).
- (3) For the small window duration of 32ms, intelligibility of magnitude-only stimuli is significantly better than the phase-only stimuli when the Hamming window function is used ( $F[1, 11] = 17.4$ ,  $p < 0.01$ ), but these are comparable when the rectangular window function is used ( $F[1, 11] = 2.91$ ,  $p < 0.01$ ). Thus, if a rectangular window function is used in the STFT analysis–modification–synthesis system, the phase spectrum carries as much information about the speech signal as the magnitude spectrum, even for small window durations, which are typically used in speech processing applications.
- (4) For a small window duration of 32ms, the Hamming window provides better intelligibility than the rectangular window for magnitude-only stimuli ( $F[1, 11] = 29.38$ ,  $p < 0.01$ ); while the rectangular window is better than the Hamming window for the construction of phase-only stimuli ( $F[1, 11] = 176.30$ ,  $p < 0.01$ ).
- (5) For a small window duration of 32ms, the best intelligibility results from magnitude-only stimuli (obtained by using a Hamming window) are significantly better than the best results from phase-only stimuli (obtained using a rectangular window) ( $F[1, 11] = 17.14$ ,  $p < 0.01$ ).

These results can be explained as follows. The multiplication of a speech signal with a window



function is equivalent to the convolution of the speech spectrum  $S(f)$  with the spectrum  $W(f)$  of the window function. The window's magnitude spectrum  $|W(f)|$  has a big main lobe and a number of side lobes. This causes two problems: (1) frequency resolution problem and (2) spectral leakage problem. The frequency resolution problem is caused by the main lobe of  $|W(f)|$ . When the main lobe is wider, a larger frequency interval of the speech spectrum gets smoothed and the frequency resolution problem becomes worse. The spectral leakage problem is caused by the side lobes; the amount of spectral leakage increases with the magnitude of the side lobes. For magnitude-only utterances, we want to preserve the true magnitude spectrum of the speech signal. For the estimation of the magnitude spectrum, frequency resolution as well as spectral leakage are serious problems. Since the Hamming window has a wider main lobe and smaller side lobes in comparison to the rectangular window, the Hamming window provides a better trade-off between frequency resolution and spectral leakage than the rectangular window and, hence, it results in higher intelligibility for the magnitude-only utterances. For the estimation of the phase spectrum, it seems that the side lobes do not cause a serious problem; the smoothing effect caused by the main lobe appears to be more serious. It is because of this that the rectangular window results in better intelligibility than the Hamming window for phase-only utterances. Reddy and Swamy (1985) have also recommended the use of a rectangular window function in the computation of the group delay spectrum, which is a frequency derivative of the phase spectrum.

For magnitude-only stimuli constructed with a small window duration, the best intelligibility is obtained for a Hamming window (type A1). For phase-only stimuli constructed with a small window duration, the best intelligibility is obtained when a rectangular window is used (type D1). In order to provide some details about the acoustic properties of these stimuli, we present, in Fig. 3,

a spectrogram<sup>10</sup> for a sentence of speech and the corresponding magnitude-only (type A1) and phase-only (type D1) spectrograms.<sup>11</sup> The magnitude-only spectrogram is visually more similar to that of the original spectrogram than the phase-only spectrogram. In keeping the magnitude information, we also maintain the frame energies; thus, in the magnitude-only reconstruction, the short-time energy contour is preserved. The image contrast, therefore, in the magnitude-only spectrogram is similar to that of the original spectrogram. In phase-only reconstruction, however, setting each frame's magnitude spectra to unity suppresses energy information, resulting in an almost constant energy contour over the duration of the reconstructed signal. This results in the silent parts at the beginning and end of the original utterance being heard as loud as the speech parts in the reconstructed signal.

Another interesting point can be observed through the time-domain plots in Fig. 4. This figure compares a 32 ms time frame of speech to its reconstructed magnitude-only (type A1) and phase-only (type D1) signals. In the phase-only reconstruction (Fig. 4(b)), pitch epochs are preserved, while they are lost in the magnitude-only reconstruction (Fig. 4(c)). Thus, the phase-only reconstruction preserves the pitch-related timing aspects.

As noted previously, the choice of analysis window type for the construction of magnitude-only and phase-only stimuli for large window durations is unimportant. For consistency, however, the authors recommend that the best analysis window functions used for constructing magnitude-only and phase-only stimuli at small window durations (as in stimuli types A1 and D1) should also be used for construction at large window durations (as in stimuli types E1 and H1). With this in mind, we introduce Fig. 5, which presents magnitude-only (type E1) and phase-only (type H1) reconstructions of the same speech. At large analysis window

<sup>9</sup> The window's phase spectrum  $\angle W(f)$  is a linear function of frequency and, hence, does not cause a problem in estimating the speech spectrum  $S(f)$ .

<sup>10</sup> Unless otherwise stated, all spectrograms in this paper are constructed using a Hamming analysis window of length 32 ms, a time shift of 1 ms, a pre-emphasis coefficient of 0.97 and a dynamic range of 50 dB.

<sup>11</sup> Refer to Appendix A for an explanation of why we can see formant structure in the "phase-only" stimuli.

