Speech Phase Spectrum; Love it or Leave it? Part I

Erfan Loweimi

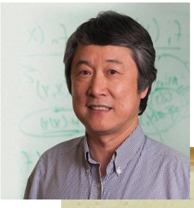
Centre for Speech Technology Research (CSTR)

My MSc Supervisors

Amirkabir
University of Technology

MSc dissertation: *On the Importance of Phase in Robust Speech Recognition*


Professor
Mohammad Ahadi




Professor
Hamid Sheikhzadeh

Speech Processing Research Lab

My PhD Supervisors

The University Of Sheffield

Professor
Jon Barker

Professor
Thomas Hain

SPANDH

PhD Thesis: *“Robust Phase-based Speech Signal Processing; From Source-Filter Separation to Model-Based Robust ASR”*

My PhD Thesis

- Phase-based Speech Signal Processing
  - *Source-filter Separation* in the Phase domain (IS15)
  - Phase Filter Component → feature extraction for ASR (ICASSP17, IS17)
  - Phase Source Component → F_0 estimation (IS18)

- Model-based Robust ASR
  - *Generalised Vector Taylor Series* (gVTS) for robust ASR (IS2016, IS2017)
  - *Blind Channel Noise Estimation* (IS2017)
  - Extension to product spectrum domain (ICASSP18)

My PhD Thesis

- Phase-based Speech Signal Processing
- Model-based Robust ASR

Outlines

- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction
- Source-filter Separation in the Phase Domain
- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F_0 Extraction

Outlines

- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction

Part I

- Source-filter Separation in the Phase Domain

Part II

- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F0 Extraction

Outlines

- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction — Part I
- Source-filter Separation in the Phase Domain — Part II
- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F0 Extraction

(1) Historical Bias

- Ohm’s acoustic (phase) law (1843)

Georg Simon
Ohm
(1789 - 1854)

(1) Historical Bias

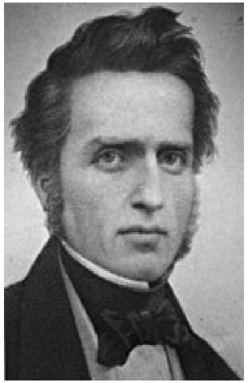
- Ohm's acoustic (phase) law (1843)

"*The perceived quality of a* *__tone__* *depends solely on the number and* relative ***strength*** *of its partial simple tones, and not on their* relative ***phases***"

(1) Historical Bias

- Ohm's acoustic (phase) law (1843)

Georg Simon Ohm
(1789 - 1854)



August Seebeck
(1805 - 1849)

THE OHM-SEEBECK DISPUTE, HERMANN VON HELMHOLTZ, AND THE ORIGINS OF PHYSIOLOGICAL ACOUSTICS

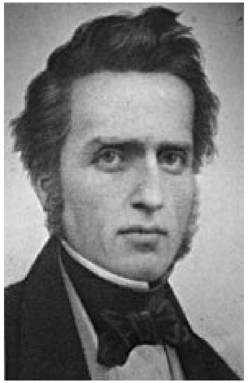
R. STEVEN TURNER*

THE term 'Ohm's law' traditionally denotes the formula of Georg Simon Ohm relating voltage, current, and resistance in metallic conductors. But to students of sensory physiology and its history, 'Ohm's law' also denotes another relationship: the fundamental principle of auditory perception that Ohm announced in 1843. This aspect of Ohm's science has attracted very little attention, partly because his galvanic researches so thoroughly eclipsed it in success and importance, and partly because

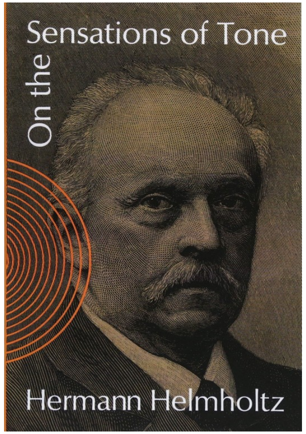
(1) Historical Bias

- Ohm’s acoustic (phase) law (1843)
- Verified by Helmholtz (1875)

Georg Simon Ohm
(1789 - 1854)



August Seebeck
(1805 - 1849)



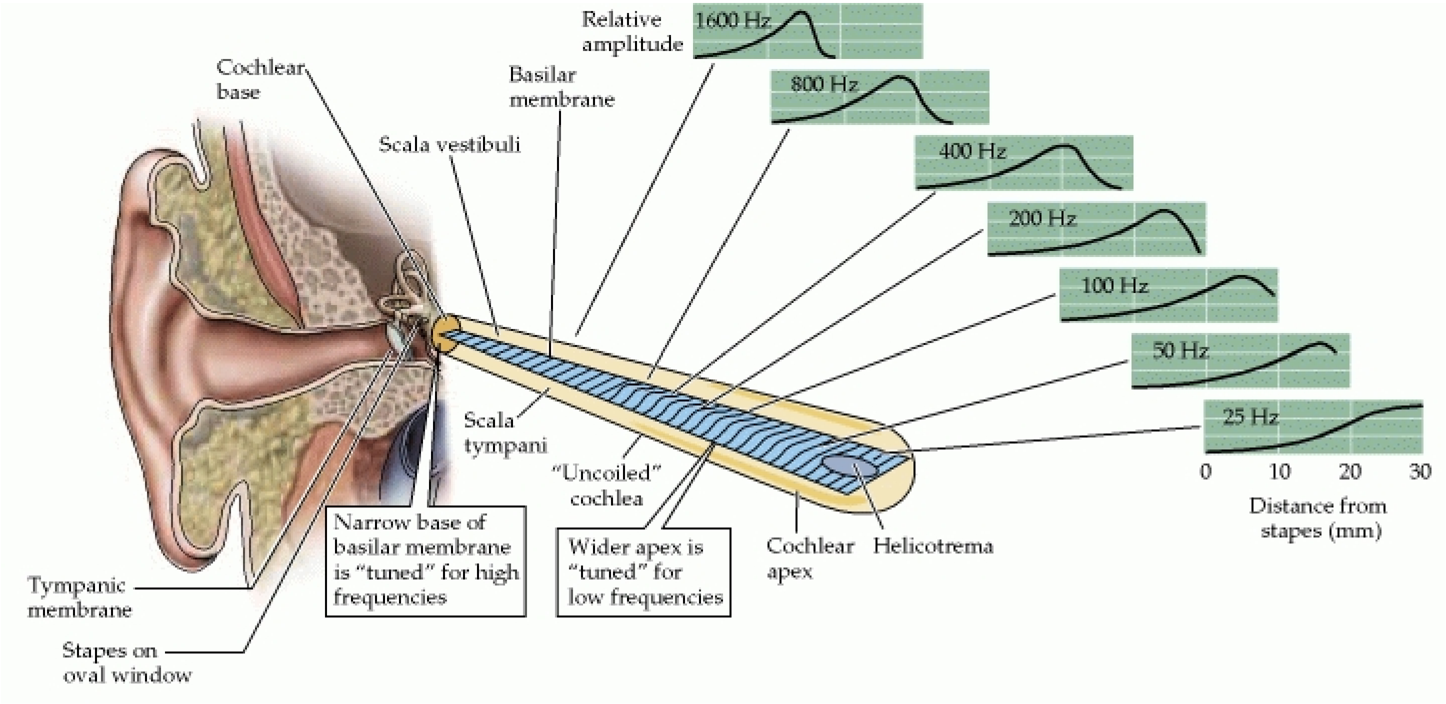
Hermann von Helmholtz
(1821 - 1894)

(1) Historical Bias

- Ohm’s acoustic (phase) law (1843)
- Verified by Helmholtz (1875)

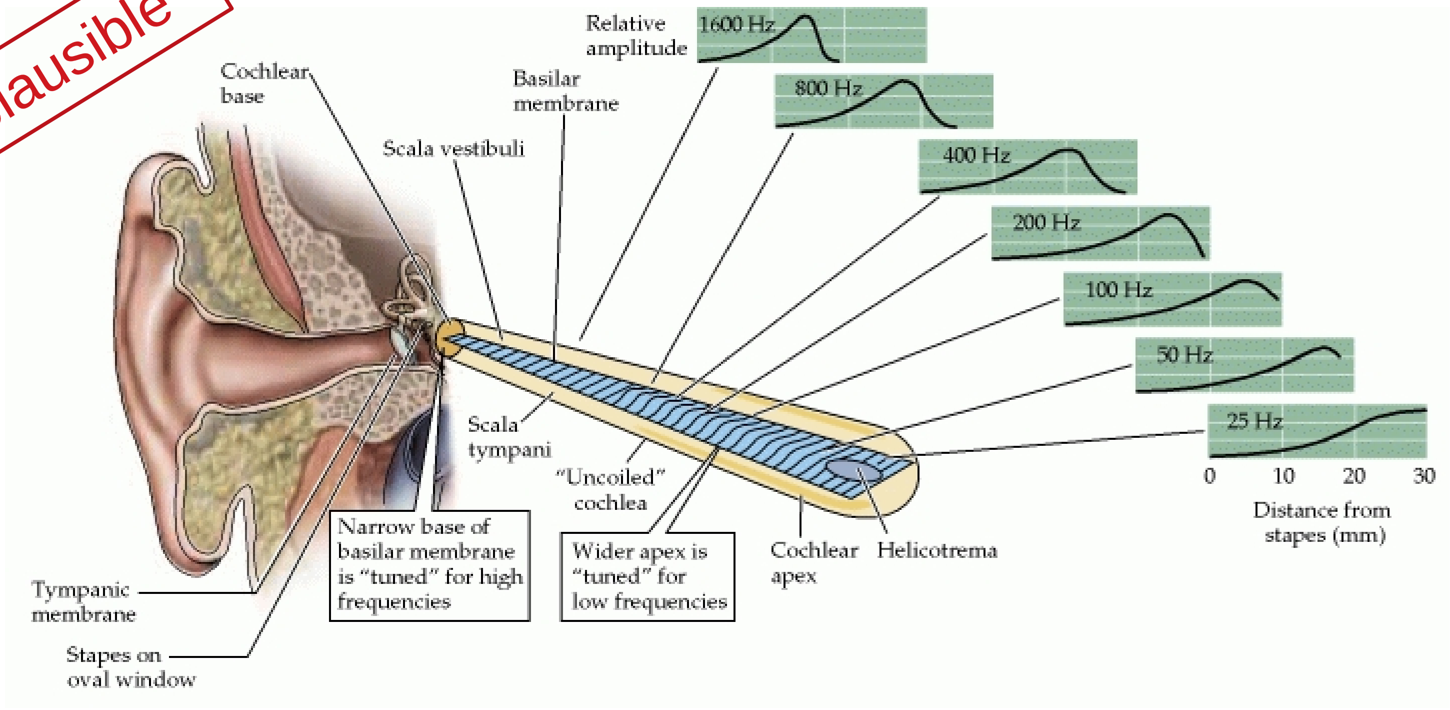
Human auditory system is *phase-deaf*

Cochlea, *approximately*, computes the magnitude spectrum ...



Cochlea, *approximately*, computes the magnitude spectrum ...

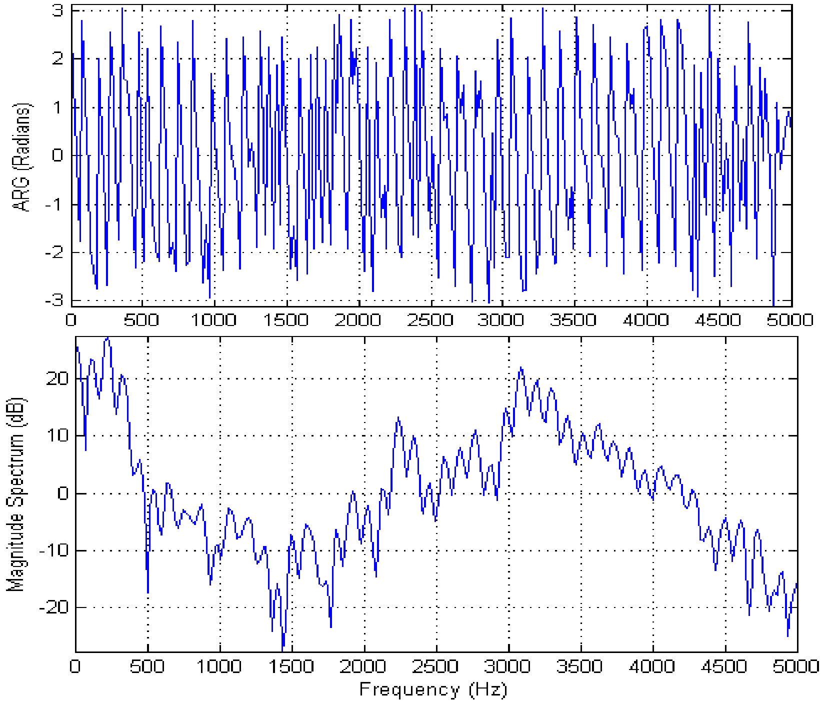
Biologically plausible



(2) Chaotic and Noise-like Shape

Principle Phase (ARG)

Magnitude Spectrum (dB)



(2) Chaotic and Noise-like Shape

- Cause
  - Phase wrapping (due to *arctan* function)

$$\phi_X = arctan(\frac{X_{Im}(\omega)}{X_{Re}(\omega)})$$

(2) Chaotic and Noise-like Shape

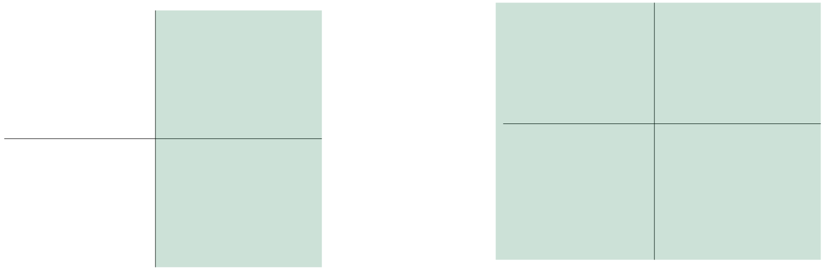
- Cause
  - Phase wrapping (due to *arctan2* function)

$$\phi_X(\omega) = arctan2(X_{Im}(\omega), X_{Re}(\omega))$$

(2) Chaotic and Noise-like Shape

- Cause
  - Phase wrapping (due to *arctan2* function)

$$\phi_X(\omega) = arctan2(X_{Im}(\omega), X_{Re}(\omega))$$

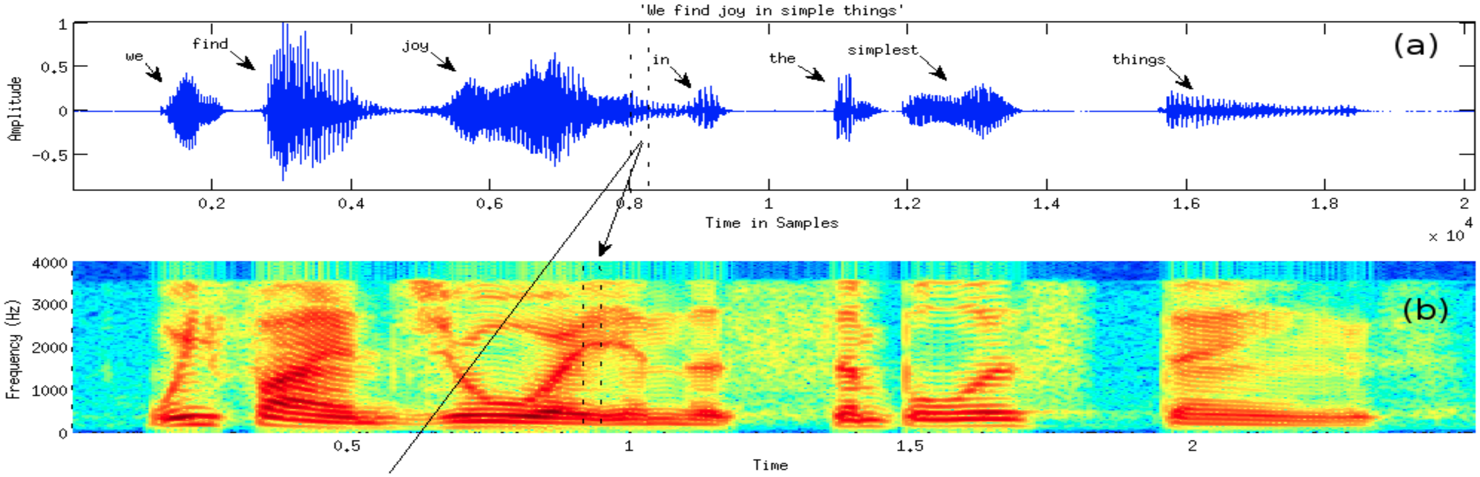


$(-\frac{\pi}{2}, \frac{\pi}{2}] \longleftarrow$ arctan(Im/Re)   arctan2(Im,Re) $\longrightarrow (-\pi, \pi]$

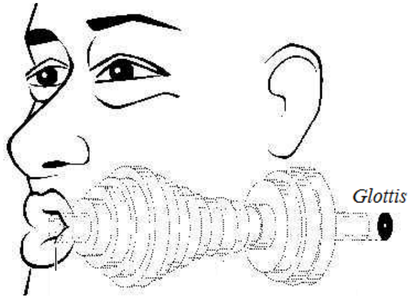
(2) Chaotic and Noise-like Shape

- Cause
  - Phase wrapping (due to arctan2 function)

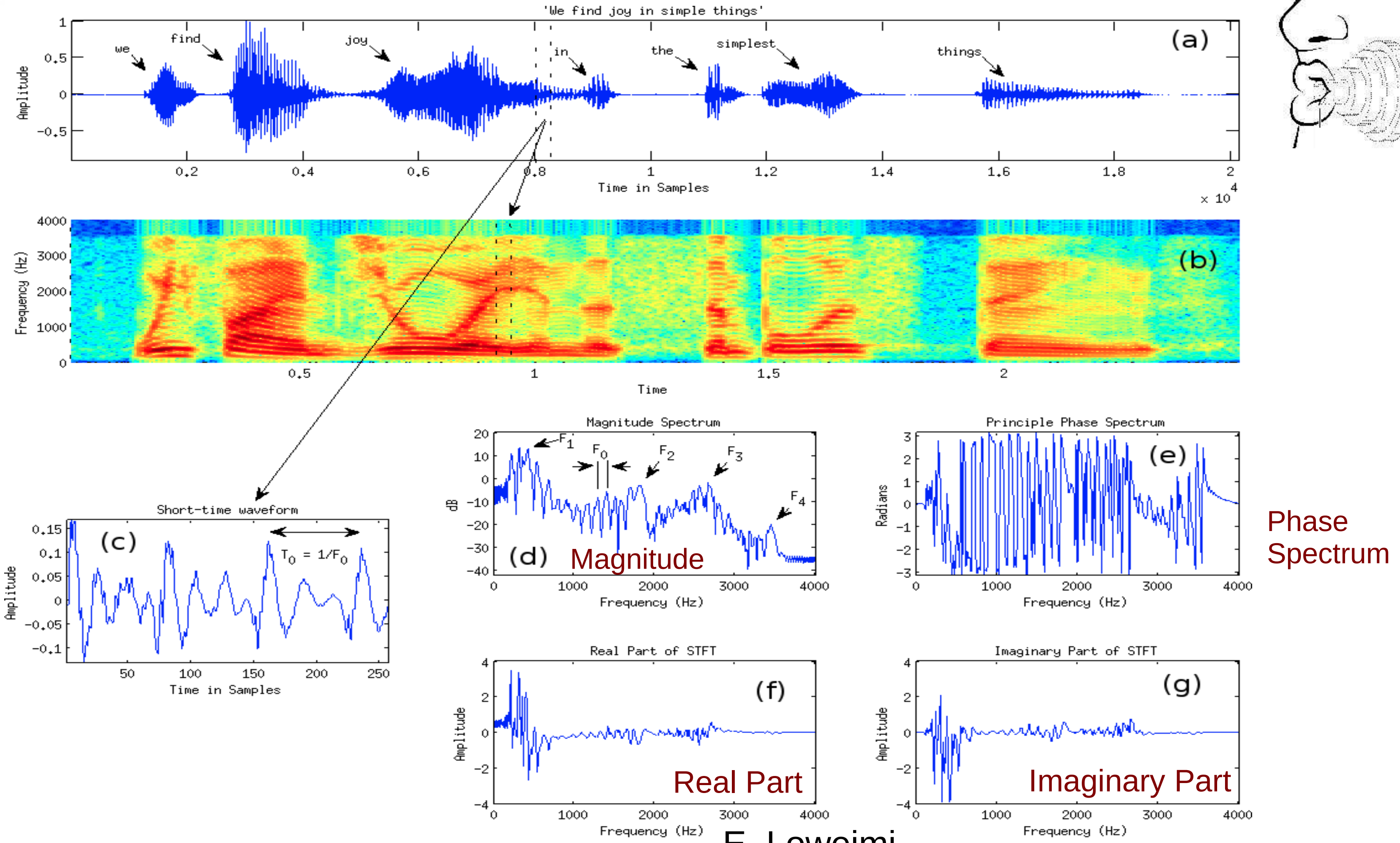
- Effect → Complicates ...
  - Physical interpretation
  - Mathematical modelling

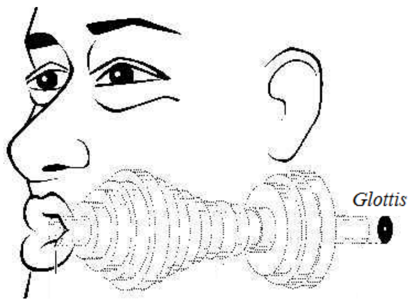


'We find joy in simple things'
we
find
joy
in
the
simplest
things
(a)
Amplitude
Time in Samples
× 10^4
(b)
Frequency (Hz)
Time



Glottis

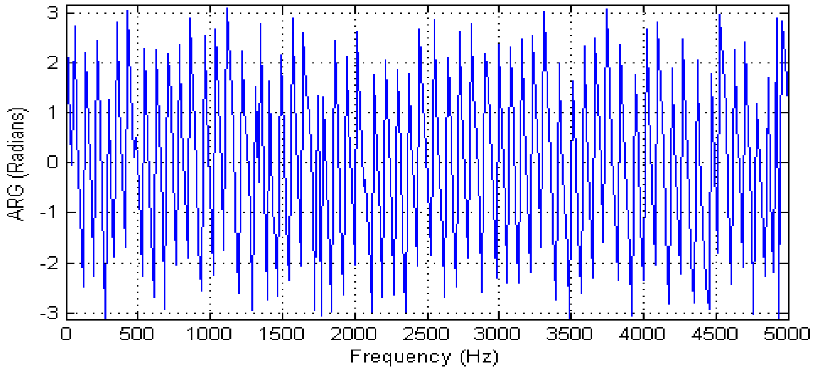

'We find joy in simple things'
we
find
joy
in
the
simplest
things
(a)
Amplitude
Time in Samples
× 10^4
(b)
Frequency (Hz)
Time
Short-time waveform
(c)
$T_0 = 1/F_0$
Amplitude
Time in Samples
Magnitude Spectrum
F_1
F_0
F_2
F_3
F_4
(d)
Magnitude
dB
Frequency (Hz)
Principle Phase Spectrum
(e)
Radians
Frequency (Hz)
Phase Spectrum
Real Part of STFT
(f)
Real Part
Amplitude
Frequency (Hz)
Imaginary Part of STFT
(g)
Imaginary Part
Amplitude
Frequency (Hz)

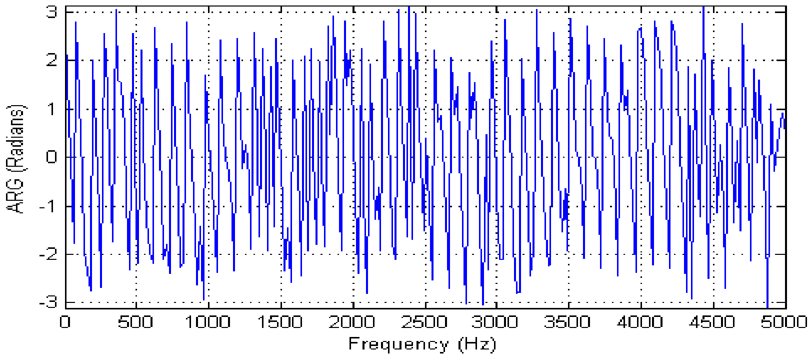

Glottis

E. Loweimi

(2) Noise-like from Correlation Viewpoint

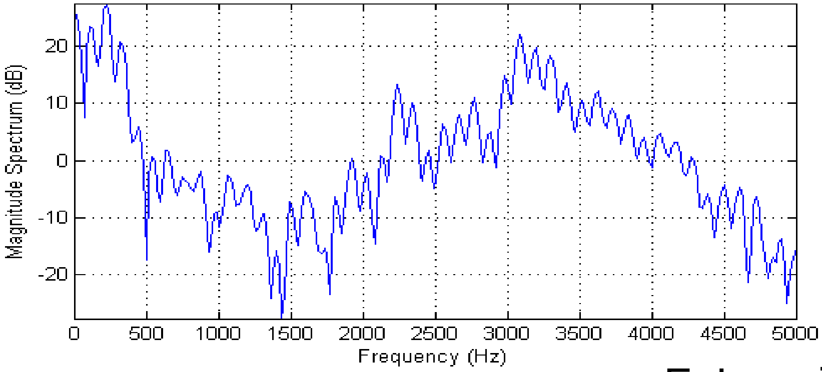
Corr-coef:
-0.0272

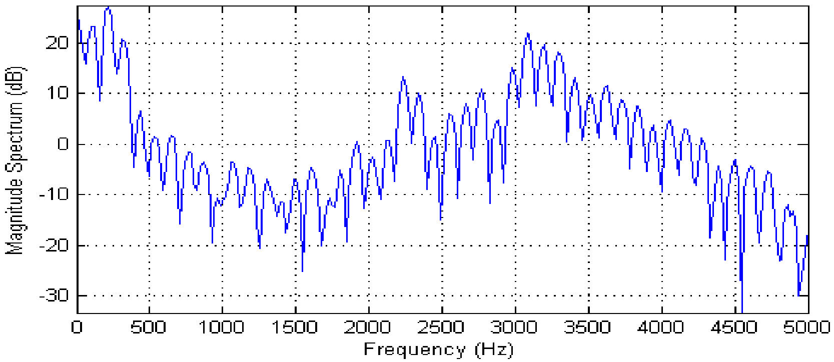




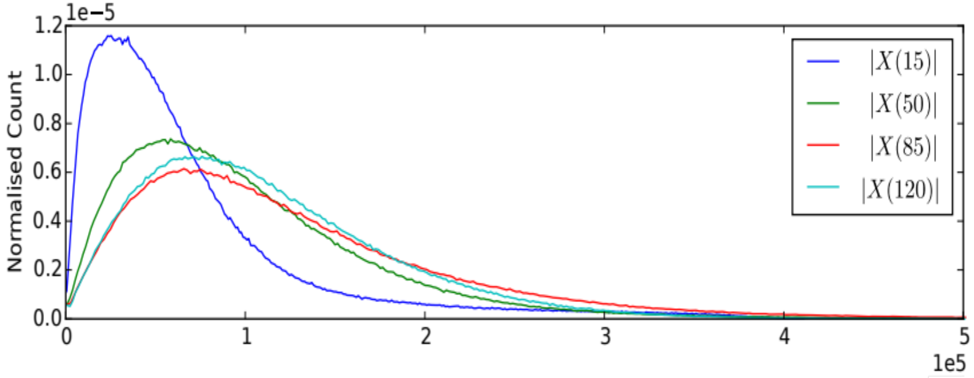
Successive frames with 1 ms shift

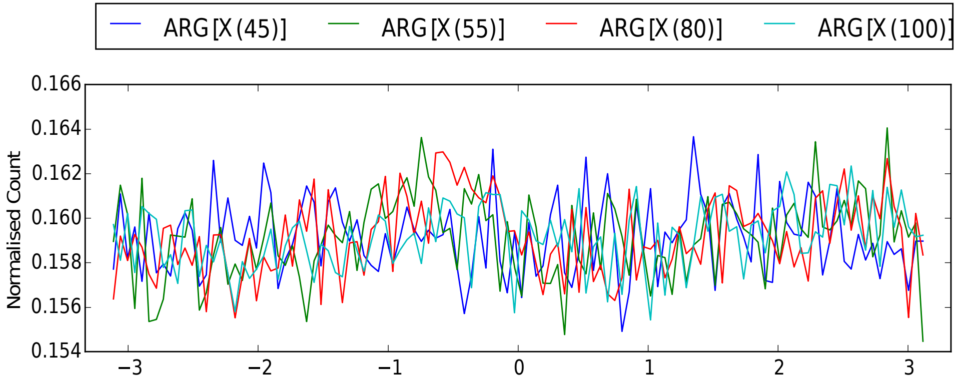
Corr-coef:
+0.9904





(2) Noise-like from Statistical Viewpoint



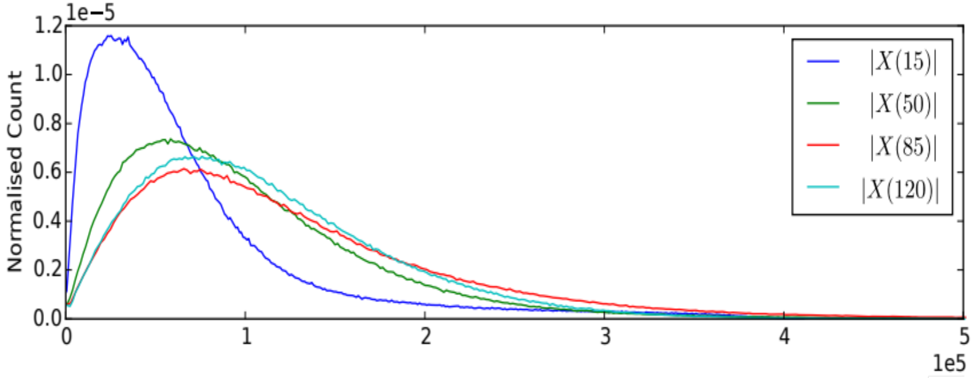


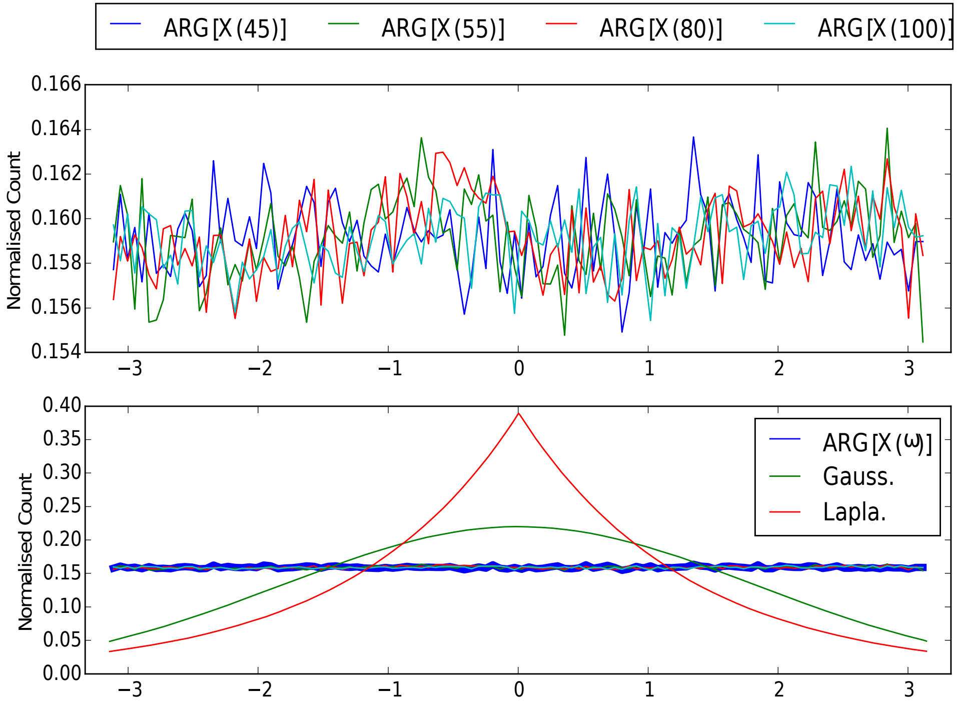
Histogram of
Magnitude Spectrum

Histogram of
Phase Spectrum

Estimated using 1.47 M frames

(2) Noise-like from Statistical Viewpoint





Uniform Distribution → Maximum Entropy

(3) Low Information Content in Short-term Analysis

- Approach: Phase-only reconstructed signal vs Original

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. 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

ICASSP 1979

PROCEEDINGS OF THE IEEE, VOL. 69, NO. 5, MAY 1981 529

The Importance of Phase in Signals

ALAN V. OPPENHEIM, FELLOW, IEEE, AND JAE S. LIM, MEMBER, IEEE

Invited Paper

Abstract—In the Fourier representation of signals, spectral magnitude and phase tend to play different roles and in some situations many of the important features of a signal are preserved if only the phase is retained. Furthermore, under a variety of conditions, such as when a signal is of finite length, phase information alone is sufficient to completely reconstruct a signal to within a scale factor. In this paper, we review and discuss these observations and results in a number of different contexts and applications. Specifically, the intelligibility of phase-only reconstruction for images, speech, and crystallographic structures are illustrated. Several approaches to justifying the relative importance of phase through statistical arguments are presented, along with a number of informal arguments suggesting reasons for the importance of phase. Specific conditions under which a sequence can be exactly reconstructed from phase are reviewed, both for one-dimensional and multidimensional sequences, and algorithms for both approximate and exact reconstruction of signals from phase information are presented. A number of applications of the observations and results in this paper are suggested.

I. INTRODUCTION

IN THE FOURIER representation of signals, spectral magnitude and phase tend to play different roles and in some situations, many of the important features of a signal are but not in the magnitude-only image. Similar observations have also been made in the context of speech signals and X-ray crystallography. Specifically, for speech it has been shown that the intelligibility of a sentence is retained if the phase of the Fourier transform of a long segment of speech is combined with unity magnitude. In the context of X-ray crystallography, details of the crystallographic structure are often inferred from X-ray diffraction data. The Fourier synthesis of the structure from only the correct magnitude of the diffraction data with zero phase in general does not preserve the atomic structure whereas Fourier synthesis using only the correct phase with unity magnitude does reflect the correct atomic structure. These examples, elaborated on in Section II, suggest very strongly the fact that in many contexts the phase contains much of the essential "information" in a signal.

The above discussion relates to the fact that if the true magnitude information is eliminated many of the important characteristics of the signal are nevertheless retained. In the experiments outlined above, the true magnitude information is simply replaced by a standard magnitude. With so much intelligibility incorporated in the phase, it is natural to consider the possi-

1981

(3) Low Information Content in Short-term Analysis

- Approach: Phase-only signal vs Original

- For speech, opposite to the magnitude spectrum ...
  - Low information content in short-term
  - High information content in long-term

PROCEEDINGS OF THE IEEE, VOL. 69, NO. 5, MAY 1981 529

The Importance of Phase in Signals

ALAN V. OPPENHEIM, FELLOW, IEEE, AND JAE S. LIM, MEMBER, IEEE

Invited Paper

Abstract–In the Fourier representation of signals, spectral magnitude and phase tend to play different roles and in some situations many of the important features of a signal are preserved if only the phase is retained. Furthermore, under a variety of conditions, such as when a signal is of finite length, phase information alone is sufficient to completely reconstruct a signal to within a scale factor. In this paper, we review and discuss these observations and results in a number of different contexts and applications. Specifically, the intelligibility of phase-only reconstruction for images, speech, and crystallographic structures are illustrated. Several approaches to justifying the relative importance of phase through statistical arguments are presented, along with a number of informal arguments suggesting reasons for the importance of phase. Specific conditions under which a sequence can be exactly reconstructed from phase are reviewed, both for one-dimensional and multidimensional sequences, and algorithms for both approximate and exact reconstruction of signals from phase information are presented. A number of applications of the observations and results in this paper are suggested.

I. Introduction

IN THE FOURIER representation of signals, spectral magnitude and phase tend to play different roles and in some situations, many of the important features of a signal are

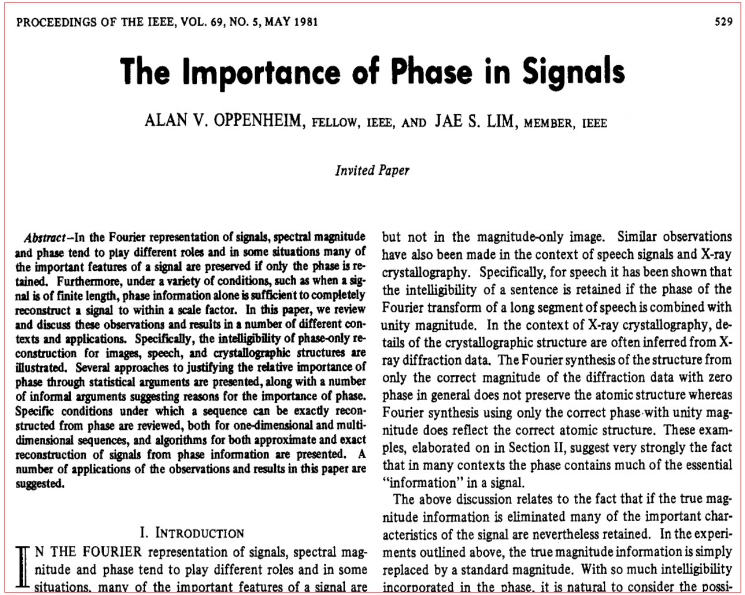
but not in the magnitude-only image. Similar observations have also been made in the context of speech signals and X-ray crystallography. Specifically, for speech it has been shown that the intelligibility of a sentence is retained if the phase of the Fourier transform of a long segment of speech is combined with unity magnitude. In the context of X-ray crystallography, details of the crystallographic structure are often inferred from X-ray diffraction data. The Fourier synthesis of the structure from only the correct magnitude of the diffraction data with zero phase in general does not preserve the atomic structure whereas Fourier synthesis using only the correct phase with unity magnitude does reflect the correct atomic structure. These examples, elaborated on in Section II, suggest very strongly the fact that in many contexts the phase contains much of the essential "information" in a signal.

The above discussion relates to the fact that if the true magnitude information is eliminated many of the important characteristics of the signal are nevertheless retained. In the experiments outlined above, the true magnitude information is simply replaced by a standard magnitude. With so much intelligibility incorporated in the phase, it is natural to consider the possi-

1981

(3) Low Information Content in Short-term Analysis

- Approach: Phase-only signal vs Original
- For speech, opposite to the magnitude spectrum ...
  - Low information content in short-term
  - High information content in long-term
- Phase is MORE important than magnitude for image reconstruction!