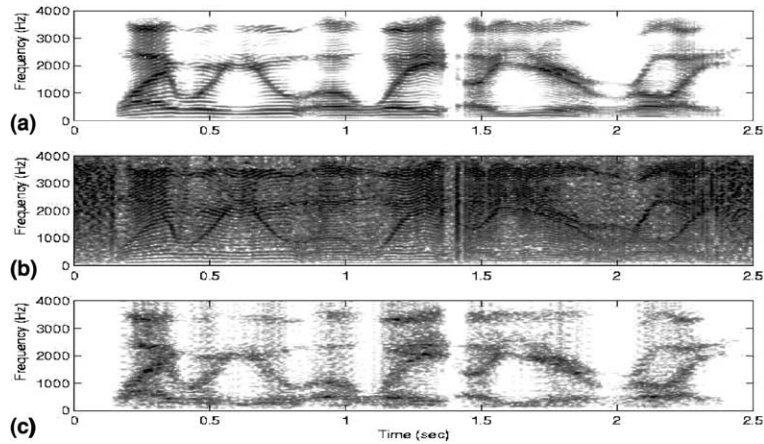


Fig. 3. (a) Spectrogram of the original speech sentence “Why were you away a year Roy?”, (b) phase-only (type D1) spectrogram, and (c) magnitude-only (type A1) spectrogram.

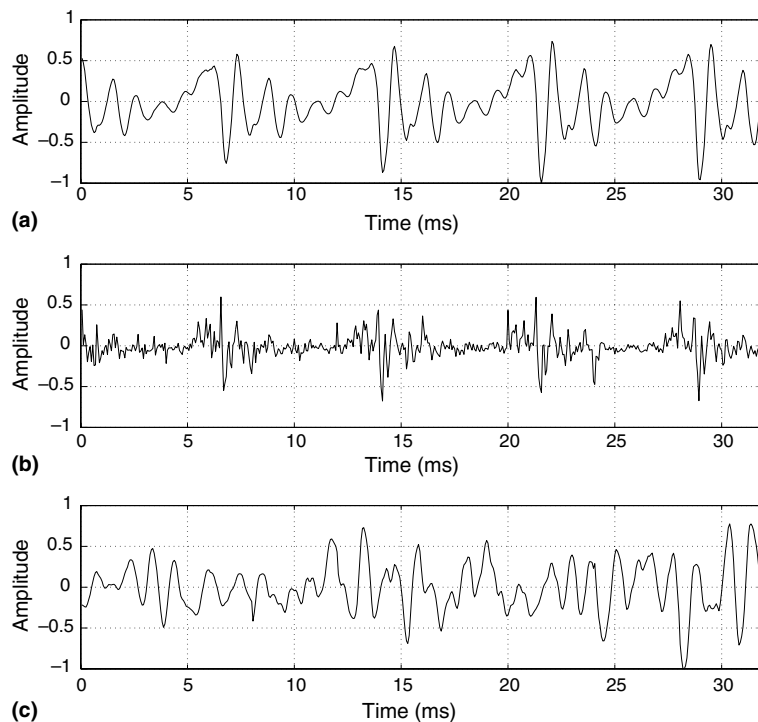


Fig. 4. (a) 32ms segment of speech, (b) phase-only (type D1) reconstruction, and (c) magnitude-only (type A1) reconstruction.

durations (we use  $T_w = 1024$  ms), phase-only stimuli (type H1) provide much better intelligibility than magnitude-only stimuli (type E1). Formant tracks are visible in the phase-only spectrogram

(Fig. 5(b)), and are absent in the magnitude-only spectrogram (Fig. 5(c)). This explains the better intelligibility of phase-only stimuli over magnitude-only stimuli for large window durations.



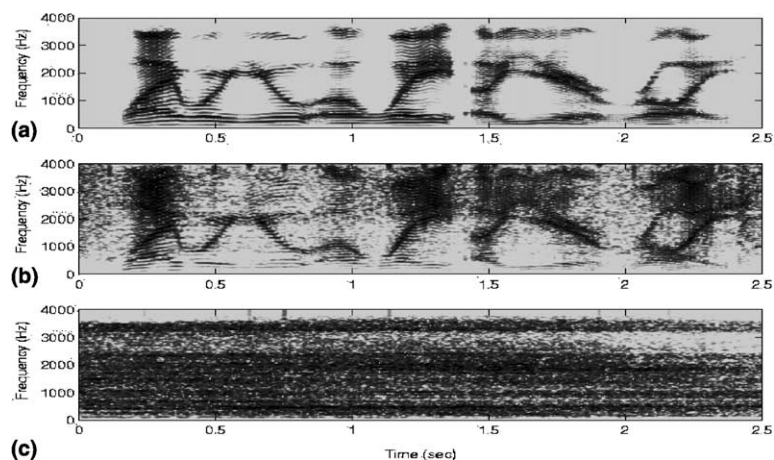


Fig. 5. (a) Spectrogram of the original speech sentence “Why were you away a year Roy?”, (b) phase-only (type H1) spectrogram, and (c) magnitude-only (type E1) spectrogram.

### 3.2. Experiment 2

In this experiment, we investigate more closely how intelligibility varies with window duration for magnitude-only and phase-only stimuli. We construct magnitude-only and phase-only stimuli using the analysis–modification–synthesis parameters used by Liu et al. (1997) and compare the intelligibility scores to stimuli constructed using our best parameter settings suggested in Experiment 1. This comparison is made over a number of analysis window durations, ranging from 16ms to 2048 ms.

#### 3.2.1. Stimuli

In their experiments, Liu et al. used a Hamming window and a frame shift of  $T_w/2$  for construction of both phase-only and magnitude-only stimuli. These parameter selections differ to those suggested in this paper, where we use a rectangular window for phase-only reconstruction, a Hamming window for magnitude-only reconstruction, and a frame shift of  $T_w/8$  for both types of stimuli.

In this experiment we have four types of stimuli to compare at eight analysis window durations (16, 32, 64, 128, 256, 512, 1024, and 2048 ms). Table 6 details the parameters used to construct each type of stimulus and the names subsequently used to reference them. Stimuli types A2 and B2 are constructed with Liu’s settings and stimuli types C2

Table 6  
Stimuli for Experiment 2

Type of stimuli	Retained spectrum	Window type	Frame shift
A2	Phase	Hamming	$T_w/2$
B2	Magnitude	Hamming	$T_w/2$
C2	Phase	Rectangular	$T_w/8$
D2	Magnitude	Hamming	$T_w/8$

and D2 are constructed using the best settings suggested in Experiment 1 of this section.