PROCEEDINGS OF THE IEEE, VOL. 69, NO. 5, MAY 1981 529

The Importance of Phase in Signals

ALAN V. OPPENHEIM, FELLOW, IEEE, AND JAE S. LIM, MEMBER, IEEE

Invited Paper

***Abstract*–In the Fourier representation of signals, spectral magnitude and phase tend to play different roles and in some situations many of the important features of a signal are preserved if only the phase is retained. Furthermore, under a variety of conditions, such as when a signal is of finite length, phase information alone is sufficient to completely reconstruct a signal to within a scale factor. In this paper, we review and discuss these observations and results in a number of different contexts and applications. Specifically, the intelligibility of phase-only reconstruction for images, speech, and crystallographic structures are illustrated. Several approaches to justifying the relative importance of phase through statistical arguments are presented, along with a number of informal arguments suggesting reasons for the importance of phase. Specific conditions under which a sequence can be exactly reconstructed from phase are reviewed, both for one-dimensional and multidimensional sequences, and algorithms for both approximate and exact reconstruction of signals from phase information are presented. A number of applications of the observations and results in this paper are suggested.**

I. INTRODUCTION

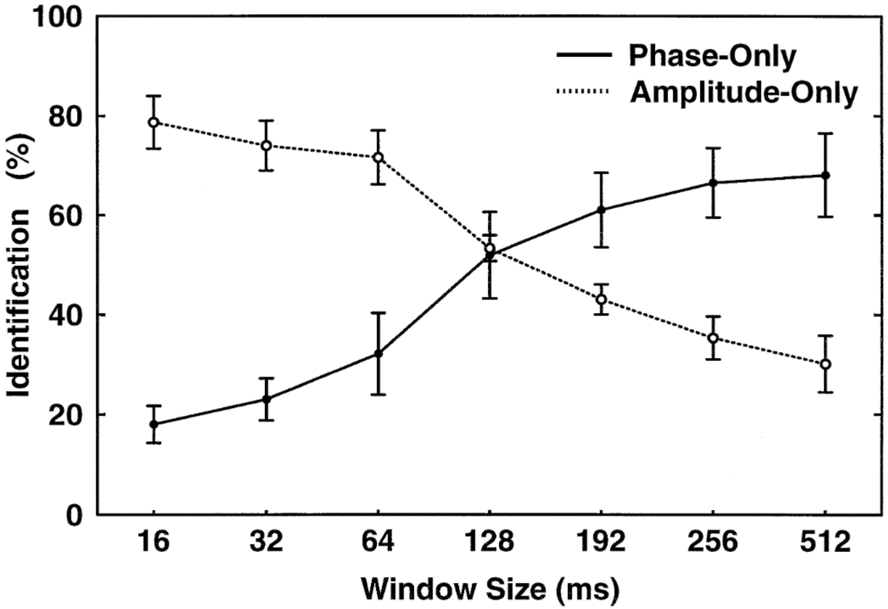
IN THE FOURIER representation of signals, spectral magnitude and phase tend to play different roles and in some situations, many of the important features of a signal are

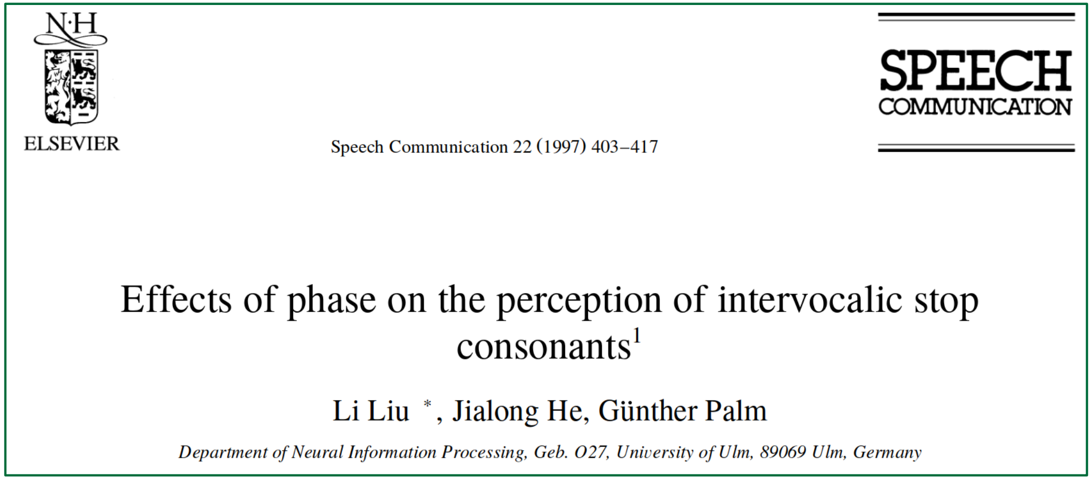
but not in the magnitude-only image. Similar observations have also been made in the context of speech signals and X-ray crystallography. Specifically, for speech it has been shown that the intelligibility of a sentence is retained if the phase of the Fourier transform of a long segment of speech is combined with unity magnitude. In the context of X-ray crystallography, details of the crystallographic structure are often inferred from X-ray diffraction data. The Fourier synthesis of the structure from only the correct magnitude of the diffraction data with zero phase in general does not preserve the atomic structure whereas Fourier synthesis using only the correct phase with unity magnitude does reflect the correct atomic structure. These examples, elaborated on in Section II, suggest very strongly the fact that in many contexts the phase contains much of the essential "information" in a signal.

The above discussion relates to the fact that if the true magnitude information is eliminated many of the important characteristics of the signal are nevertheless retained. In the experiments outlined above, the true magnitude information is simply replaced by a standard magnitude. With so much intelligibility incorporated in the phase, it is natural to consider the possi-

1981

(3) Low Information Content in Short-term Analysis





ELSEVIER

Speech Communication 22 (1997) 403–417

SPEECH COMMUNICATION

Effects of phase on the perception of intervocalic stop consonants[1]

Li Liu *, Jialong He, Günther Palm

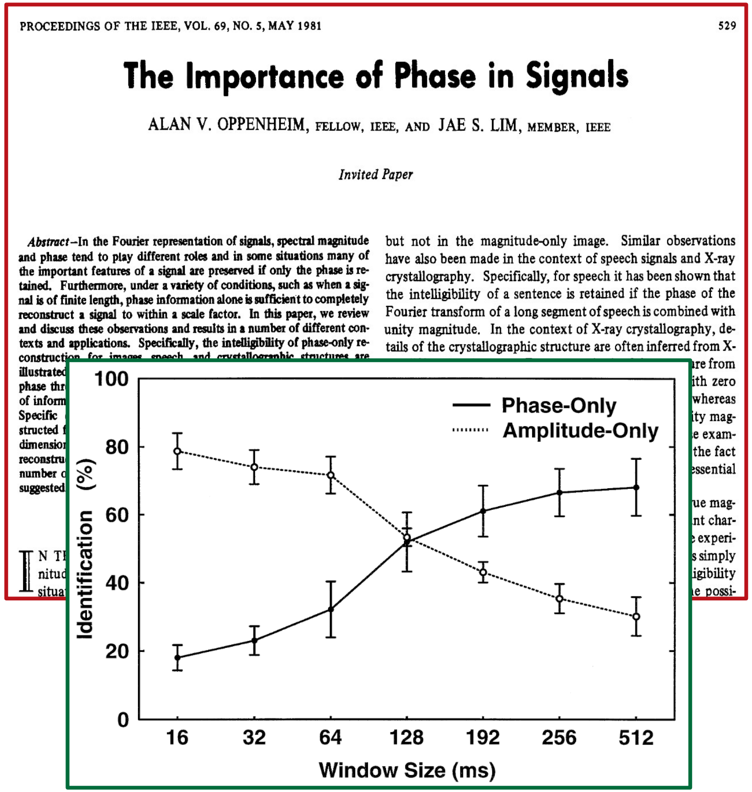
Department of Neural Information Processing, Geb. O27, University of Ulm, 89069 Ulm, Germany

1997

(3) Low Information Content in Short-term Analysis

- Approach: Phase-only signal vs Original
- Opposite to the magnitude spectrum ...
  - Low information content in short-term
  - High information content in long-term
- Due to non-stationarity ...

  … ONLY short-term is relevant!


PROCEEDINGS OF THE IEEE, VOL. 69, NO. 5, MAY 1981 529

The Importance of Phase in Signals

ALAN V. OPPENHEIM, FELLOW, IEEE, AND JAE S. LIM, MEMBER, IEEE

Invited Paper

Abstract—In the Fourier representation of signals, spectral magnitude and phase tend to play different roles and in some situations many of the important features of a signal are preserved if only the phase is retained. Furthermore, under a variety of conditions, such as when a signal is of finite length, phase information alone is sufficient to completely reconstruct a signal to within a scale factor. In this paper, we review and discuss these observations and results in a number of different contexts and applications. Specifically, the intelligibility of phase-only reconstruction for images, speech, and crystallographic structures are

but not in the magnitude-only image. Similar observations have also been made in the context of speech signals and X-ray crystallography. Specifically, for speech it has been shown that the intelligibility of a sentence is retained if the phase of the Fourier transform of a long segment of speech is combined with unity magnitude. In the context of X-ray crystallography, details of the crystallographic structure are often inferred from X-

Why bother? KEEP CALM AND …

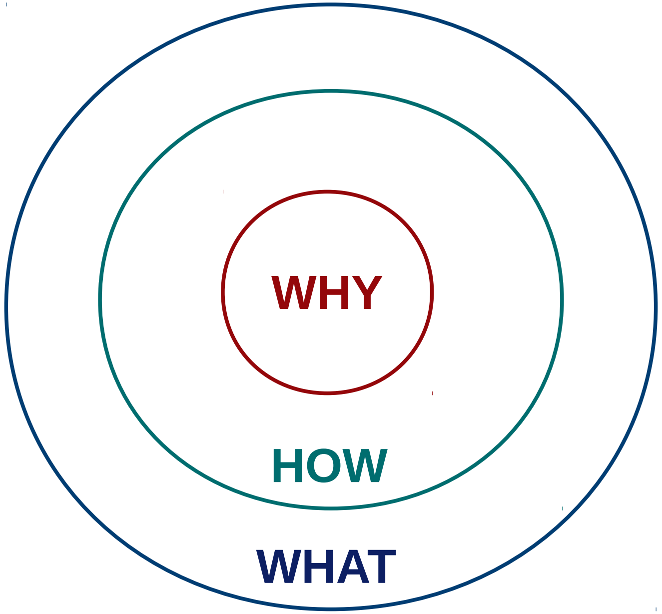
Why bother? Use Magnitude Spec.

- *Perceptually → Informative*
- *Physically → Understandable*
- *Mathematically → Modelling is easy*
- *Statistically → Tractable*

Common Theme of Most of the Phase-related Papers ..

- Use phase along with magnitude
- Report some minor improvement
- Say it carries complementary info
- Recommend for future researches

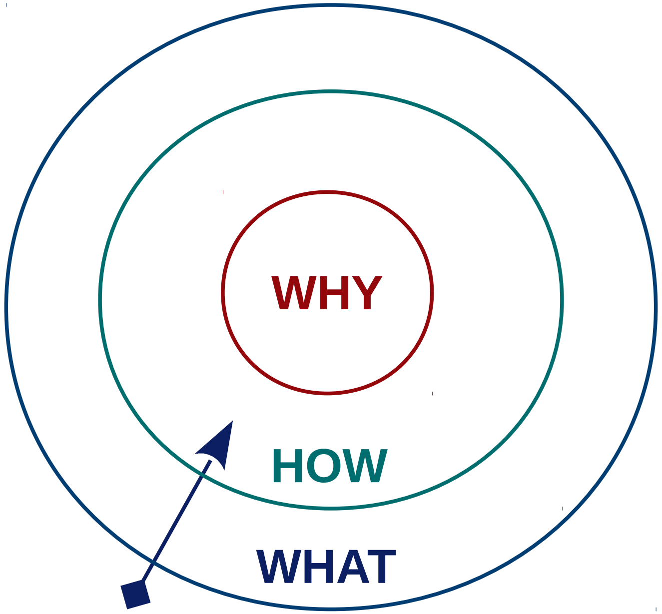
Common Theme of Most of the Phase-related Papers ..



Golden Circle

Simon Sinek

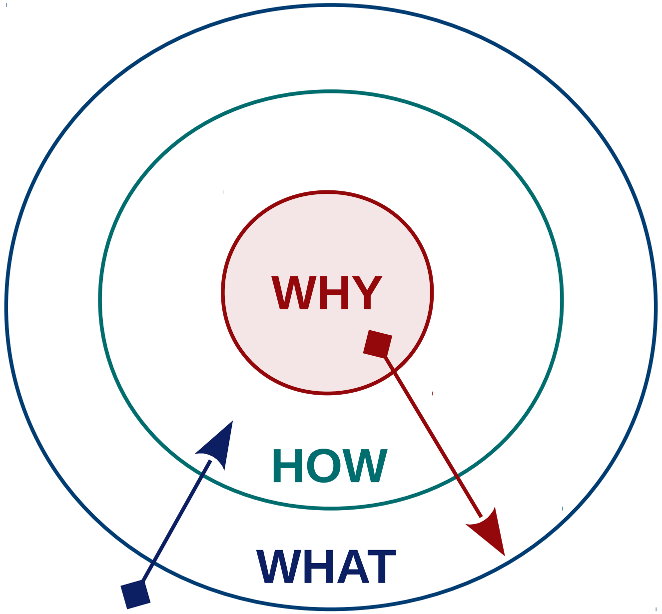
Common Theme of Most of the Phase-related Papers ..



Golden Circle

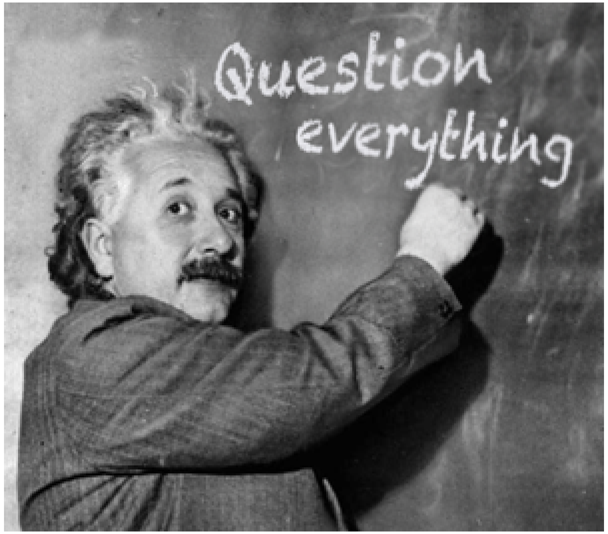
Simon Sinek

Common Theme of Most of the Phase-related Papers ..



Golden Circle

Simon Sinek



The important thing is not to stop questioning …
Never lose a holy curiosity.

Albert Einstein

(1879 - 1955)

Outlines

- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction — Part I

- Source-filter Separation in the Phase Domain — Part II
- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F0 Extraction

Central Questions ↔ Information

- How much/what kind of info gets encoded in phase?

- How is information distributed between phase/magnitude?
  - Which part of info is shared/unique to each one?

Central Questions ↔ Information

- How much/what kind of info gets encoded in phase?
- How is information distributed between phase/magnitude?
  - Which part of info is shared/unique to each one?

Central Questions ↔ Information

(1) Make sure sth is inside

(2) Look for a key to open

Phase Spectrum

Central Questions $\leftrightarrow$ Information

(1) Make sure sth is inside

(2) Look for a key to open

$\boxed{\text{Information}} \equiv$

Signal: Physical manifestation of information that changes with time and/or space

president's MESSAGE

José M.F. Moura
2008–2009 SPS President
j.moura@ieee.org

What Is Signal Processing?

I am returning to this perennial question that, in some sense, I already touched on in my March 2008 column. I am sure we are all often confronted with explaining signal processing to others, including nonexperts or the public in general. We may choose to explain it by what we and our colleagues do, as exemplified by the diversity and richness of the work of the IEEE Signal Processing Society's (SPS) technical committees and the topical coverage of our numerous solely sponsored or cosponsored conferences, workshops, and publications. Alternatively, we might turn to the cyber age oracle (Wikipedia) for an intuitive description (I suggest you read it).

societies). I am writing this in August, with you reading it in November; thus, I cannot predict whether or not all interested parties will converge.

Now, back to the challenge. The current state of our discussions (I want to acknowledge my five comrades in "crime": Leah Jamieson, Rich Cox, Mos

THE CURRENT STATE OF OUR DISCUSSIONS ARE CENTERED ON TWO OBVIOUS QUESTIONS, "WHAT DO WE MEAN BY *SIGNAL*?" AND "HOW DO WE DEFINE *PROCESSING*?"

cal, chemical, molecular, genomic, medical, musical, data, or sequences of attributes, or numerical quantities; the list goes on.

As for *processing*, it comprises operations of representing, filtering, coding, transmitting, estimating, detecting, inferring, discovering, recognizing, synthesizing, recording, or reproducing signals by digital or analog devices, techniques, or algorithms, in the form of software, hardware, or firmware (Did we leave out other important techniques?).

So, putting it together, can we say that *signal processing* is an enabling technology that encompasses the fundamental theory, applications, algorithms, and implementations of pro-

What is Information?

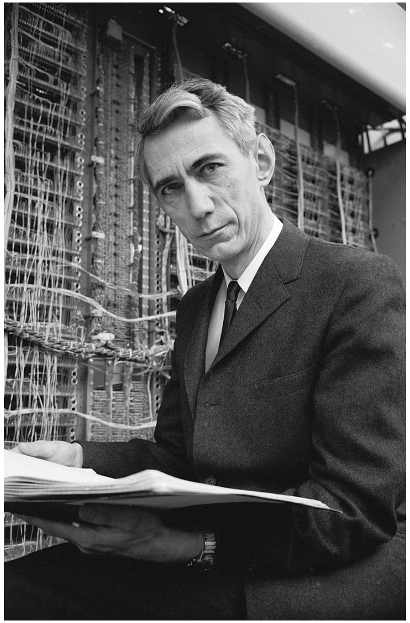
Reprinted with corrections from *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

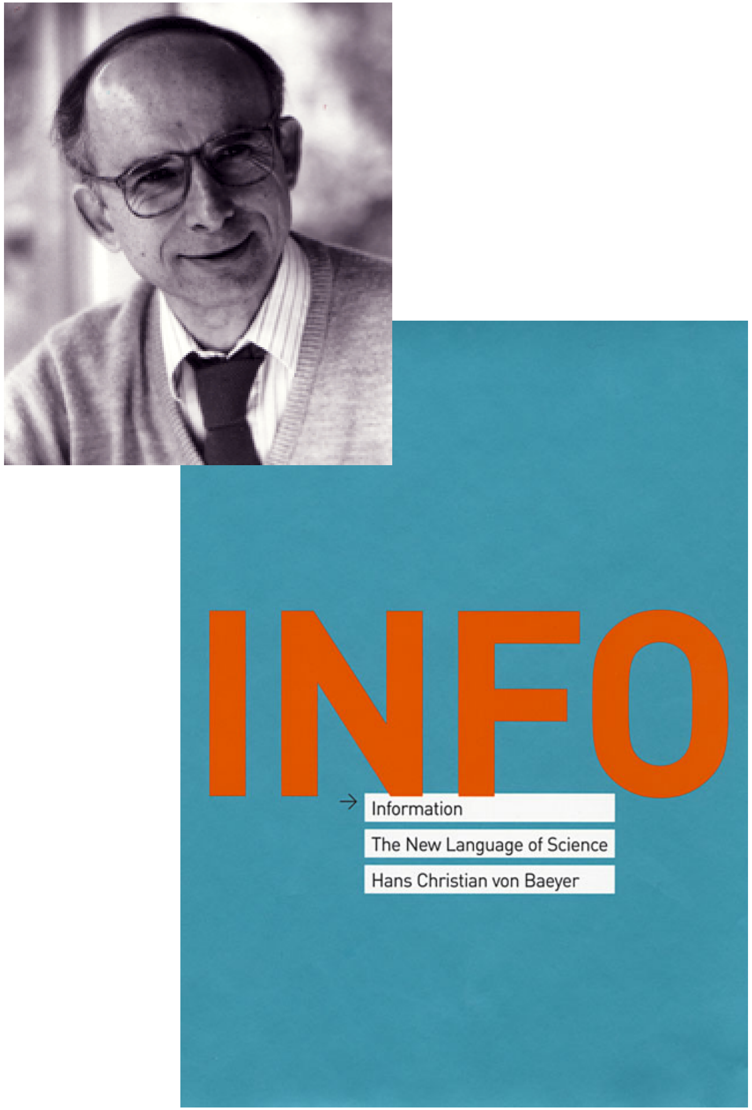
INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the



$$H(X) = -\sum_{x \in \mathcal{X}} P(x) \log P(x)$$

$$\text{Information} \equiv \text{Average Surprise}$$



Claude Shannon, the founder of information theory, invented a way to measure 'the amount of information' in a message without defining the word 'information' itself, nor even addressing the question of the meaning of the message.

Hans Christian Von Baeyer, “Information, The New Language of Science”, Chapter 4, p. 28, 2003.

What is Information?