#### 3.2.2. Procedure

The experiment is split into two parts: in the first part, the intelligibility of stimuli types A2 and B2 are compared, while in the second part we compare the intelligibility of stimuli types C2 and D2. The details of the experimental setup are the same as those used in Experiment 1.

#### 3.2.3. Results and discussion

The intelligibility of stimuli types A2 and B2 over all analysis window durations are compared in Fig. 6(a). The intelligibility of magnitude-only (type B2) stimuli is almost 2 times better than that of the phase-only (type A2) stimuli at small analysis window durations. The intelligibility of magnitude-only (type B2) stimuli decreases while intelligibility of phase-only (type A2) stimuli increases as the analysis window duration increases.

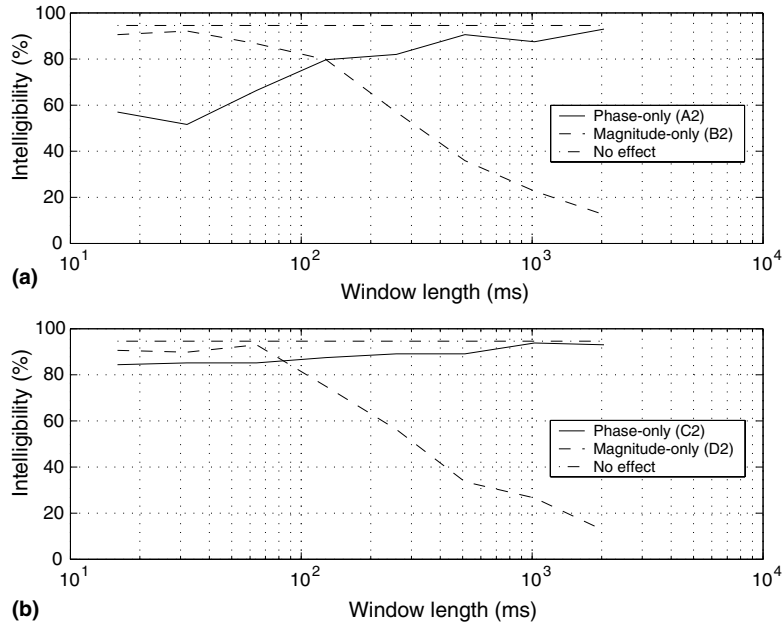


Fig. 6. Consonant identification performance (or, intelligibility) as a function of window duration for the magnitude-only and phase-only stimuli of Experiment 2. Intelligibility for the original utterances (without any modification) is shown by horizontal dot-dashed line. (a) Stimuli type A2 versus stimuli type B2, (b) Stimuli type C2 versus stimuli type D2.

The crossover point is around 128 ms. The trends observed here are similar to those observed by Liu and his colleagues (Liu et al., 1997).

Intelligibility results for stimuli types C2 and D2 are shown in Fig. 6(b). It can be observed from this figure that for magnitude-only (type D2) stimuli, the intelligibility decreases with an increase in window duration. The trend of this relationship is similar to that for type B2 stimuli (Liu's method). For phase-only stimuli of type C2 (our method of constructing phase-only stimuli), the intelligibility scores are almost the same for all the window durations. Compare this to Fig. 6(a), where the intelligibility of type A2 (Liu's method of constructing phase-only stimuli) is much worse at small window lengths. The crossover point in Fig. 6(b) is between 64 ms and 128 ms.

Table 7 compares the intelligibility scores at 32 ms of all 4 types of stimuli. There is no significant difference between the two types of magnitude-only stimuli (B2 and D2). With the help of a rectangular window and the analysis frame shift of  $T_w/8$ , the human recognition results for short-

Table 7

Experiment 2: Comparison of consonant intelligibility (or, identification accuracy) of magnitude-only and phase-only stimuli constructed with our parameter settings and those settings used in Liu et al. (1997) at 32 ms window duration

Type of stimuli	Intelligibility (in %) for	
	Our settings	Liu et al. (1997) settings
Original	94.6	94.6
Magn. only	89.8 (D2)	92.2 (B2)
Phase only	85.2 (C2)	51.6 (A2)

time phase-only stimuli is significantly improved at small window lengths.

### 3.3. Experiment 3

As seen in Experiment 2, our intelligibility results for phase-only stimuli are better than previously reported by Liu et al. (1997). We have shown that phase-only stimuli, constructed with different parameter settings than those used by

Liu et al., provide intelligibility comparable to that of magnitude-only stimuli at small analysis window durations. It will be interesting to see the reasons why we get this improvement in intelligibility. Recall the 4 analysis–modification–synthesis parameters discussed in Experiment 1: window type, window duration, window shift and zero-padding. In this experiment, we determine the contribution that each parameter setting provides towards improving the intelligibility of signals reconstructed from short-time phase spectra.

### 3.3.1. Stimuli

The previous experiment demonstrated that it is possible to attain a good intelligibility score for phase-only stimuli with a small analysis window duration of 32ms. Therefore, in this experiment, the window duration is set constant at 32ms. A number of combinations of the other parameters are tested in order to ascertain their respective contribution to the intelligibility of phase-only stimuli. Table 8 details the parameters used to construct each type of stimuli and the names we will use to refer to them in this experiment. Note that stimuli types A3, B3 and C3 are constructed without zero-padding. The original recordings and the stimuli provide a total of 320 audio files.

### 3.3.2. Procedure

The 320 audio files are presented to each subject in a single session. The details of the experimental setup are the same as those used in Experiment 1.

Table 8  
Comparison of consonant intelligibility (or, identification accuracy) for the phase-only stimuli used in Experiment 3 ( $T_w = 32\text{ms}$ )

Type of stimuli	Parameter settings	Phase-only intelligibility
A3	Hamming window, $T_w/2$ overlap	45.3%
B3	Rectangular window, $T_w/2$ overlap	76.6%
C3	Rectangular window, $T_w/8$ overlap	82.8%
D3	Rectangular window, $T_w/8$ overlap, zero-padding	85.9%

### 3.3.3. Results and discussion

The intelligibility scores are provided in Table 8. The scores indicate that the major contribution to overall intelligibility comes from the use of the rectangular window (stimuli type B3). Decreasing the frame shift provides a smaller improvement in intelligibility (stimuli type C3). The zero-padding also seems to contribute a slight improvement in intelligibility (stimuli type D3).

Fig. 7 presents the spectrogram of a sentence of speech with its reconstructed phase-only stimuli A3, B3, C3 and D3. The increasing clarity of the formant tracks in these spectrograms, from A3 through D3, is indicative of the corresponding trend in the intelligibility of these stimuli.

## 3.4. Experiment 4

In the experiments described so far, we have replaced the short-time phase spectrum by random numbers in order to reconstruct our magnitude-only stimuli. It will be interesting to see how making the phase spectrum zero for all frequencies will affect the intelligibility of the magnitude-only stimuli. Therefore, in this experiment, we address the issue of replacing phase with random-phase and zero-phase in the construction of magnitude-only stimuli. In order to determine if there exists a significant intelligibility difference between these two choices, we conduct the following experiment.

### 3.4.1. Stimuli

Two sets of magnitude-only stimuli are constructed; one with zero-phase and the other with random-phase. A short window duration of 32ms, frame shift of  $T_w/8 = 4\text{ms}$ , and a Hamming analysis window are used. The 64 original utterances and the reconstructed stimuli provide a total of 192 audio files.

### 3.4.2. Procedure

The 192 audio files are presented to each subject in a single session. The details of the experimental setup are the same as those used previously.

### 3.4.3. Results and discussion

The results of this experiment are shown in Table 9. The intelligibility of the random-phase

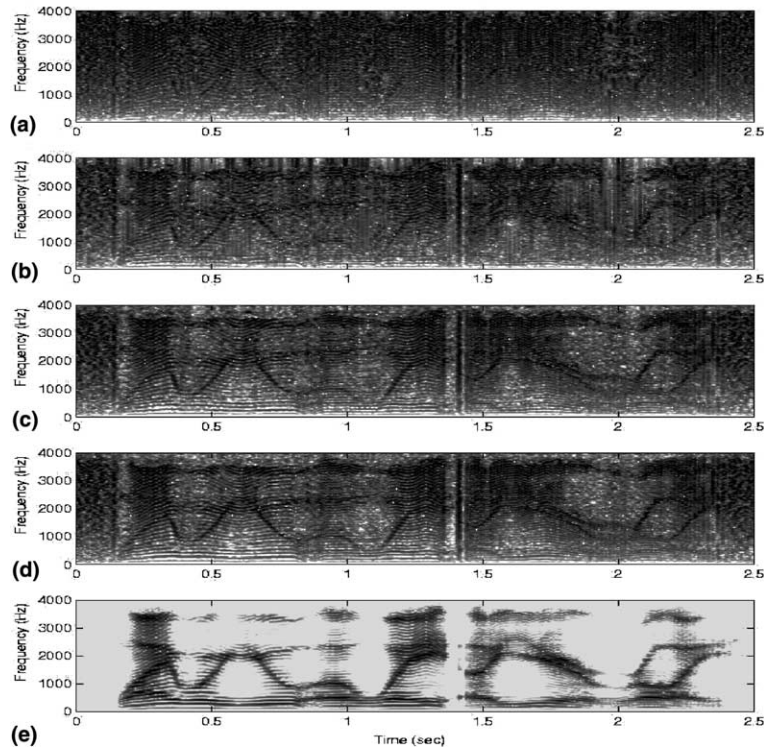


Fig. 7. The spectrograms of phase-only stimuli at an analysis window duration of 32ms: (a) stimulus type A3, (b) stimulus type B3, (c) stimulus type C3, (d) stimulus type D3, and (e) spectrogram of the original speech sentence “Why were you away a year Roy?”. Stimulus construction parameters are given in Table 8.

Table 9

Experiment 4: Consonant intelligibility (or, identification accuracy) of magnitude-only stimuli constructed with random-phase and zero-phase (with  $T_w/8$  frame shift, and  $T_w = 32$ ms)

Type of stimuli	Intelligibility (in %)
Original	91.1
Random phase	86.4
Zero phase	75.4

stimuli is significantly higher than those of the zero-phase stimuli.

We can explain this result by observing the spectrograms of Fig. 8. Setting the phase of each frame to zero introduces a periodicity (with a period equal to that of the frame shift) which manifests itself as horizontal lines on the spectrogram (Fig. 8(b)), producing a high pitched, unnatural sounding speech. The resulting stimuli is so unnatural that it seems to have an adverse effect on intel-

ligibility. The subjects described the zero-phase stimuli as ‘harsh’ and ‘robotic’.