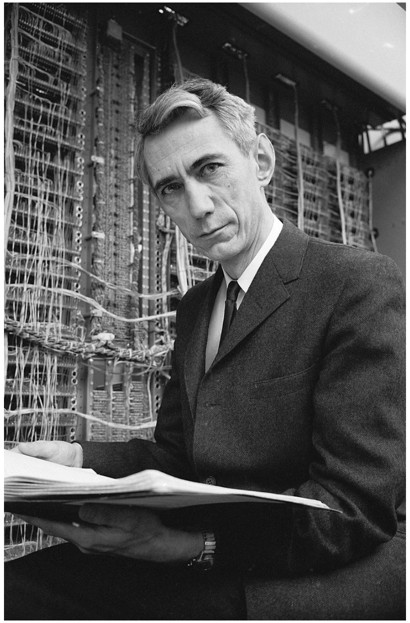
Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the

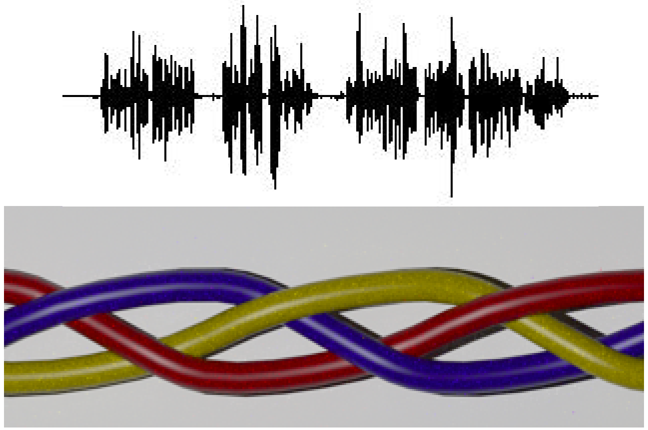


$$H(X) = -\sum_{x\in\mathcal{X}} P(x)\log P(x)$$

Objective definition of Information

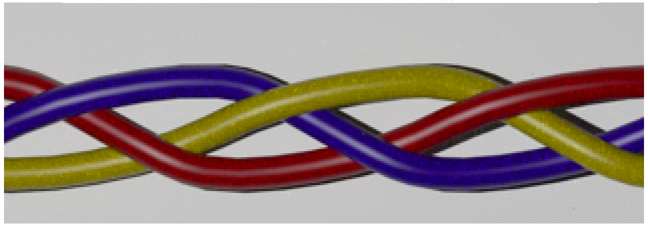
Info in Speech is rather Subjective ...

- Lingual Content
- Speaker
- Environment/Channel



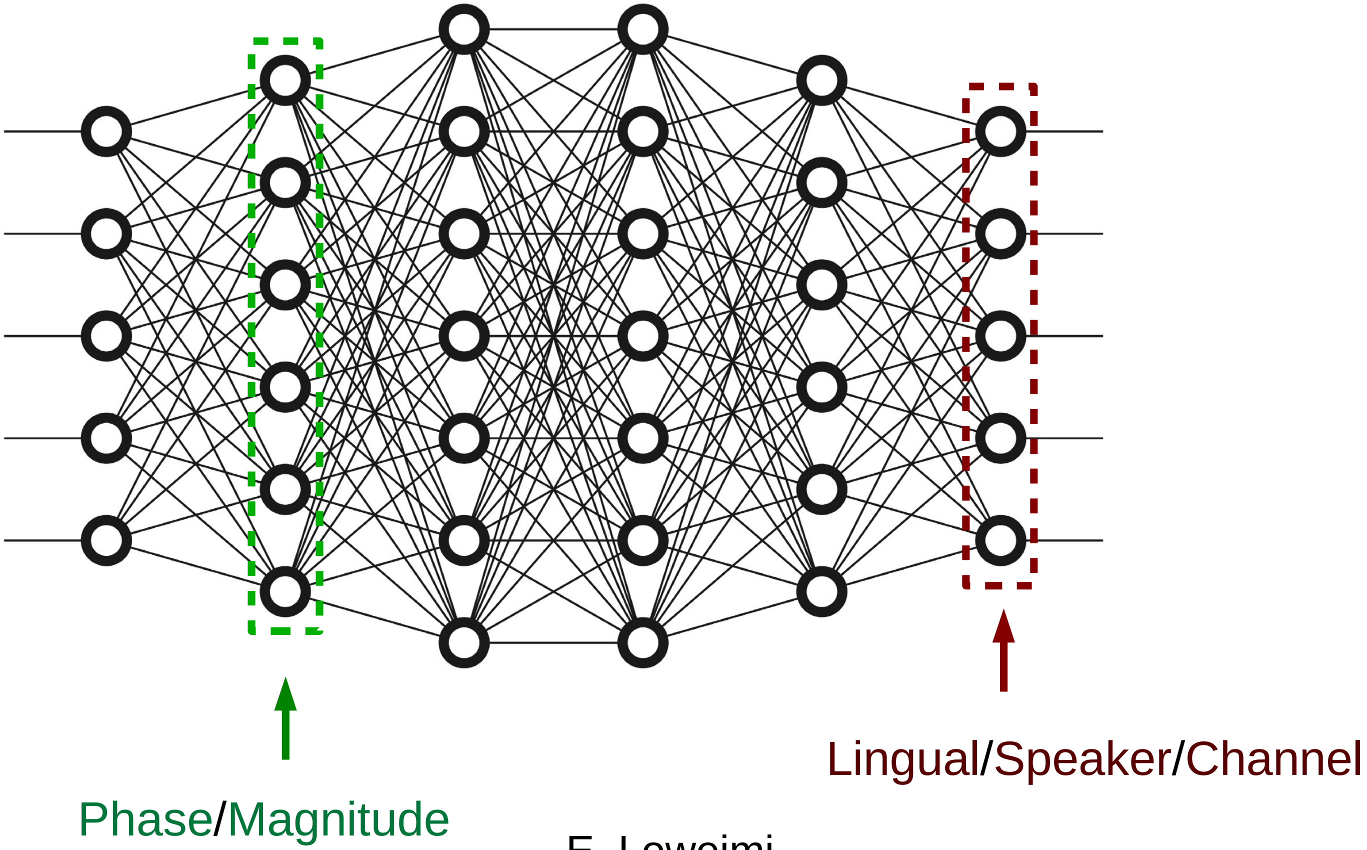
Info in Speech is rather Subjective ...

- Lingual Content
- Speaker
- Environment/Channel



Too abstract!

Abstraction Levels ...



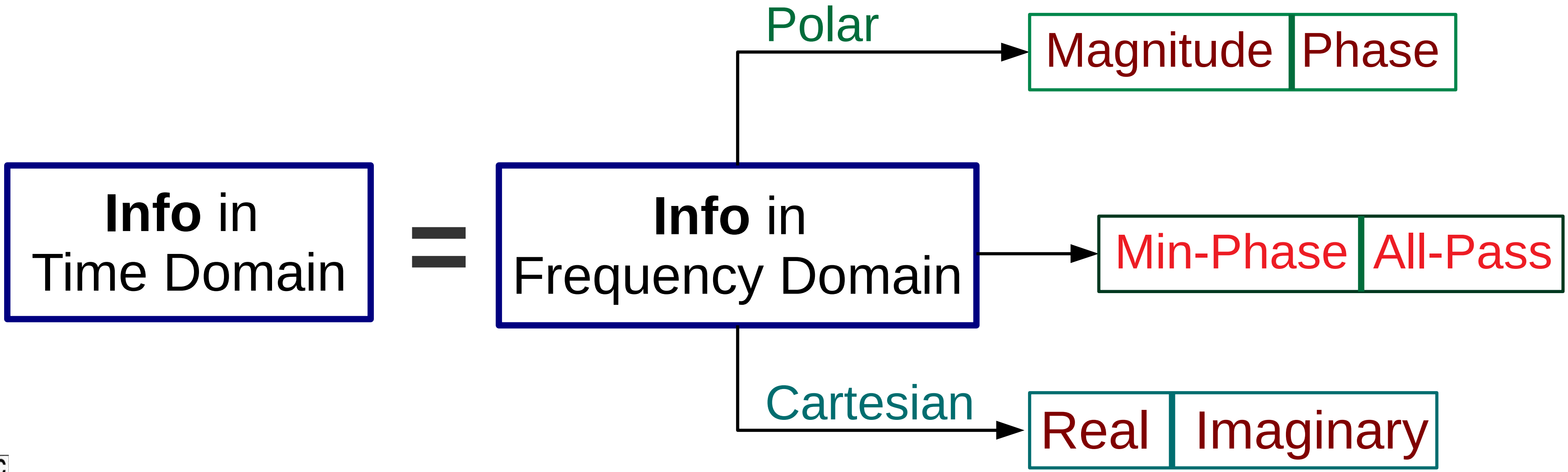
Lower Level View to Info

- Conservation of Information under Fourier Transform

Info in Time Domain	=	**Info** in Frequency Domain

Lower Level View to Info

- Conservation of Information under Fourier Transform



Information in Frequency Domain

- $I_{Total} = I_{Time\ domain} = I_{\ Frequency\ domain}$
  - $I_{Frequency\ domain} = I_{Re} \cup I_{Im}$
  - $I_{Frequency\ domain} = I_{Magnitude} \cup I_{Phase}$
  - $I_{Frequency\ domain} = I_{All\text{-}pass} \cup I_{Min\text{-}phase}$

I
N
F
O
R
M
A
T
I
O
N

Information in Frequency Domain

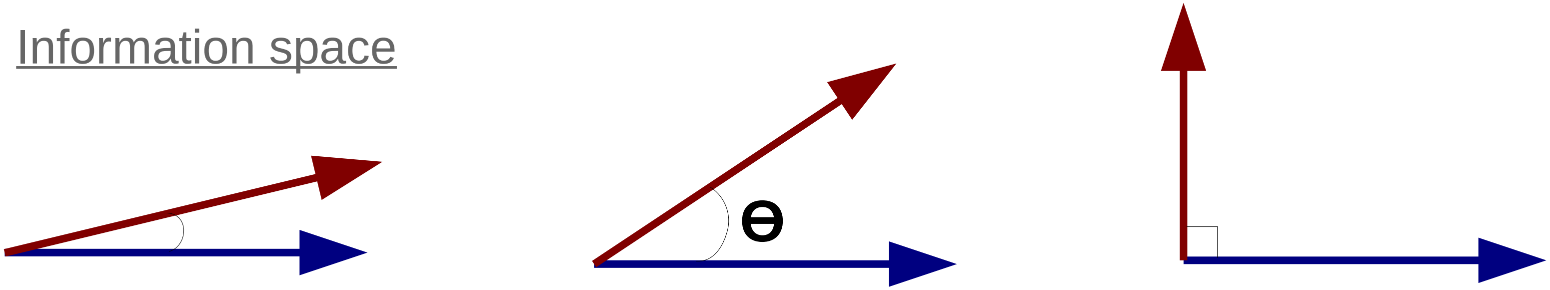
- $I_{Total} = I_{Time\ domain} = I_{\ Frequency\ domain}$
  - $I_{Frequency\ domain} = I_{Re} \cup I_{Im}$
  - $I_{Frequency\ domain} = I_{Magnitude} \cup I_{Phase}$
  - $I_{Frequency\ domain} = I_{All\text{-}pass} \cup I_{Min\text{-}phase}$

INFORMATION

Each pair *spans* the information space

Information Distribution in Frequency Domain

Information space



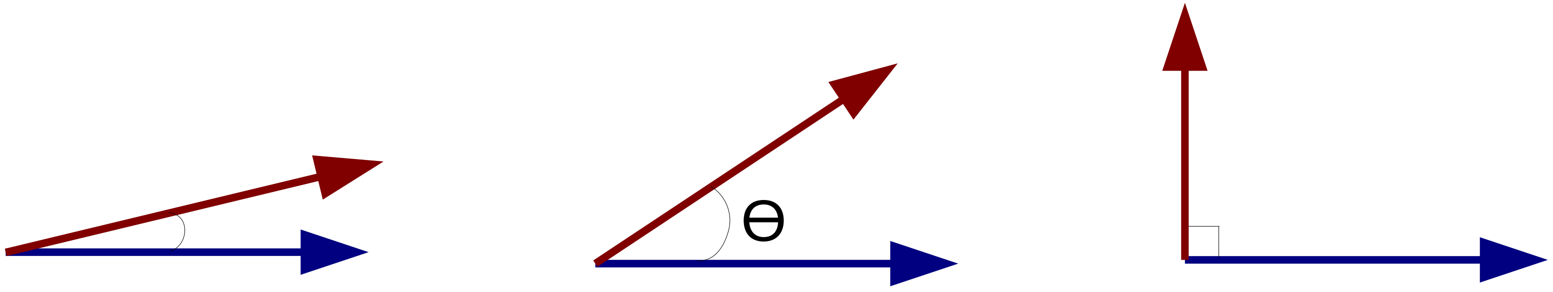
Real / Imaginary Parts

Phase / Magnitude Spectra

Minimum-phase / All-pass Components

θ reflects information *correlation*
(orthogonality ↔ *sharing* no info)

Information Distribution in Frequency Domain



Real / Imaginary
Parts

Phase / Magnitude
Spectra

Minimum-phase /
All-pass Components

$$\boxed{I_{Real} \cap I_{Imaginary} > I_{Magnitude} \cap I_{Phase}}$$

$$\boxed{I_{All\text{-}pass} \cap I_{min\text{-}phase} = \{\}}$$

Phase and Magnitude Relationship

$$X_{Re} \rightleftharpoons X_{Im}$$

$$Magnitude \rightleftharpoons Phase$$

Phase and Magnitude Relationship

$$X_{Re} \boxed{\rightleftharpoons} X_{Im}$$

The milder the requirement
the higher the correlation

$$Magnitude \boxed{\rightleftharpoons} Phase$$

Phase and Magnitude Relationship

$$X_{Re} \boxed{\rightleftharpoons} X_{Im}$$

Hilbert Transform

$$Magnitude \boxed{\rightleftharpoons} Phase$$

Phase and Magnitude Relationship

$$X_{Re} \boxed{\rightleftharpoons} X_{Im}$$

Hilbert Transform ← **Causality**

$$Magnitude \boxed{\rightleftharpoons} Phase$$

Phase and Magnitude Relationship

$$X_{Re} \rightleftharpoons X_{Im}$$

Time domain

Hilbert Transform

Causality

$$Magnitude \rightleftharpoons Phase$$

Complex Cepstrum domain

Phase and Magnitude Relationship

$$X_{Re} \boxed{\rightleftharpoons} X_{Im}$$

Hilbert Transform

$$Magnitude \boxed{\rightleftharpoons} Phase$$

EASY

Time domain

Causality

Complex Cepstrum domain

Difficult

Phase and Magnitude Relationship

$$X_{Re} \boxed{\rightleftharpoons} X_{Im}$$

Hilbert Transform

$$Magnitude \boxed{\rightleftharpoons} Phase$$

EASY

Time domain

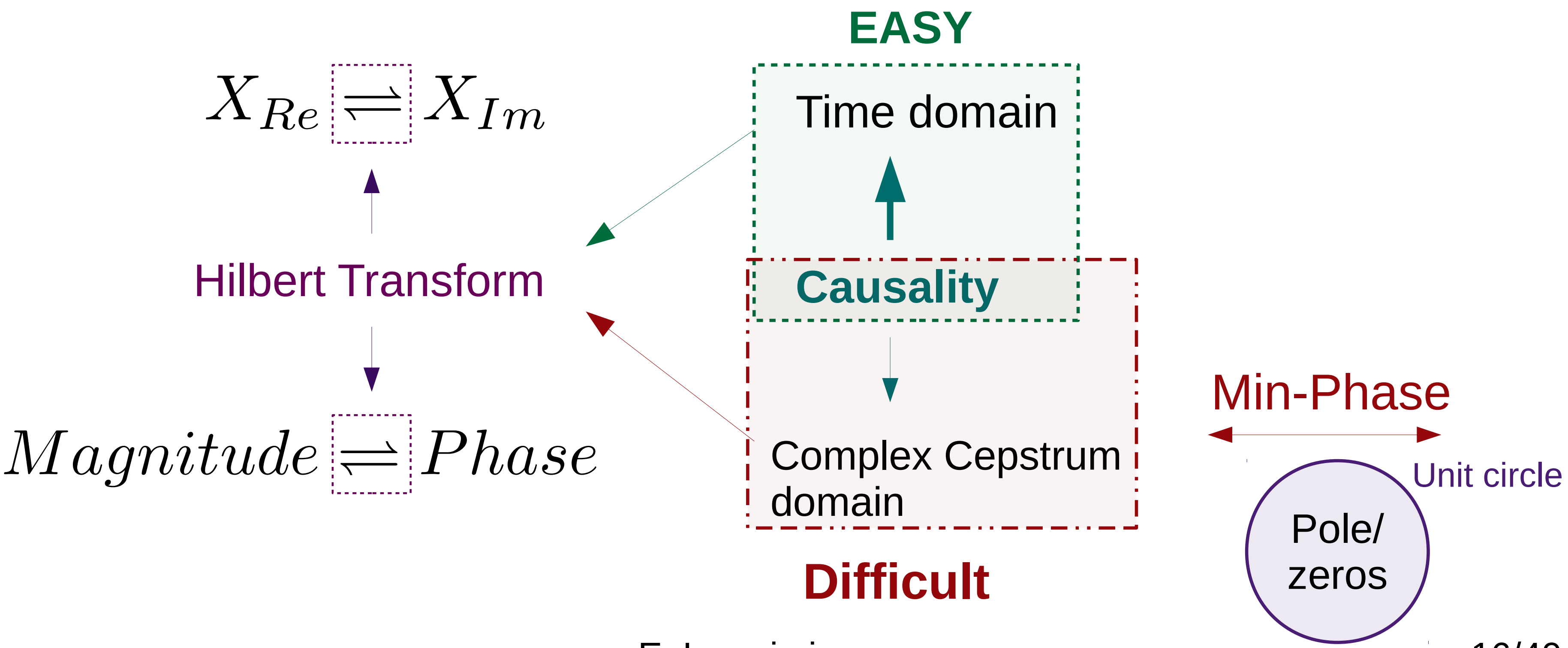
Causality

Complex Cepstrum domain

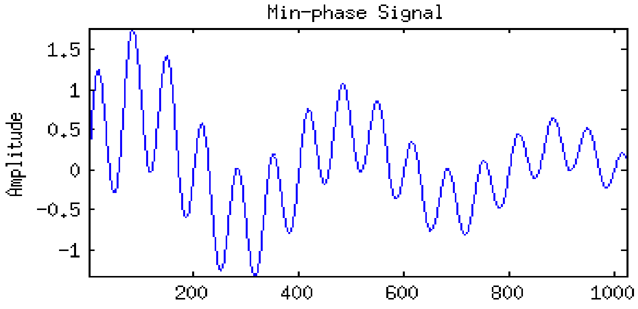
Difficult

Min-Phase

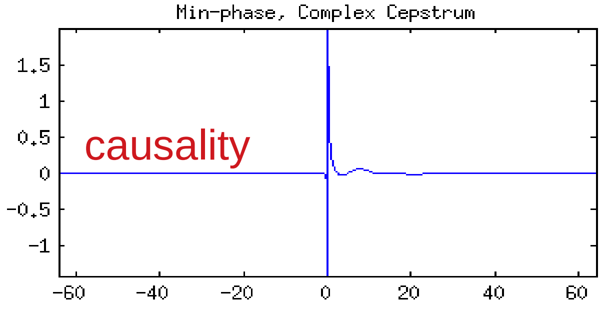
Phase and Magnitude Relationship



Minimum-Phase Signals

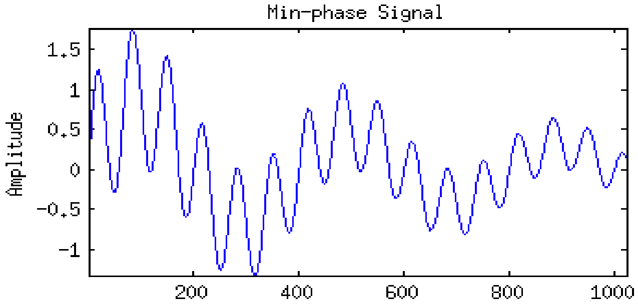


Time in Samples

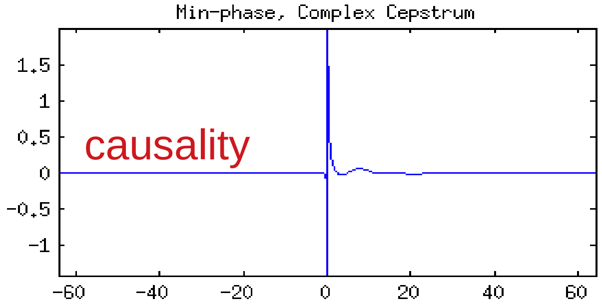


Quefrency in Samples

Minimum-Phase Signals



Time in Samples



Quefrency in Samples

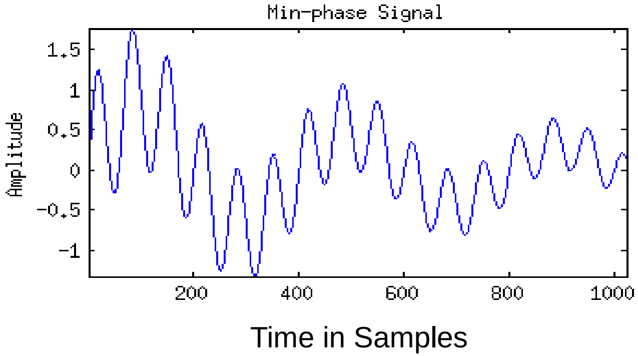
Hilbert Transform

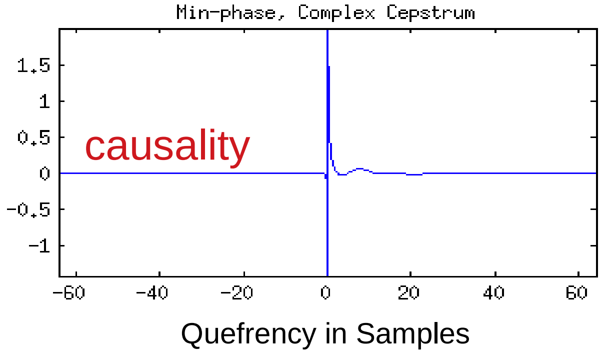
$$|X(\omega)| \xrightarrow{\text{HiL.Tran}} \phi_X(\omega)$$

$$\phi_X(\omega) \xrightarrow{\text{HiL.Tran}} |X(\omega)|/c$$

$$c = \exp(mean(\log|X(\omega)|))$$

Minimum-Phase Signals





Hilbert Transform $\longrightarrow$

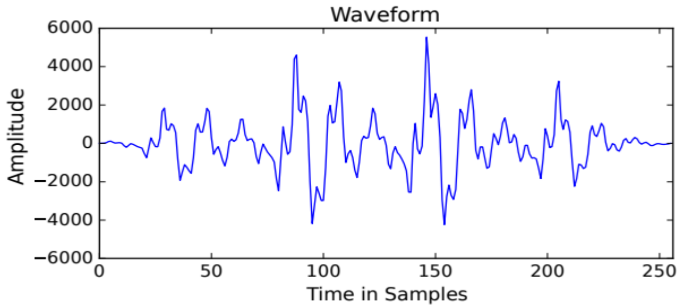
$$|X(\omega)| \xrightarrow{\text{HiL.Tran}} \phi_X(\omega)$$