To explain the periodicity in the zero-phase stimuli, we refer to Fig. 9. To create the waveform in Fig. 9(b), we perform the STFT analysis on the 32ms frame of speech shown in Fig. 9(a), replace its phase spectrum with zero-phase, then apply an inverse Fourier transform. As expected, we get an autocorrelation-like sequence. Thus, zero-phase puts more energy towards the beginning of the frame. Note that the first sample in the frame, which is similar to the zeroth order autocorrelation coefficient, has the largest value. The waveform in Fig. 9(c) is constructed in a similar manner except that the phase spectrum is replaced by random-phase. The random-phase distributes the energy across the time-axis. Fig. 10 illustrates the results of using zero-phase and random-phase for the reconstruction of a magnitude-only stimuli. The large peak of the zero-phase signal in Fig. 9(b)

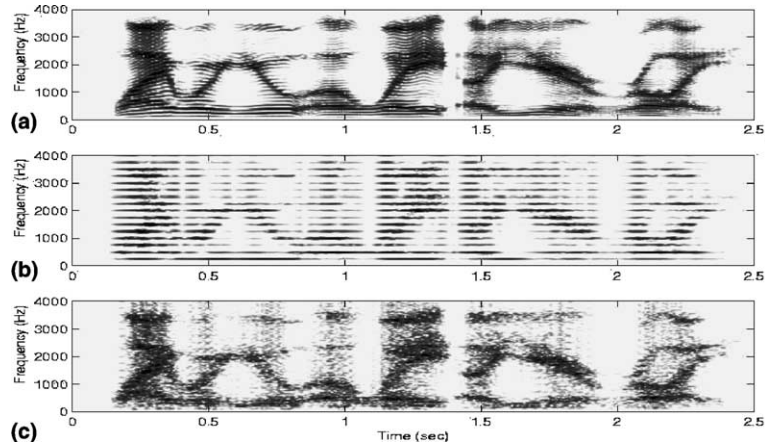


Fig. 8. Spectrograms of (a) the original speech sentence “Why were you away a year Roy?”, (b) magnitude-only, zero-phase stimuli, and (c) magnitude-only, random-phase stimuli (with  $T_w/8$  frame shift, and  $T_w = 32$  ms).

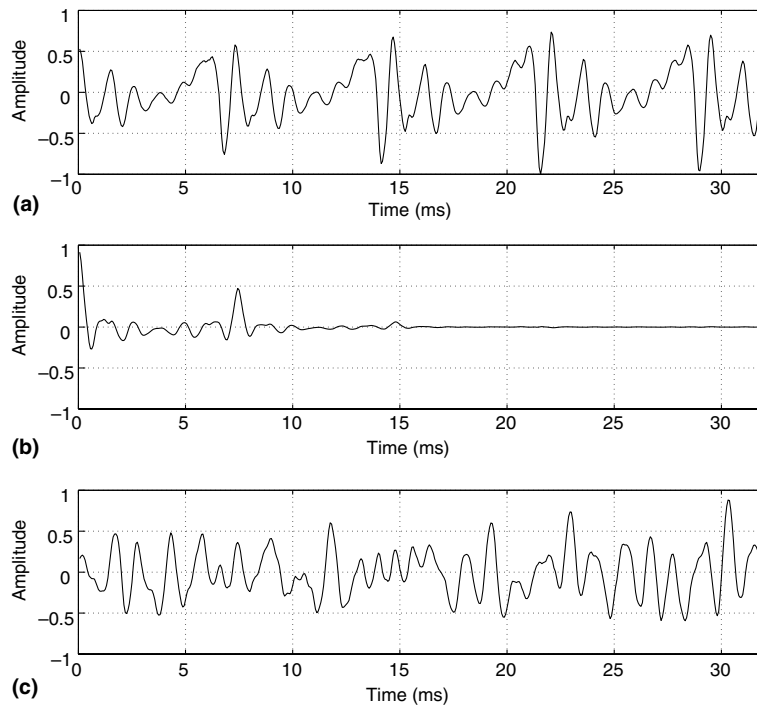


Fig. 9. The result of processing one frame of speech, retaining only its magnitude information. (a) Original speech segment, (b) zero-phase and (c) random-phase.

now repeats itself every 4 ms ( $1/8$  of 32 ms) in Fig. 10(b). This periodicity is not present in the random-phase signal of Fig. 10(c). Consequently, the random-phase stimuli are ‘easier to listen to’. Sub-

jects describe the random-phase stimuli as ‘natural’, ‘mellow’, and ‘breathy’. This naturalness is most likely attributed to the lower peak factor of the random-phase signal (Schroeder and Strube, 1986).

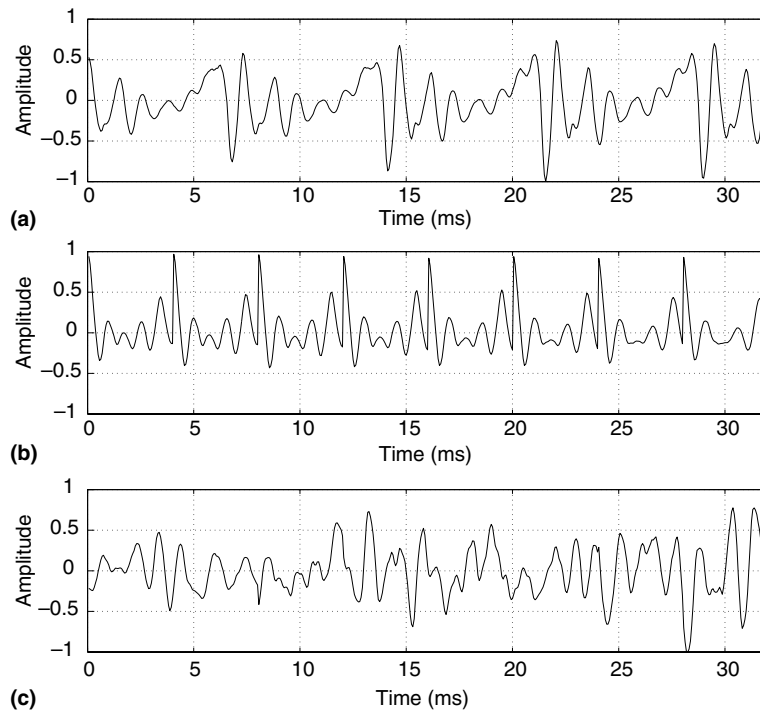


Fig. 10. (a) 32ms segment of speech, and its magnitude-only reconstruction using, (b) zero-phase and (c) random-phase.

#### 4. Conclusion

In this paper, the relative importance of short-time magnitude and phase spectra on speech perception is investigated. Human perception experiments are conducted to measure intelligibility of speech stimuli reconstructed either from magnitude spectra or phase spectra. The experiments reported here demonstrate that even for small window durations, phase spectra can contribute to speech intelligibility as much as magnitude spectra if the analysis–modification–synthesis parameters are properly selected.<sup>12</sup>

Although the STFT phase spectrum is as important for speech intelligibility as the STFT magnitude spectrum, it is not clear whether these

two spectra contribute to speech intelligibility in a complementary (or independent) fashion. In order to answer this question, we have done a detailed analysis of confusion matrices for consonant identification obtained from Experiment 1. However, we have not been able to observe any consistent pattern. We intend to explore this issue further in the future.<sup>13</sup>

In this work, we reconstruct intelligible speech from short-time magnitude spectra by making the short-time phase spectra random. We also reconstruct intelligible speech from short-time phase spectra by setting the short-time magnitude spectra to unity. It may be possible to improve the intelligibility of the reconstructed stimuli if we make some assumptions about the speech signal.

<sup>12</sup> During the revision of this work, the authors noticed a paper by Cox and Robinson (1980) which tends to confirm the findings of some of the experiments reported in the present paper.

<sup>13</sup> To maintain clarity, the confusion matrices are not provided in this paper. The interested reader can view the confusion matrices at <http://maxwell.me.gu.edu.au/spl/research/phase/project.htm>.



For example, if we assume speech to be a minimum phase signal, the phase spectra and magnitude spectra are related through the Hilbert transform (Oppenheim and Schaffer, 1975). This makes it possible to reconstruct the phase spectrum of a speech frame given its magnitude spectrum, or to reconstruct the magnitude spectrum given its phase spectrum. A number of reconstruction algorithms, which utilise the constraints imposed by certain assumptions made about the speech signal, have been proposed in the literature (Espy and Lim, 1983; Griffin and Lim, 1984; Hayes et al., 1980; Izraelevitz, 1985; Merchant and Parks, 1983; Nawab et al., 1983; Quatieri and Oppenheim, 1981; Thomas and Hayes, 1984; Van Hove et al., 1983; Yegnanarayana et al., 1984; Yegnanarayana et al., 1987). It is the intention of the authors to implement some of these alternative reconstruction procedures to ascertain if more intelligibility can be drawn from short-time phase spectra of speech. The stimuli for this work were created from recordings made in clean conditions. The usefulness of phase information in the presence of noise and other adverse influences will also be investigated in the future.

The goal of this work is to gain a better understanding of the role that the short-time phase spectrum plays in human speech perception. The results of our experiments in this paper have shown that the phase spectrum not only contributes to speech intelligibility at large analysis window durations ( $T_w = 1024$ ms), but also at small analysis window durations ( $T_w = 32$ ms). Since the speech processing in ASR applications is done frame-wise over small analysis window durations (20–40ms), it is logical to investigate the use of phase spectrum to extract features for these applications. Some preliminary results have already been reported earlier (Paliwal, 2003; Paliwal and Atal, 2003), which show the usefulness of phase spectrum for ASR. More detailed results will be reported in the future.

## Acknowledgements

This work was partly supported by ARC (Discovery) grant (No. DP0209283). The authors

also wish to thank the volunteers who took part in the subjective listening tests reported in this paper.

## Appendix A

Throughout this paper, we have explained intelligibility results through the comparison of spectrograms. There seems to be a direct correlation between intelligibility and the presence of formant-like structure in these spectrograms. One may ask how do we get this formant structure in the spectrograms of our “phase-only” stimuli. This may happen either due to the overlap-add procedure used in the reconstruction of phase-only stimuli, or it may come as an artifact of spectrogram computation. These issues are addressed in the following discussion.

We construct a phase-only signal using a rectangular window with duration of  $T_w = 32$ ms and frame shift of  $T_w$  (i.e., no overlap). We compute the spectrogram<sup>14</sup> for this signal with a window duration of 32ms and a frame shift of 32ms. This is shown in Fig. A.1(b). As expected, we attain a flat spectrogram. Although one can hear speech in the signal (albeit with little intelligibility), it is an interesting observation that the spectrogram provides no information whatsoever. If we change the spectrogram frame shift to 1ms, we obtain Fig. A.1(c). The reason we can now see some formant structure is because the magnitude spectrum is not unity for all frames used in the spectrogram computation. The unity-magnitude spectrum constraint only exists if the spectrogram frame duration is 32ms (the same as the reconstruction frame duration) and its ends coincide with the ends of a reconstruction frame (e.g., Fig. A.2(a)). Thus, wherever the spectrogram frame does not line up with a reconstruction frame, the unity-magnitude spectrum constraint

<sup>14</sup> This spectrogram is created with a rectangular analysis window and no pre-emphasis in order to visualise the effect of the unity-magnitude constraint. For consistency, all other spectrograms for this discussion are created in the same manner.