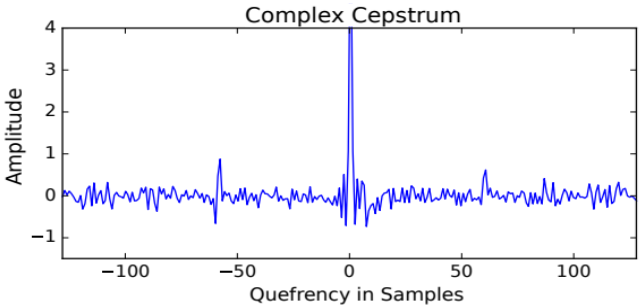
$$\phi_X(\omega) \xrightarrow{\text{HiL.Tran}} |X(\omega)|/c$$

$$c = \exp(mean(\log|X(\omega)|))$$

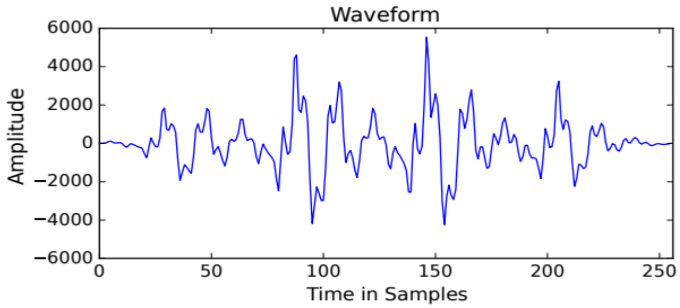
Both spectra carry (almost) the same amount of information!

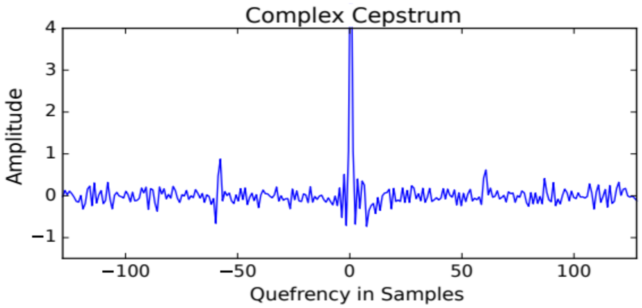
Speech is a Mixed-Phase Signal!





Speech is a Mixed-Phase Signal!





No one-to-one relationship between Phase & Magnitude!

Phase and Magnitude Relationship

- In general, there is no relationship

$$Phase \rightleftharpoons Magnitude$$

ill-posed

Phase and Magnitude Relationship

- In general, there is no relationship

$$Phase \rightleftharpoons Magnitude$$

ill-posed

#knowns < #unknowns
(under-determined)

Phase and Magnitude Relationship

- In general, there is no relationship

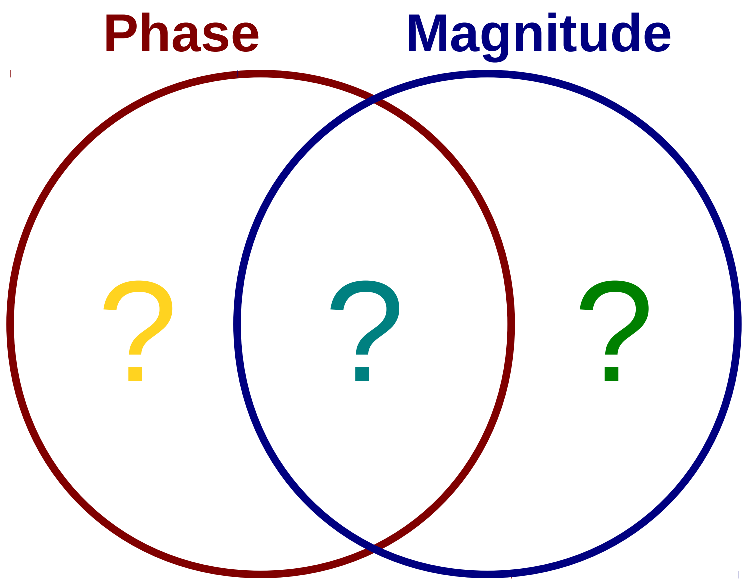
$$Phase \rightleftharpoons Magnitude$$

ill-posed

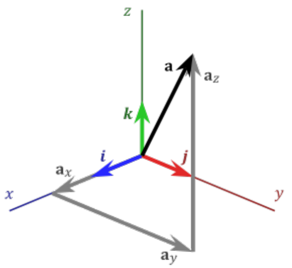
#knowns $\approx$ #unknowns + Constraints

Information Distribution between Phase and Magnitude

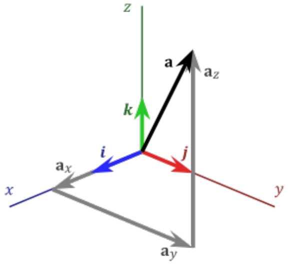
- Unique to phase ?
- Shared ?
- Unique to magnitude ?

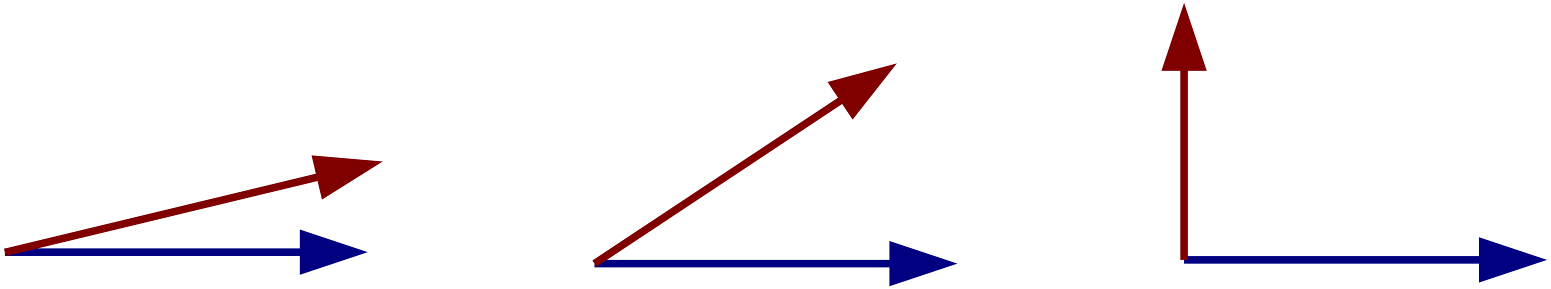


Signal Decomposition



Signal Decomposition



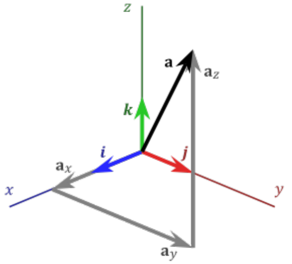


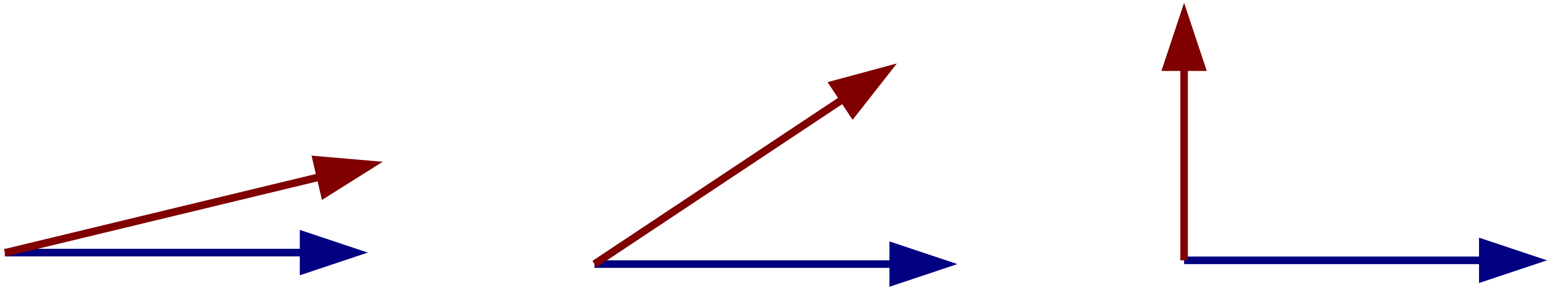
Real / Imaginary
Parts

Phase / Magnitude
Spectra

Minimum-phase /
All-pass Components

Signal Decomposition





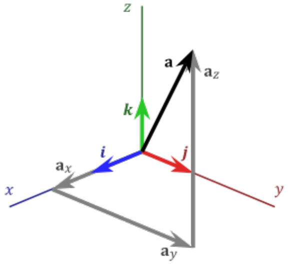
Real / Imaginary Parts

Phase / Magnitude Spectra

Minimum-phase / All-pass Components

Which one is better for signal information decomposition?

Signal Information Distribution



$$X(\omega) = |X(\omega)|\, e^{j\phi_X(\omega)}$$

$$X(\omega) = \; X_{MinPh}(\omega)\; X_{AllP}(\omega)$$

$$X(\omega) = |X_{MinPh}(\omega)|\; e^{j\phi_{MinPh}(\omega)}\;\; e^{j\phi_{AllP}(\omega)}$$

Min-Phase/All-Pass Decomposition

Signal Information Distribution

$$X(\omega) = |X(\omega)|\, e^{j\phi_X(\omega)}$$

$$X(\omega) = \ X_{MinPh}(\omega)\ X_{AllP}(\omega)$$

$$X(\omega) = |X_{MinPh}(\omega)|\ e^{j\phi_{MinPh}(\omega)}\ \ e^{j\phi_{AllP}(\omega)}$$

Min-Phase/All-Pass Decomposition

$$|X(\omega)| = \ |X_{MinPh}(\omega)|$$

$$arg[X(\omega)] = \ arg[X_{MinPh}(\omega)]\ +\ arg[X_{AllP}(\omega)]$$

Signal Information Distribution

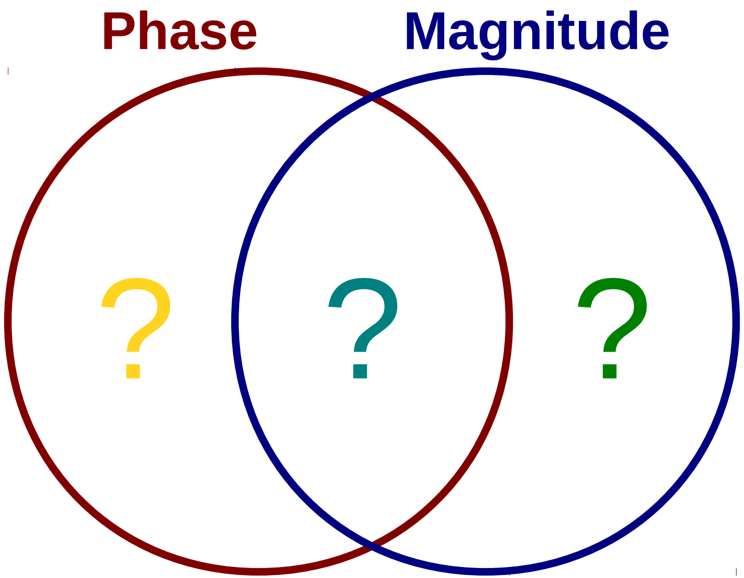
$$|X(\omega)| = |X_{MinPh}(\omega)|$$
$$arg[X(\omega)] = arg[X_{MinPh}(\omega)] + arg[X_{AllP}(\omega)]$$

log

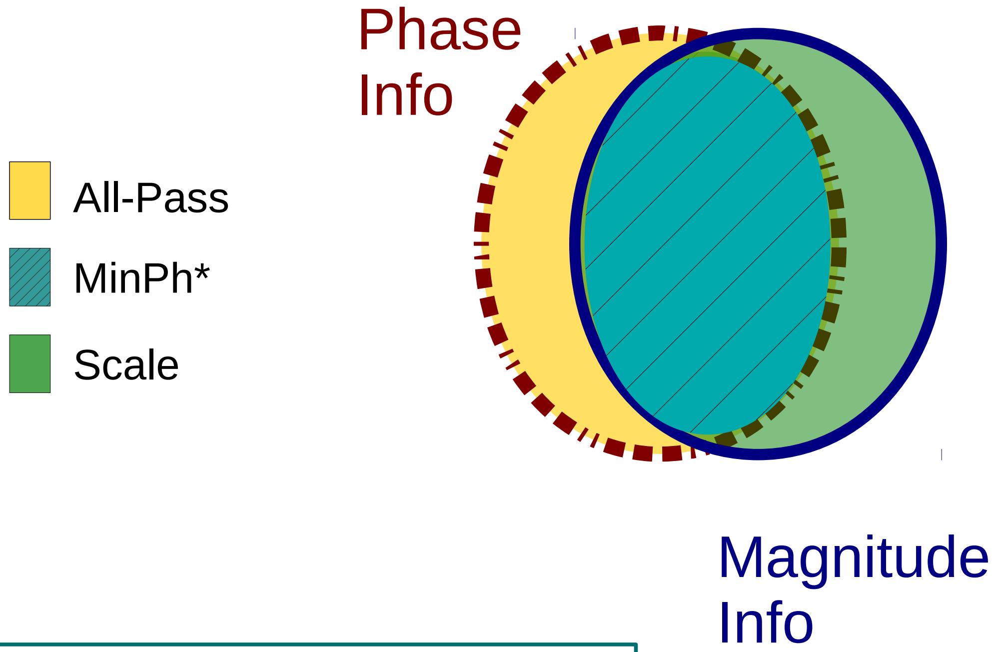
$$|X_{MinPh}(\omega)| \quad \underleftrightarrow{\text{HiL.Tran}} \quad \arg[X_{MinPh}(\omega)]$$

Hilbert Transform

Signal Information Distribution

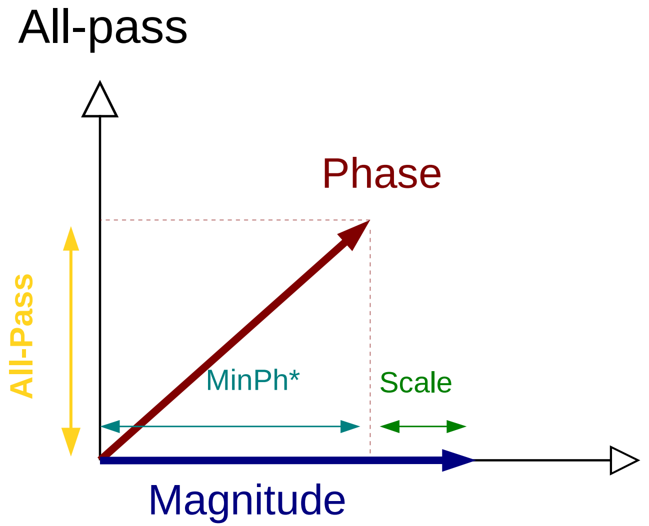


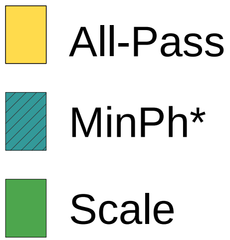
Signal Information Distribution

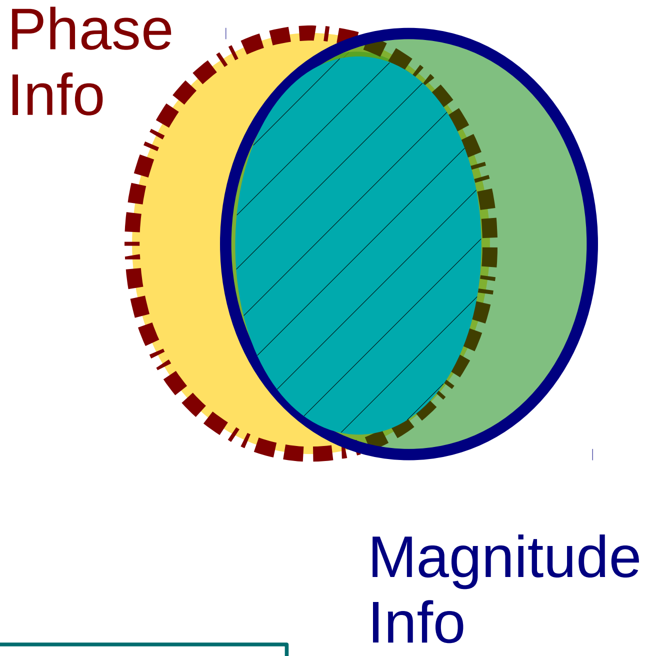


$$\boxed{|X_{MinPh}(\omega)| \quad \overset{\text{HiL.Tran}}{\longleftrightarrow} \quad \arg[X_{MinPh}(\omega)]}$$

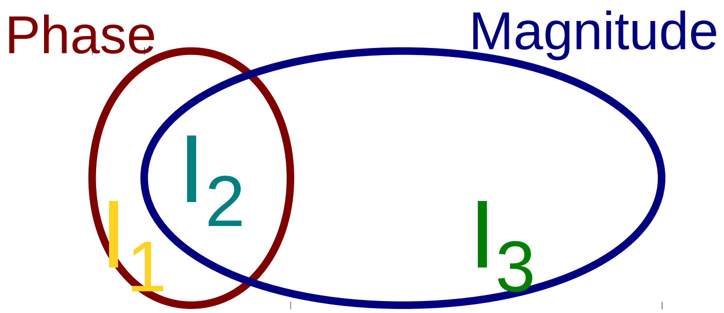
Signal Information Distribution

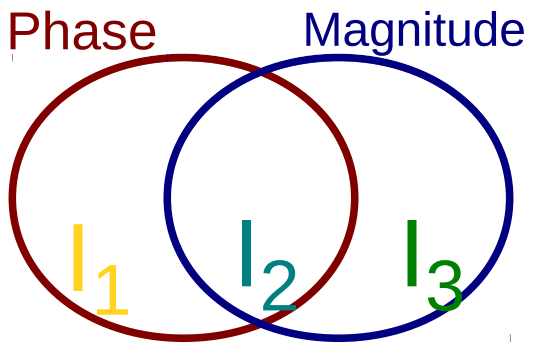




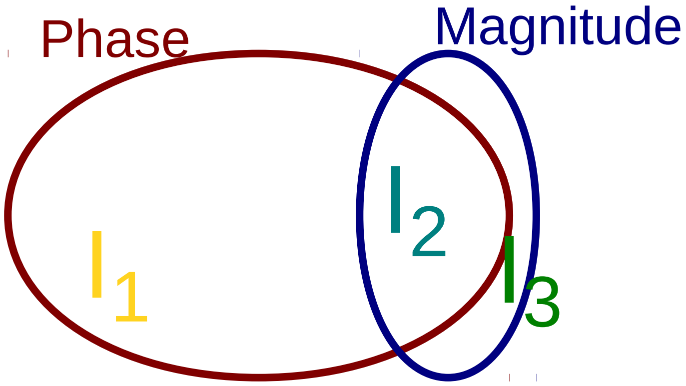


$$\boxed{|X_{MinPh}(\omega)| \;\xleftrightarrow{\text{HiL.Tran}}\; \arg[X_{MinPh}(\omega)]}$$





Relative importance?



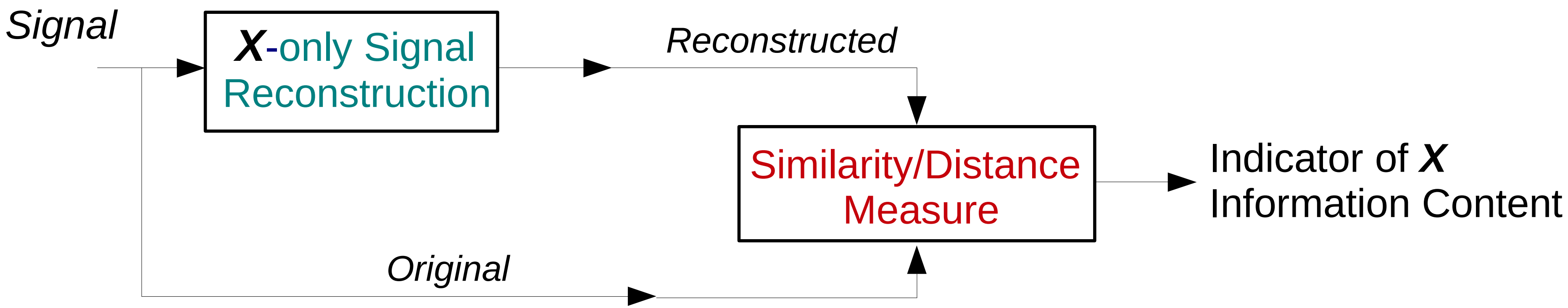
E. Loweimi

Outlines

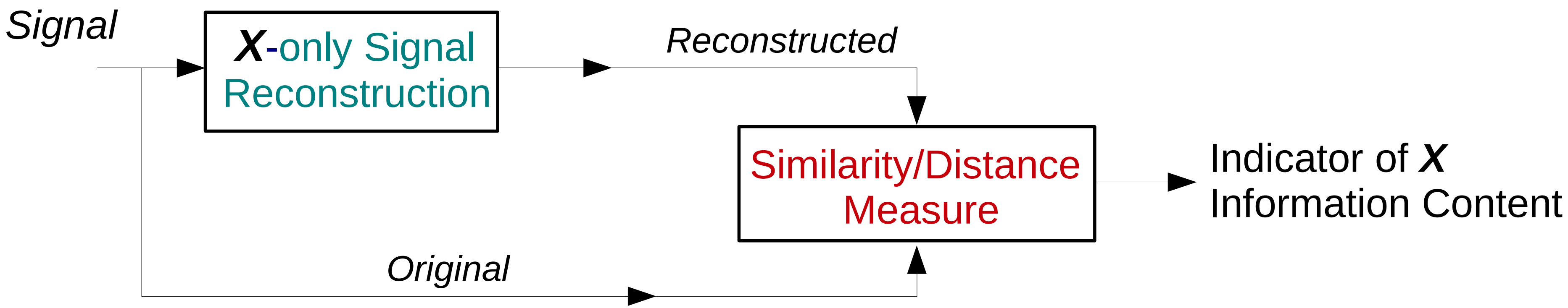
- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction — Part I

- Source-filter Separation in the Phase Domain — Part II
- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F_0 Extraction

Signal Reconstruction

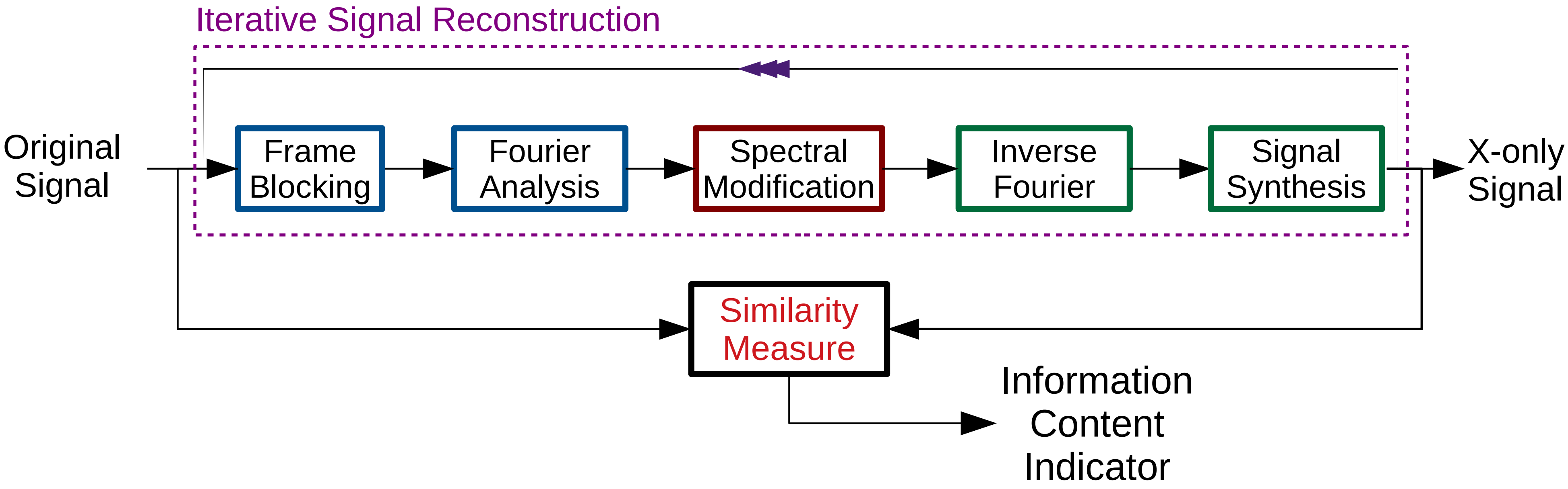


Signal Reconstruction

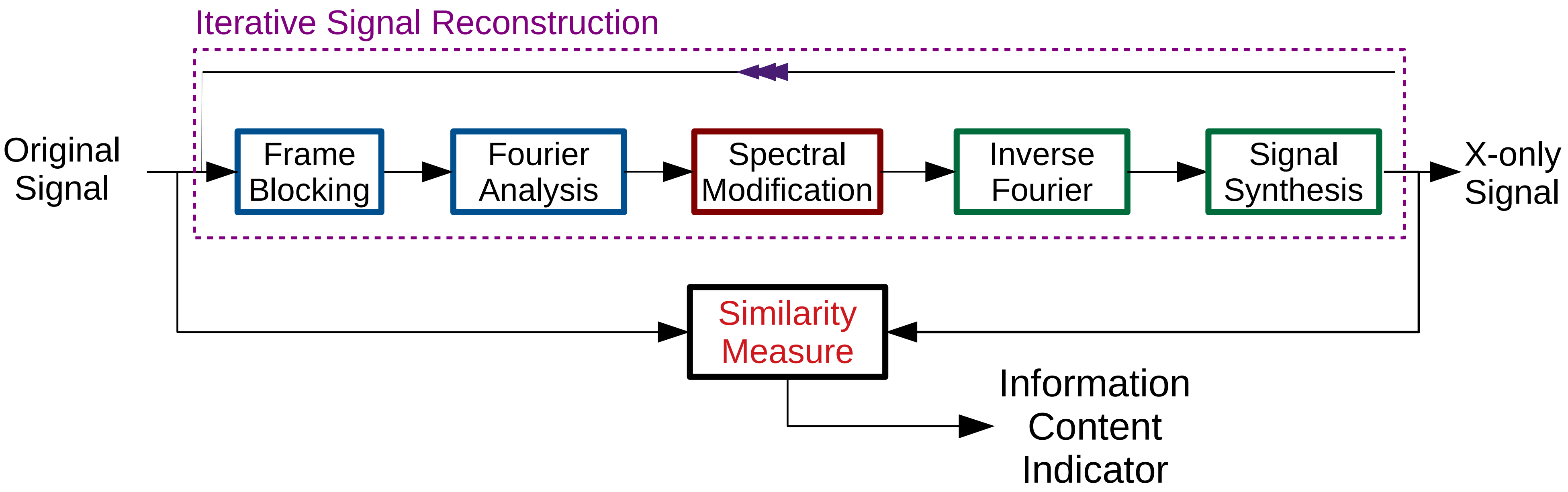


– *X:* Magnitude, Phase, …
– *X*-only signal reconstruction is *ill-posed*
– Solution → *constraints* + *iteration*

Signal Reconstruction – AMS

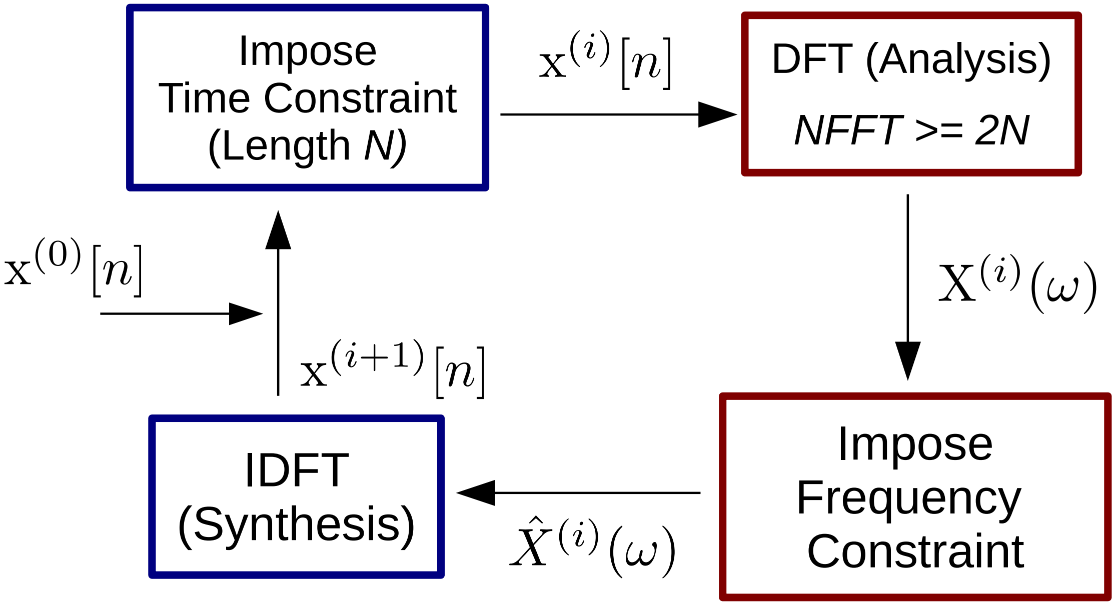


Signal Reconstruction – AMS

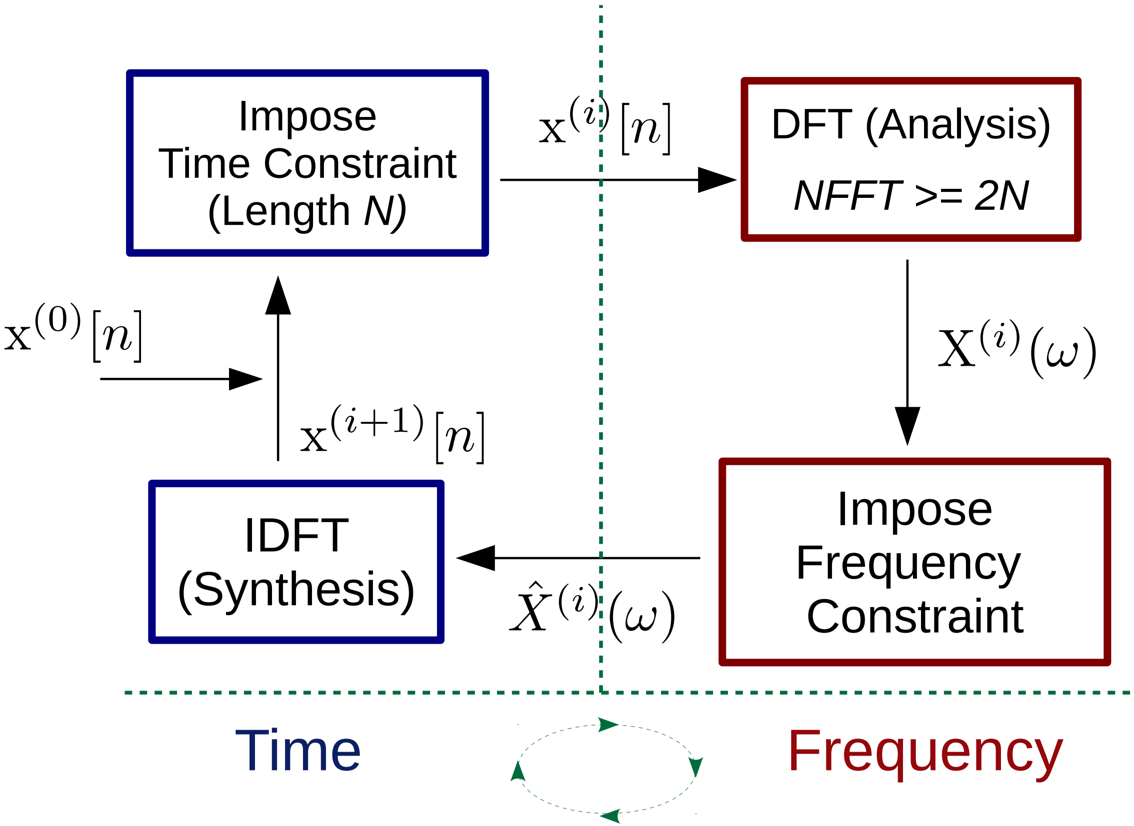


Analysis-*Modification*-*Synthesis* (AMS) framework

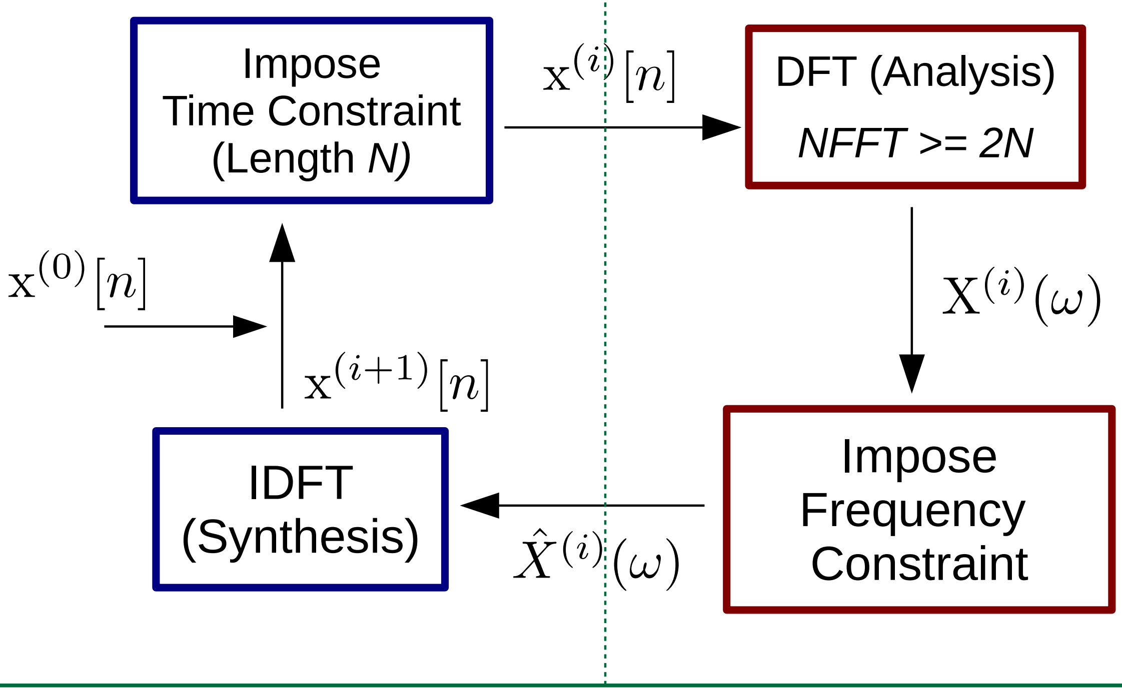
Iterative Signal Reconstruction



Iterative Signal Reconstruction



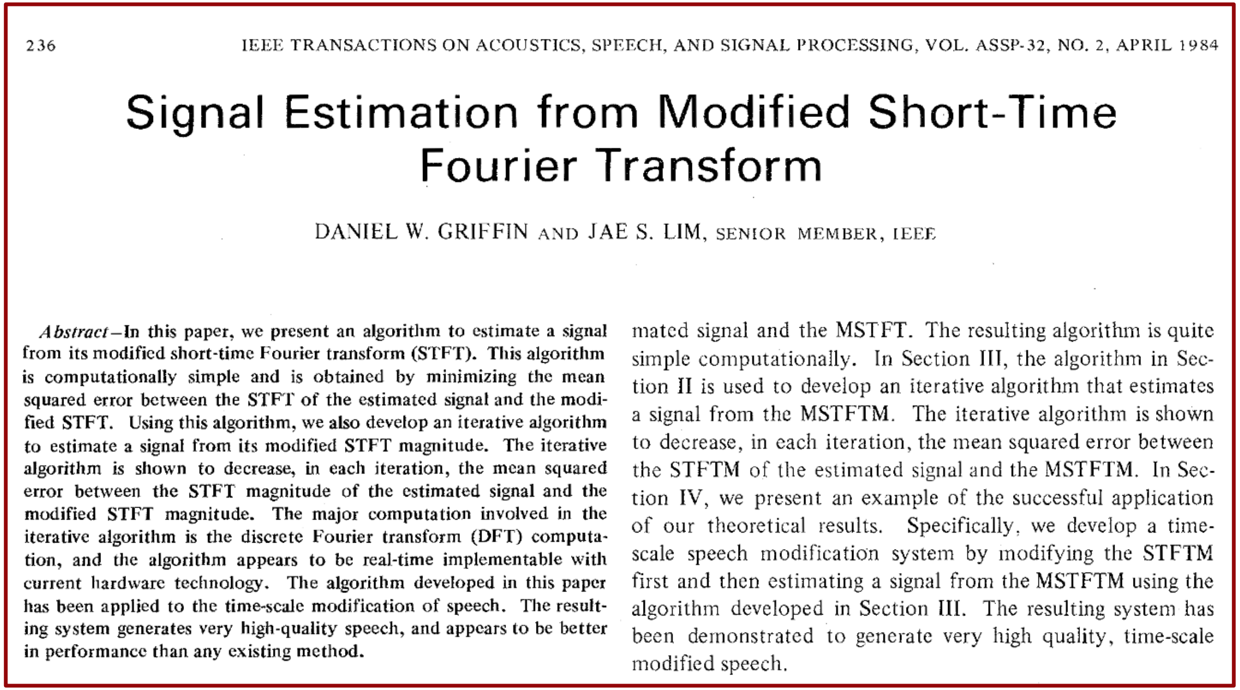
Iterative Signal Reconstruction



– Switch back and forth between time and frequency domains

– Impose domain-specific constraints

Griffin and Lim Algorithm



236 IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-32, NO. 2, APRIL 1984

Signal Estimation from Modified Short-Time Fourier Transform

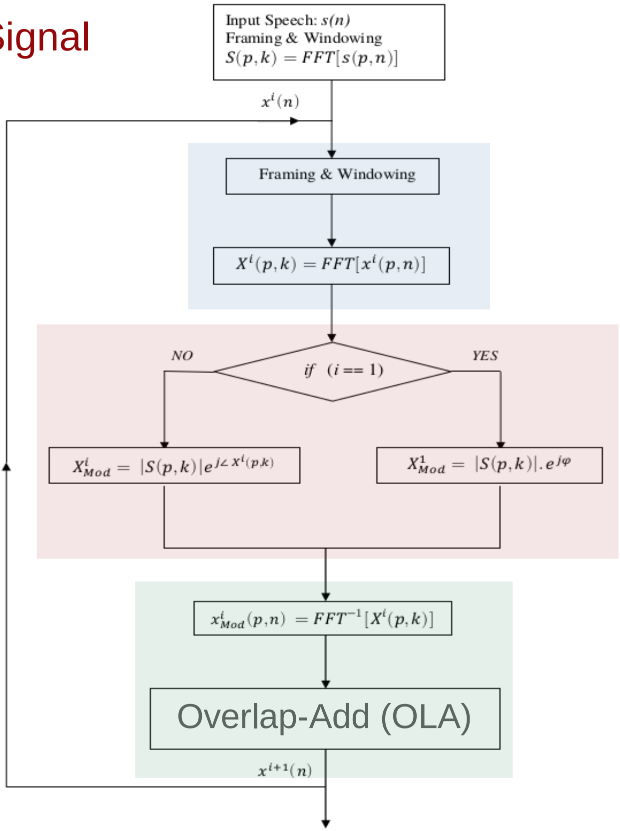
DANIEL W. GRIFFIN AND JAE S. LIM, SENIOR MEMBER, IEEE

Abstract—In this paper, we present an algorithm to estimate a signal from its modified short-time Fourier transform (STFT). This algorithm is computationally simple and is obtained by minimizing the mean squared error between the STFT of the estimated signal and the modified STFT. Using this algorithm, we also develop an iterative algorithm to estimate a signal from its modified STFT magnitude. The iterative algorithm is shown to decrease, in each iteration, the mean squared error between the STFT magnitude of the estimated signal and the modified STFT magnitude. The major computation involved in the iterative algorithm is the discrete Fourier transform (DFT) computation, and the algorithm appears to be real-time implementable with current hardware technology. The algorithm developed in this paper has been applied to the time-scale modification of speech. The resulting system generates very high-quality speech, and appears to be better in performance than any existing method.

mated signal and the MSTFT. The resulting algorithm is quite simple computationally. In Section III, the algorithm in Section II is used to develop an iterative algorithm that estimates a signal from the MSTFTM. The iterative algorithm is shown to decrease, in each iteration, the mean squared error between the STFTM of the estimated signal and the MSTFTM. In Section IV, we present an example of the successful application of our theoretical results. Specifically, we develop a time-scale speech modification system by modifying the STFTM first and then estimating a signal from the MSTFTM using the algorithm developed in Section III. The resulting system has been demonstrated to generate very high quality, time-scale modified speech.