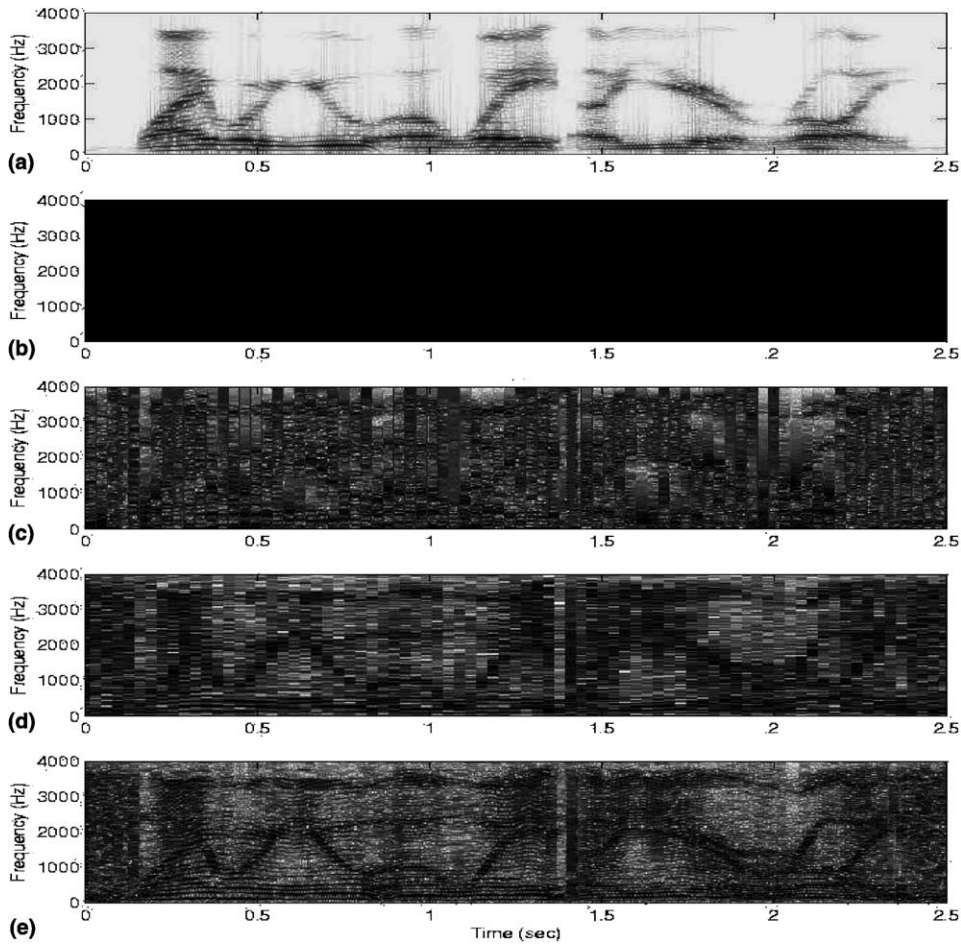


Fig. A.1. Spectrograms of (a) the original speech sentence “Why were you away a year Roy?”, and the phase-only stimuli, where (b) the stimulus is constructed with frame duration of  $T_w$  and frame shift of  $T_w$  and the spectrogram is created with the same frame duration and shift, (c) as in b, but the spectrogram uses a frame shift of 1 ms, (d) the stimulus is constructed with frame duration of  $T_w$  and frame shift of  $T_w/8$  and the spectrogram is created with frame duration of  $T_w$  and frame shift of  $T_w$ , and (e) as in d, but the spectrogram uses a frame shift of 1 ms.

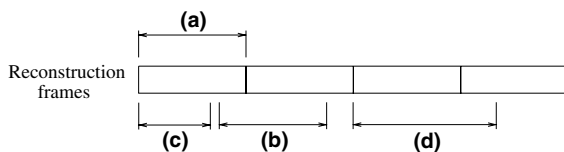


Fig. A.2. Reconstructed frames placed end to end (i.e., no overlap). For the spectrogram computation, the reconstructed signal can be analysed (a) in synchrony, (b) out of synchrony, (c) at a smaller frame duration, or (d) at a longer frame duration.

is not enforced (e.g., Fig. A.2(b)). In addition, no unity-magnitude spectrum constraint exists at

spectrogram frame durations less than and greater than 32 ms (e.g., Fig. A.2(c) and (d)).

Next, we construct another phase-only signal using a rectangular window with the same duration ( $T_w = 32$  ms), but change the frame shift to  $T_w/8$ . Again, we create a spectrogram for this signal with a window duration of 32 ms and frame shift of 32 ms (Fig. A.1(d)). Unlike Fig. A.1(b), we see formant structure, which comes from overlapping and adding of the reconstructed frames. By using a spectrogram frame shift of 1 ms, we obtain a better view of the formant structure (Fig. A.1(e)).