Griffin and Lim Algorithm (AMS)

Magnitude-only Signal Reconstruction

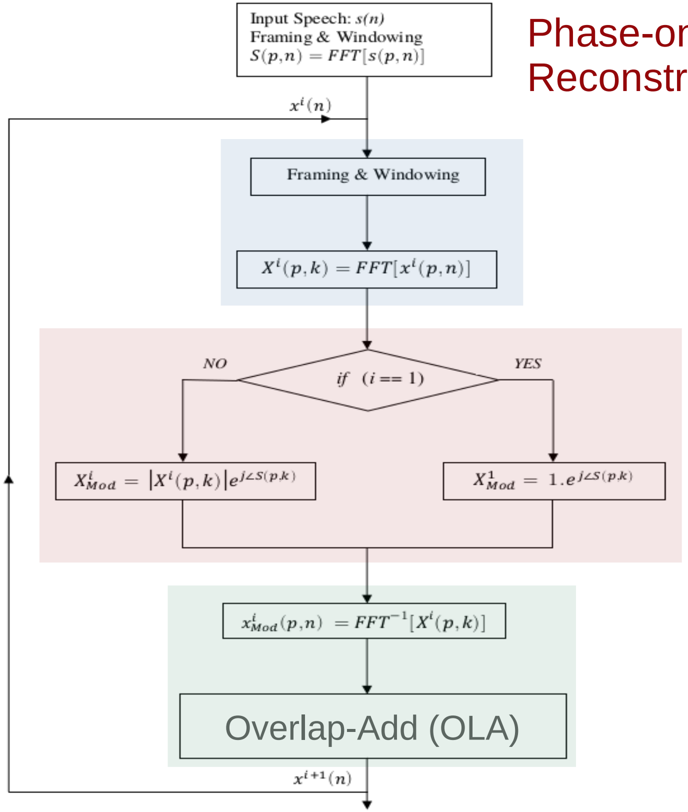


A

M

S

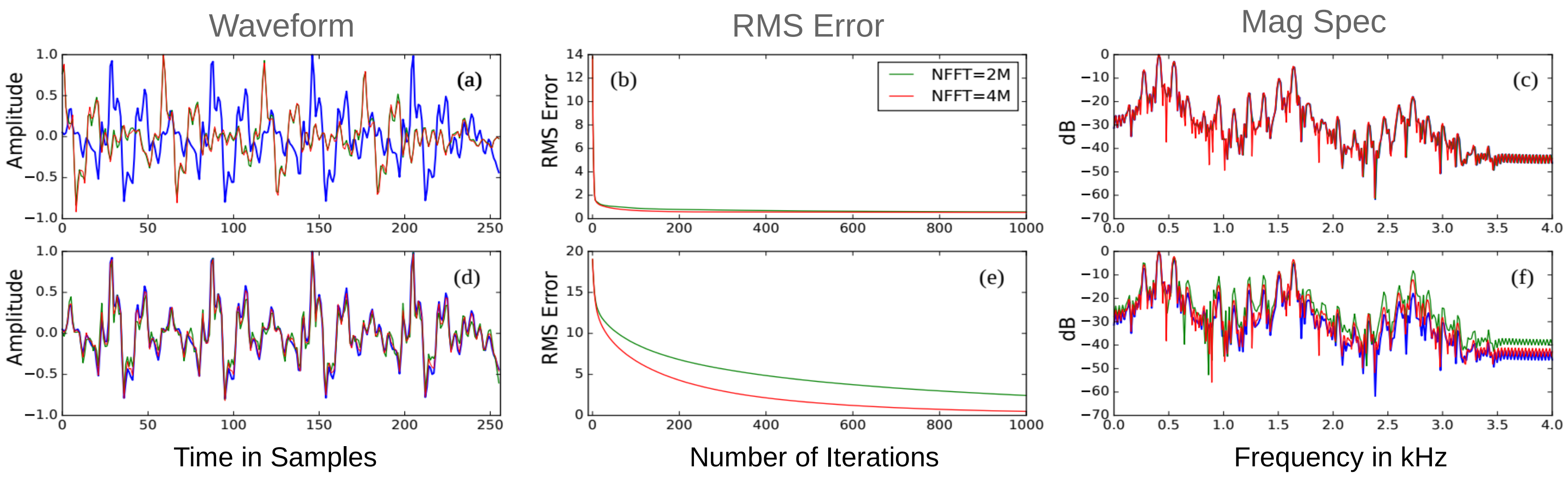
Phase-only Signal Reconstruction



X-only Signal Reconstruction

X: Magnitude

X: Phase

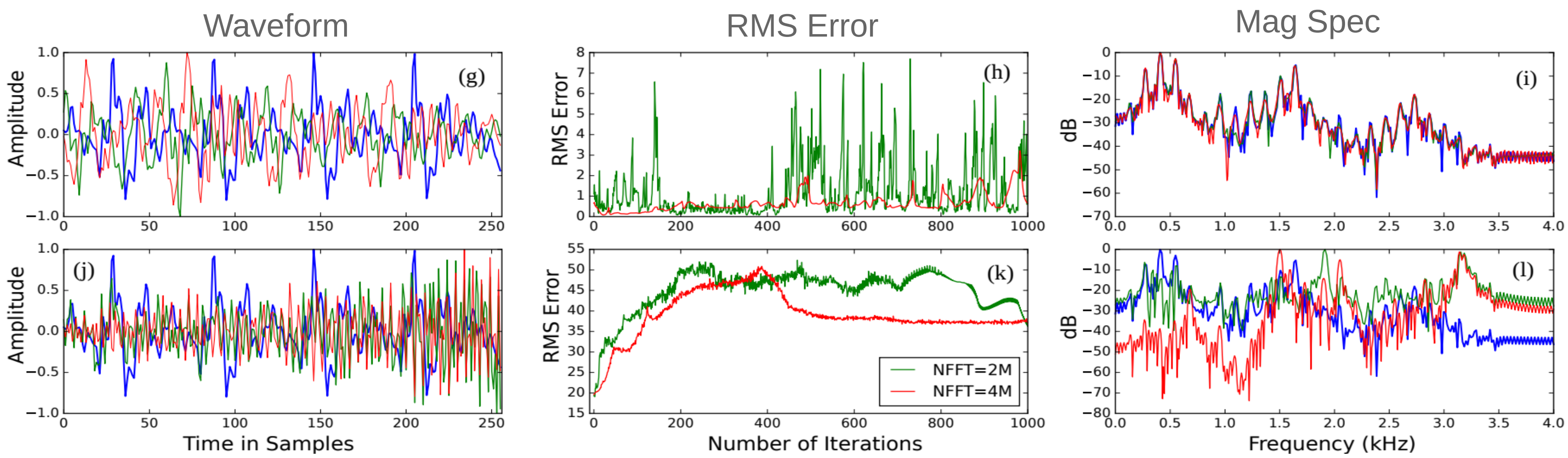


Blue ↔ Original

X-only Signal Reconstruction

X: Min-phase

X: All-pass

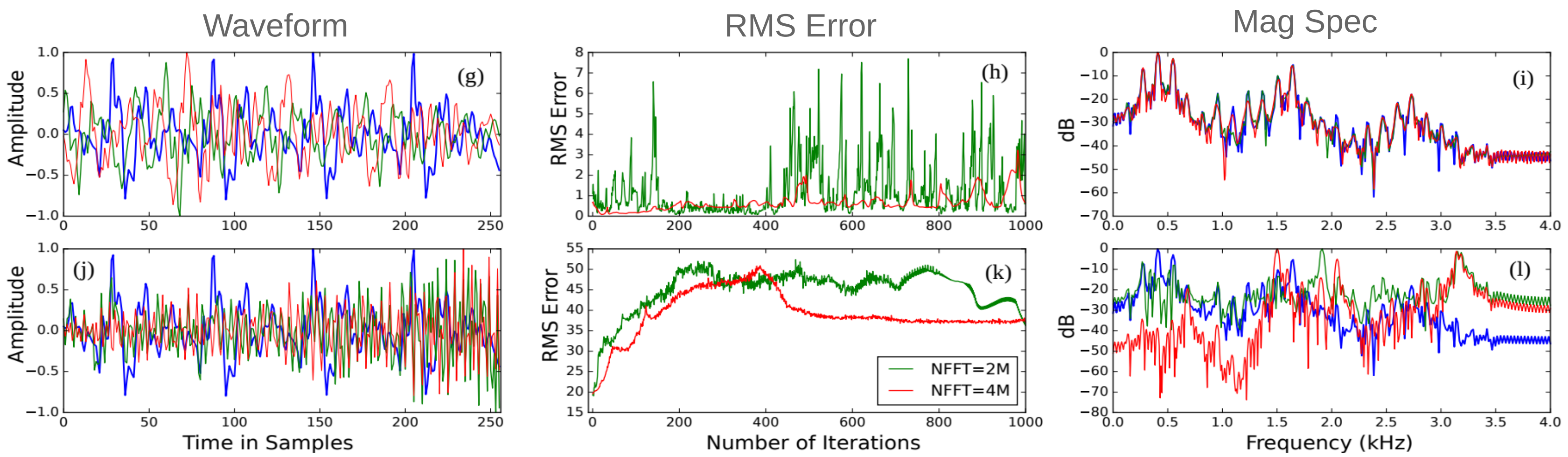


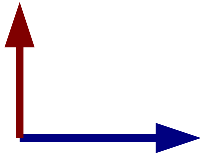
Blue ↔ Original

X-only Signal Reconstruction

X: Min-phase

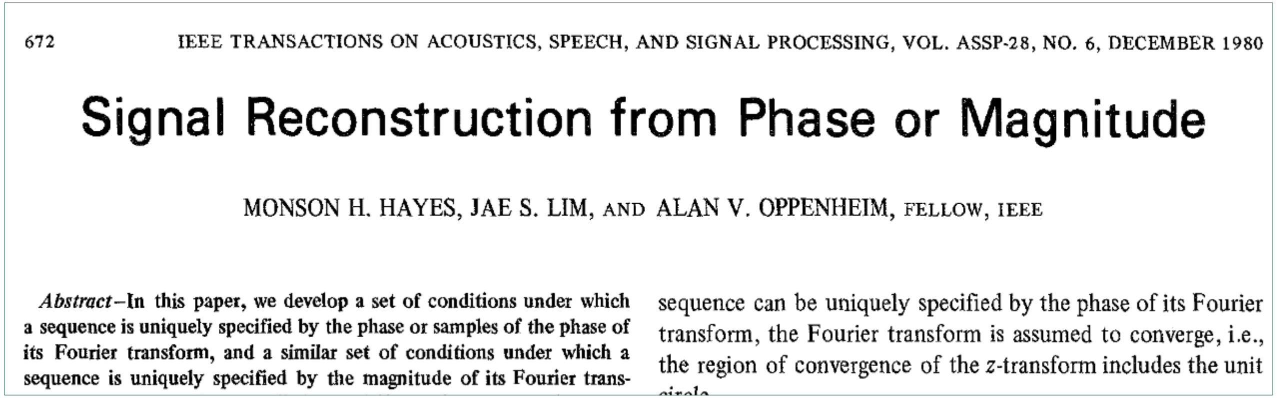
X: All-pass





Blue ↔ Original

Phase-only Signal Reconstruction

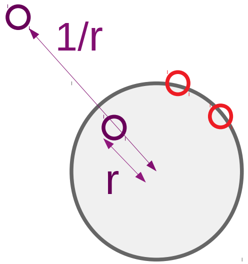

672 IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-28, NO. 6, DECEMBER 1980

Signal Reconstruction from Phase or Magnitude

MONSON H. HAYES, JAE S. LIM, AND ALAN V. OPPENHEIM, FELLOW, IEEE

Abstract—In this paper, we develop a set of conditions under which a sequence is uniquely specified by the phase or samples of the phase of its Fourier transform, and a similar set of conditions under which a sequence is uniquely specified by the magnitude of its Fourier trans-

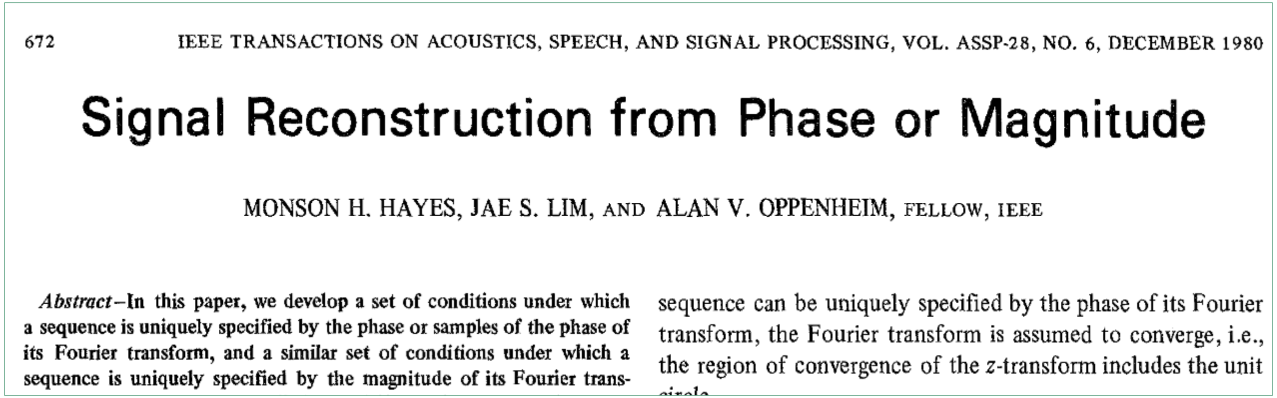
sequence can be uniquely specified by the phase of its Fourier transform, the Fourier transform is assumed to converge, i.e., the region of convergence of the z-transform includes the unit circle.



Theorem 3

"A sequence which is known to be zero outside the interval $0 \le n \le N-1$ is uniquely specified to within a scale factor by $N-1$ distinct samples of its phase spectrum in the interval $0 \le \omega \le \pi$ if it has a z-transform with no zeros on the unit circle or in conjugate reciprocal pairs"

Phase-only Signal Reconstruction

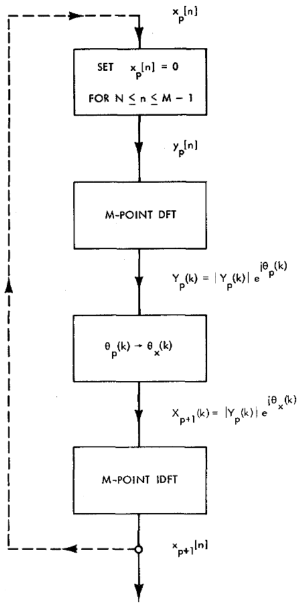

672 IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-28, NO. 6, DECEMBER 1980

Signal Reconstruction from Phase or Magnitude

MONSON H. HAYES, JAE S. LIM, AND ALAN V. OPPENHEIM, FELLOW, IEEE

Abstract—In this paper, we develop a set of conditions under which a sequence is uniquely specified by the phase or samples of the phase of its Fourier transform, and a similar set of conditions under which a sequence is uniquely specified by the magnitude of its Fourier trans-

sequence can be uniquely specified by the phase of its Fourier transform, the Fourier transform is assumed to converge, i.e., the region of convergence of the z-transform includes the unit

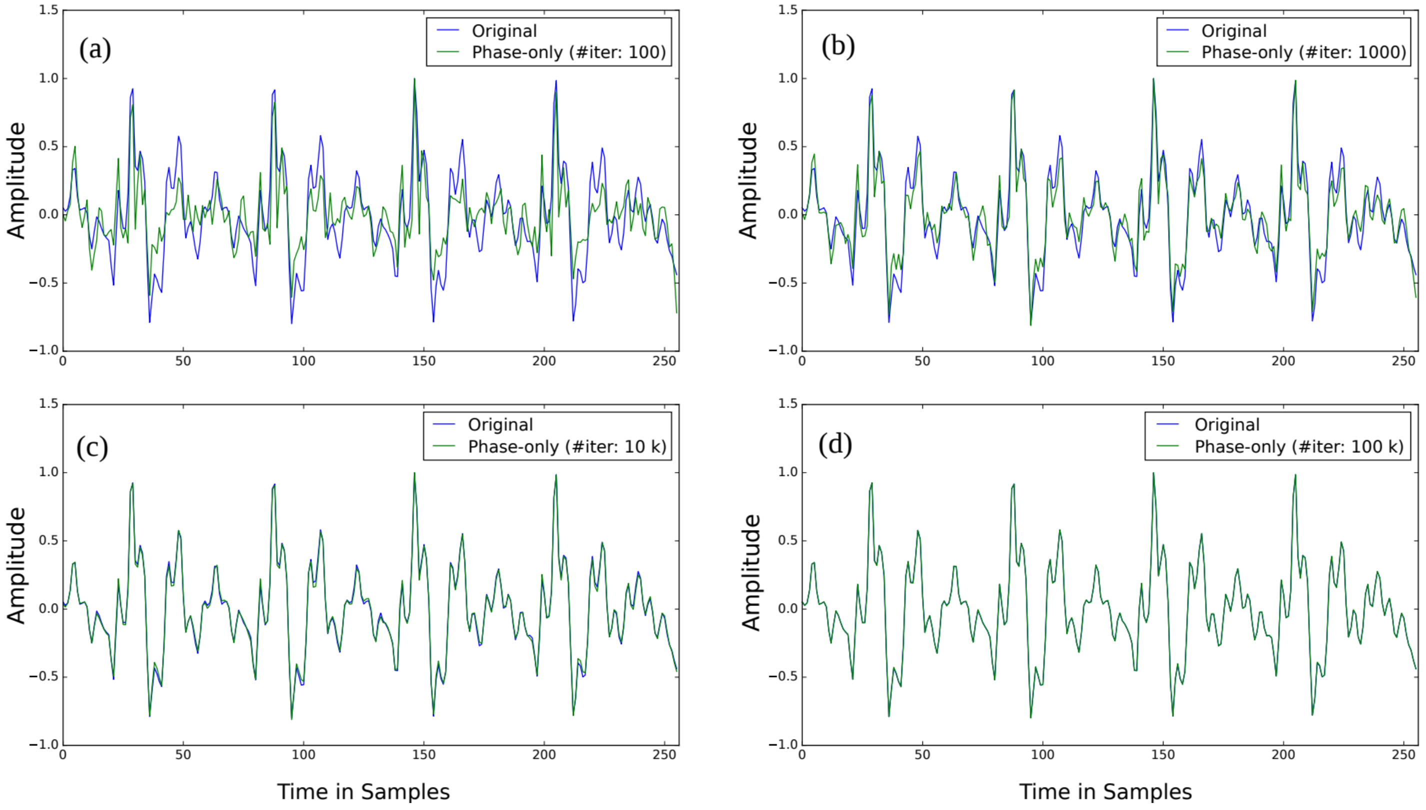


<u>Theorem 3</u>

(1) MOST signals can be reconstructed from the phase spectrum, upto a scale error.