## References

- Allen, J.B., 1977. Short-term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* ASSP-25 (3), 235–238.
- Allen, J.B., Rabiner, L.R., 1977. A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE* 65 (11), 1558–1564.
- Alsteris, L.D., Paliwal, K.K., 2004. Importance of window shape for phase-only reconstruction of speech. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, Montreal, Canada, May 2004, vol. I, pp. 573–576.
- Cox, R.C., Robinson, D.M., 1980. Some notes on phase in speech signals. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, April 1980, pp. 150–153.
- Crochiere, R.E., 1980. A weighted overlap-add method of short-time Fourier analysis/synthesis. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-28 (1), 99–102.
- Espy, C.Y., Lim, J.S., 1983. Effects of additive noise on signal reconstruction from Fourier transform phase. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-31 (4), 894–898.
- Flanagan, J.L., Golden, R.M., 1966. Phase vocoder. *Bell Syst. Tech.* 45, 1493–1509.
- Goldstein, J.L., 1967. Auditory spectral filtering and monaural phase perception. *J. Acoust. Soc. Amer.* 41 (2), 458–479.
- Griffin, D.W., Lim, J.S., 1984. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-32 (2), 236–243.
- Hayes, M.H., Lim, J.S., Oppenheim, A.V., 1980. Signal reconstruction from phase or magnitude. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-28 (6), 672–680.
- Izraelevitz, D., 1985. Some results on the time–frequency sampling of the short-time Fourier transform magnitude. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-33 (6), 1611–1613.
- Kim, D.S., 2000. Perceptual phase redundancy in speech. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, June 2000, pp. 1383–1386.
- Lim, J.S., Oppenheim, A.V., 1979. Enhancement and bandwidth compression of noisy speech. *Proc. IEEE* 67, 1586–1604.
- Liu, L., He, J., Palm, G., 1997. Effects of phase on the perception of intervocalic stop consonants. *Speech Comm.* 22 (4), 403–417.
- Mathes, R.C., Miller, R.L., 1947. Phase effects in monaural perception. *J. Acoust. Soc. Amer.* 19 (5), 780–797.
- Merchant, G.A., Parks, T.W., 1983. Reconstruction of signals from phase: efficient algorithms, segmentation, and generalisations. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-31 (5), 1135–1147.
- Nawab, S.H., Quatieri, T.F., Lim, J.S., 1983. Signal reconstruction from short-time Fourier transform magnitude. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-31 (4), 986–998.
- Ohm, G.S., 1843. Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann. Phys. Chem.* 59, 513–565.
- Oppenheim, A.V., Lim, J.S., 1981. The importance of phase in signals. *Proc. IEEE* 69, 529–541.
- Oppenheim, A.V., Schaffer, R.W., 1975. *Digital Signal Process.*. Prentice-Hall, Englewood Cliffs, NJ.
- Paliwal, K.K., 2003. Usefulness of phase in speech processing. *Proc. IPSJ Spoken Language Process. Workshop*, Gifu, Japan, February 2003, pp. 1–6.
- Paliwal, K.K., Alsteris, L., 2003. Usefulness of phase spectrum in human speech perception. In: *Proc. Eurospeech*, Geneva, Switzerland, September 2003, pp. 2117–2120.
- Paliwal, K.K., Atal, B.S., 2003. Frequency-related representation of speech. In: *Proc. Eurospeech*, Geneva, Switzerland, September 2003, pp. 65–68.
- Patterson, R.D., 1987. A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Amer.* 82 (5), 1560–1586.
- Picone, J.W., 1993. Signal Modeling techniques in speech recognition. *Proc. IEEE* 81 (9), 1215–1247.
- Plomp, R., Steeneken, H.J.M., 1969. Effect of phase on the timbre of complex tones. *J. Acoust. Soc. Amer.* 46, 409–421.
- Pobloth, H., Kleijn, W.B., 1999. On phase perception in speech. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, March 1999, pp. 29–32.
- Portnoff, M.R., 1976. Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-24 (3), 243–248.
- Portnoff, M.R., 1979. Magnitude-phase relationships for short-time Fourier transforms based on Gaussian analysis windows. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, Washington, DC, April 1979, pp. 186–189.
- Portnoff, M.R., 1980. Time–frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-28 (1), 55–69.
- Portnoff, M.R., 1981. Short-time Fourier analysis of sampled speech. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-29 (3), 364–373.
- Portnoff, M.R., 1981. Time-scale modification of speech based on short-time Fourier analysis. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-29 (3), 374–390.
- Quatieri, T.F., 2002. *Discrete-time speech signal processing*. Prentice-Hall, Upper Saddle River, NJ.
- Quatieri, J.E., Oppenheim, A.V., 1981. Iterative techniques for minimum phase signal reconstruction from phase or magnitude. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-29 (6), 1187–1193.
- Rabiner, L.R., Schaffer, R.W., 1978. *Discrete-time Speech Signal Processing, Principles and Practice*. Prentice-Hall, Englewood Cliffs, NJ.
- Reddy, N.S., Swamy, M.N.S., 1985. Derivative of phase spectrum of truncated autoregressive signals. *IEEE Trans. Circuits Syst. CAS-32* (6).
- Schafer, R.W., Rabiner, L.R., 1973. Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis. *IEEE Trans. Audio Electroacoust.* AU-21, 165–174.

- Schroeder, M.R., 1959. New results concerning monaural phase sensitivity. *J. Acoust. Soc. Am.* 31, 1579.
- Schroeder, M.R., 1975. Models of hearing. *Proc. IEEE* 63, 1332–1350.
- Schroeder, M.R., Strube, H.W., 1986. Flat-spectrum speech. *J. Acoust. Soc. Am.* 79 (5), 1580–1583.
- Thomas, D.M., Hayes, M.H., 1984. Procedures for signal reconstruction from noisy phase. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, March 1984.
- Van Hove, P.L., Hayes, M.H., Lim, J.S., Oppenheim, A.V., 1983. Signal reconstruction from signed Fourier transform magnitude. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-31 (5), 1286–1293.
- von Helmholtz, H.L.F., 1912. *On the Sensations of Tone*, 1875 (Ellis, A.J., Longmans, Green and Co., London, English Trans.).
- Wang, D.L., Lim, J.S., 1982. The unimportance of phase in speech enhancement. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-30 (4), 679–681.
- Yegnanarayana, B., Saikia, D.K., Krishnan, T.R., 1984. Significance of group delay functions in signal reconstruction from spectral magnitude or phase. *IEEE Trans. Acoust., Speech Signal Process.* ASSP-32 (3), 610–623.
- Yegnanarayana, B., Tanveer Fathima, S., Murthy, H.A., 1987. Reconstruction from Fourier transform phase with applications to speech analysis. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.*, April 1987, pp. 301–304.