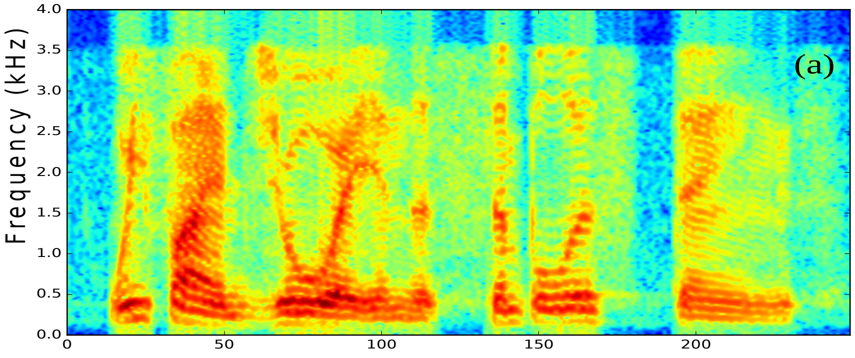
(2) #Iterations is unknown

Phase-only Signal Reconstruction

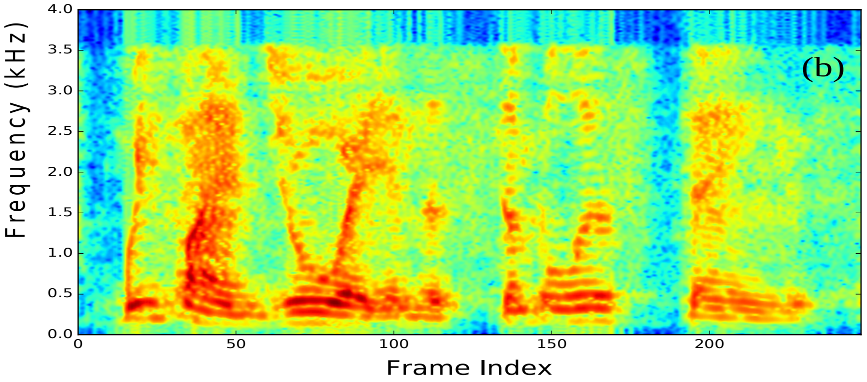


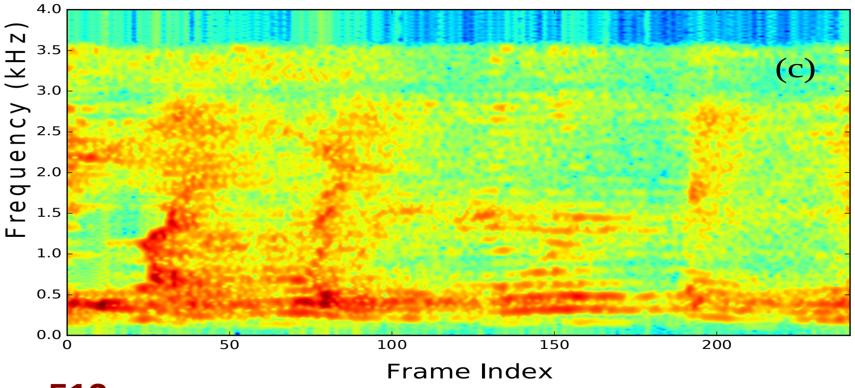
Minimum Phase-only Reconstructed Speech

32 ms



128 ms

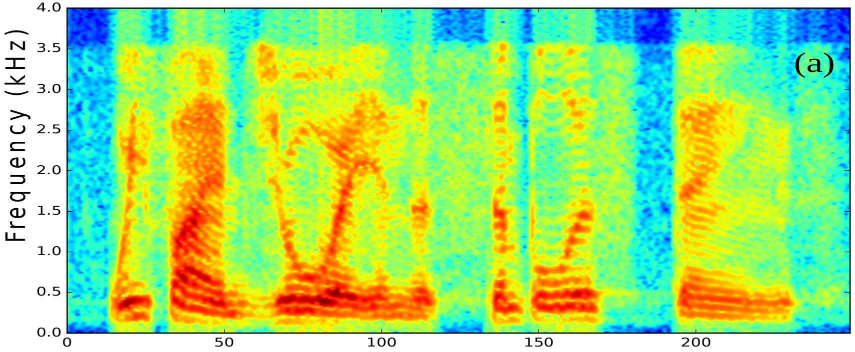




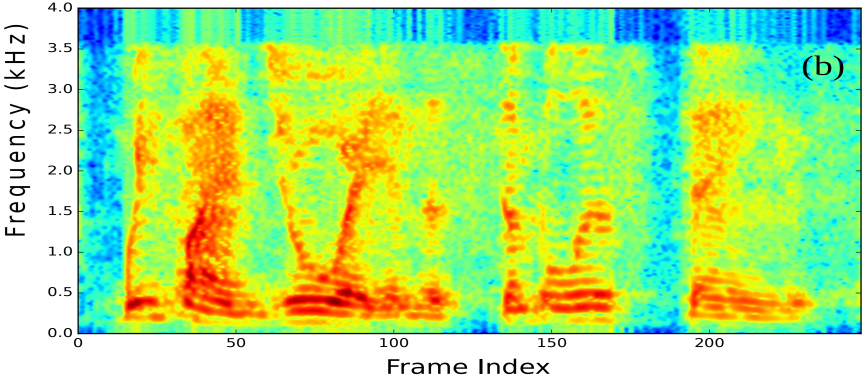
512 ms

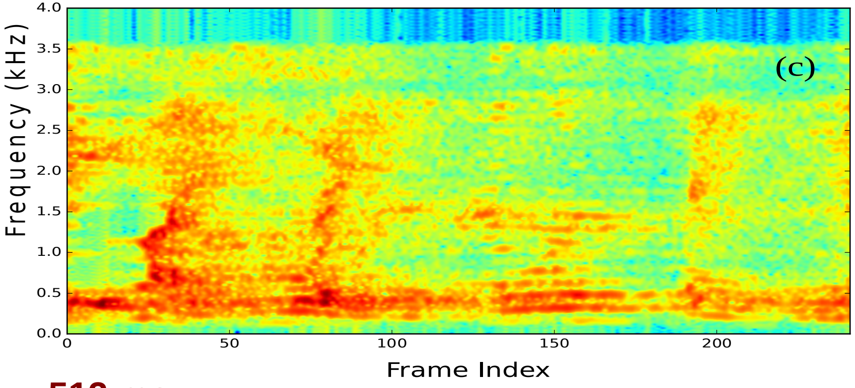
Minimum Phase-only Reconstructed Speech

32 ms



128 ms



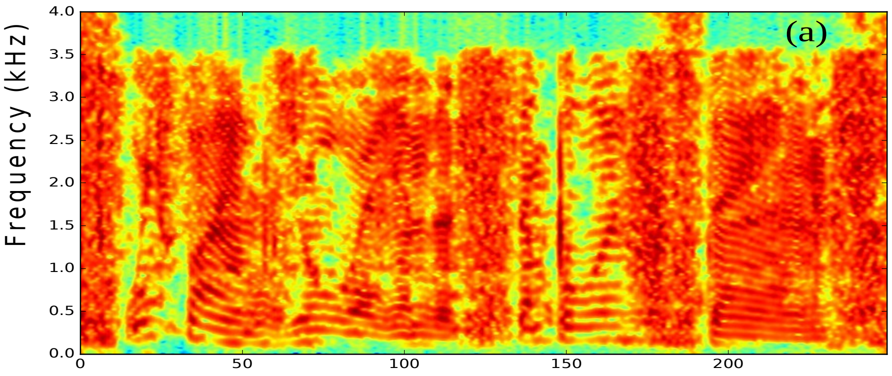


512 ms

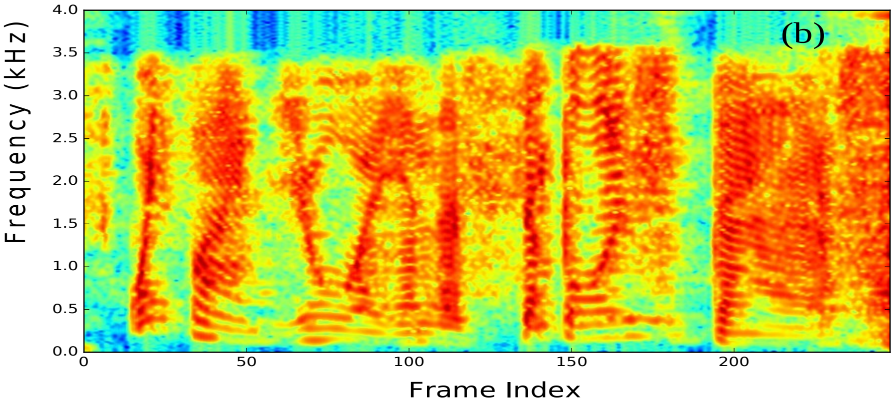
Min-ph component is informative in short-term

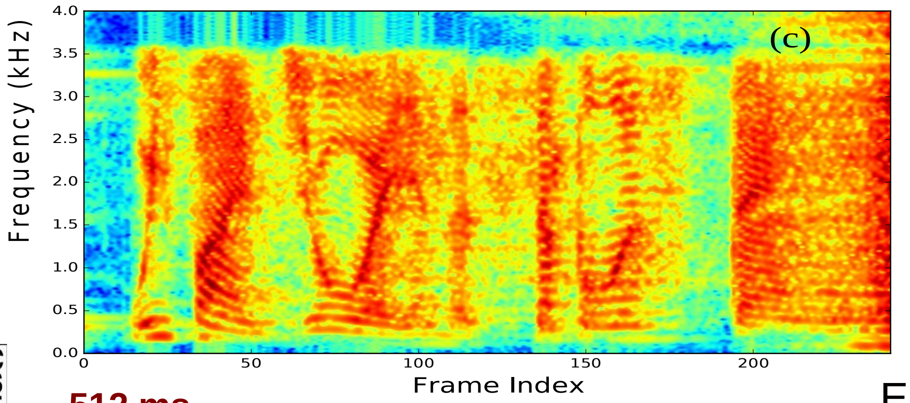
All-Pass-only Reconstructed Speech

32 ms



128 ms

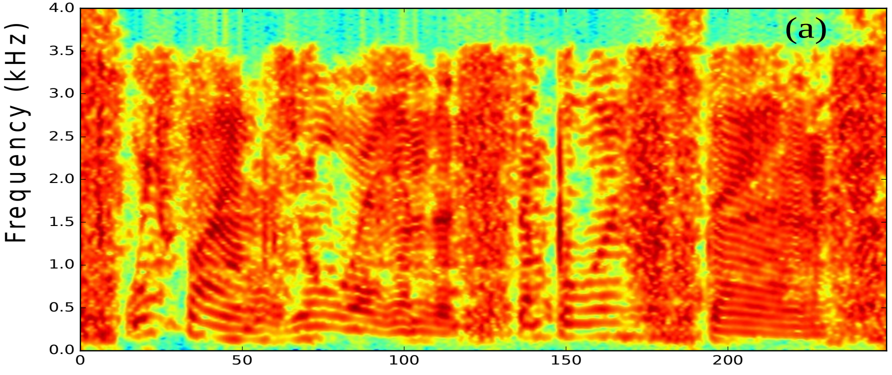




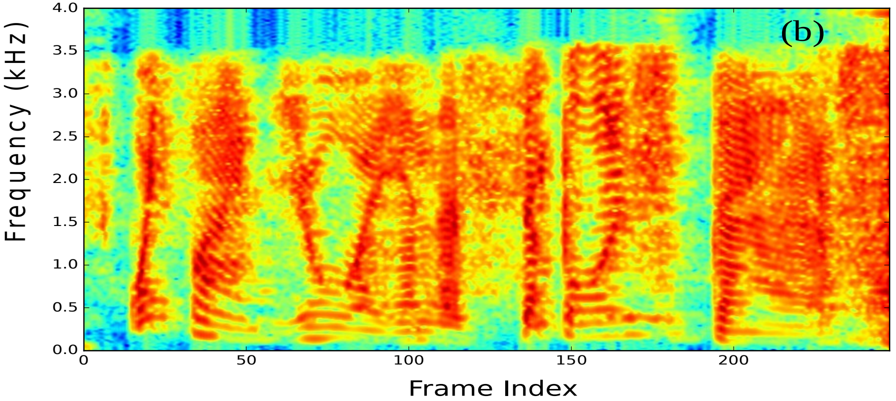
512 ms

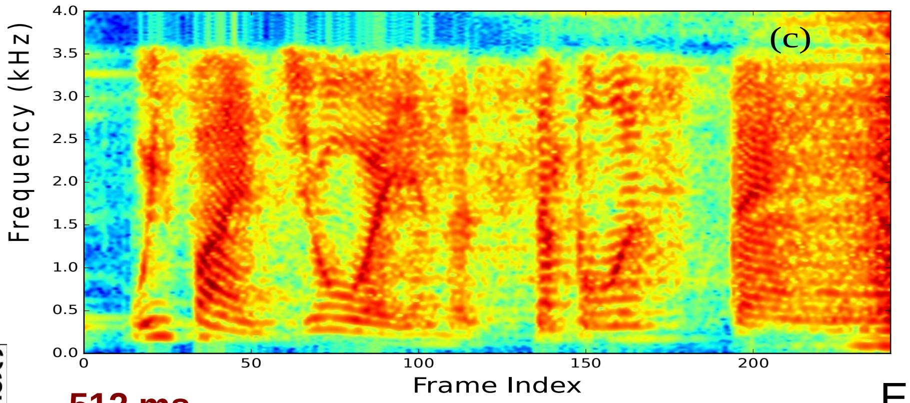
All-Pass-only Reconstructed Speech

32 ms



128 ms



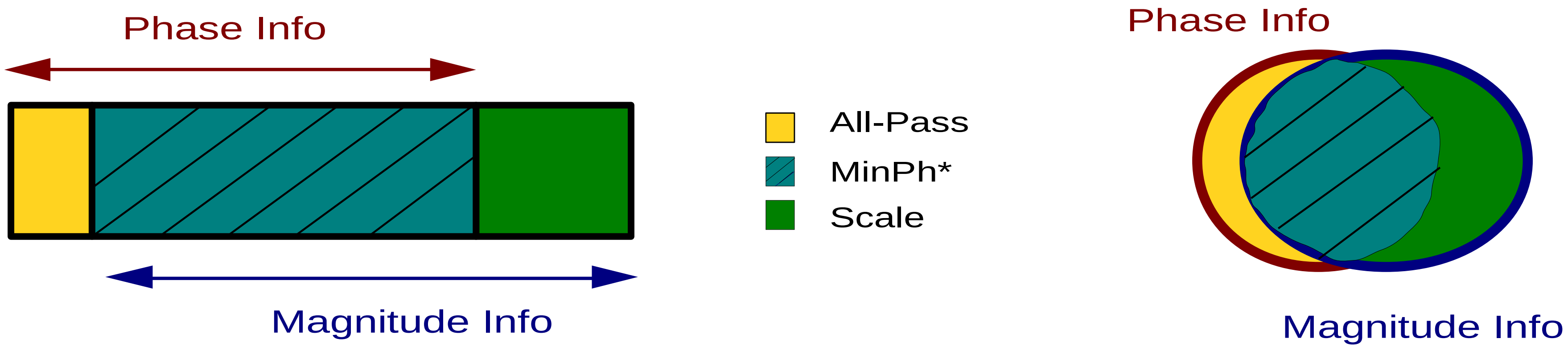


512 ms

– Events are localised exactly, in time

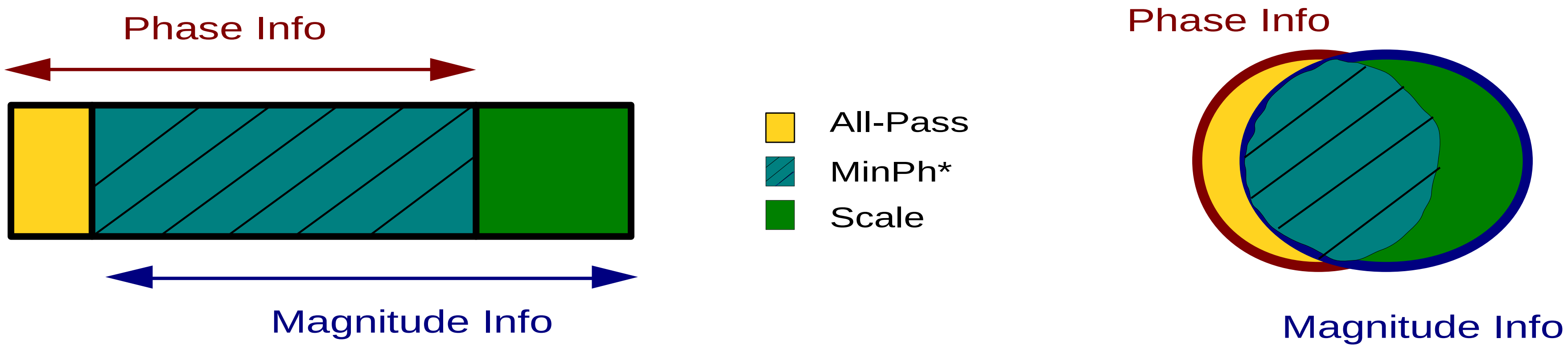
– All-pass part is dominant in long-term

What is the relative importance of each region in short-term?



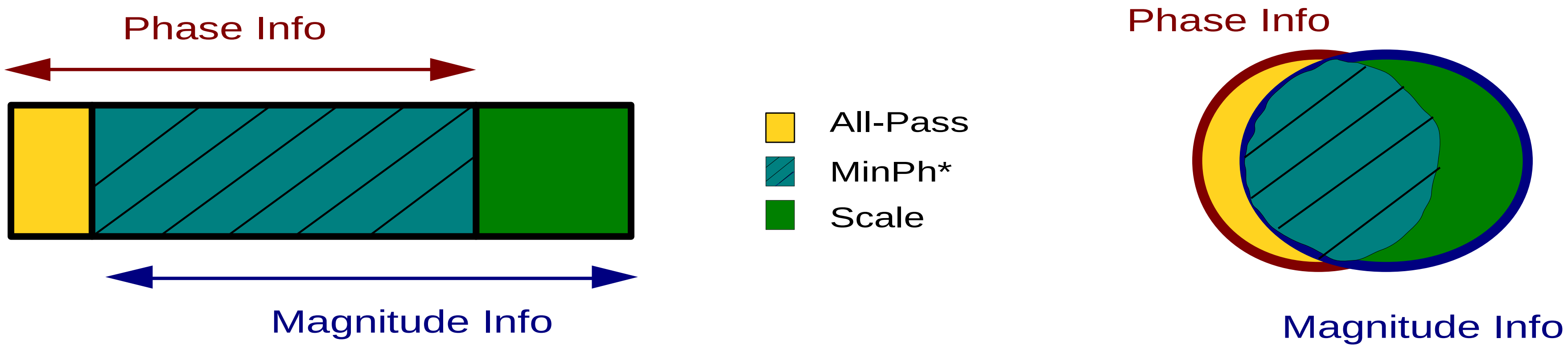
- The Minimum-phase component is dominant

What is the relative importance of each region in short-term?



- The Minimum-phase component is dominant
  - The magnitude spectrum is informative

What is the relative importance of each region in short-term?



- The Minimum-phase component is dominant
  - The magnitude spectrum is informative
  - **AND** also the phase!

Two Important Questions ...

- Is using phase in short-term along with mag. useful?
- If YES, How much? Why?

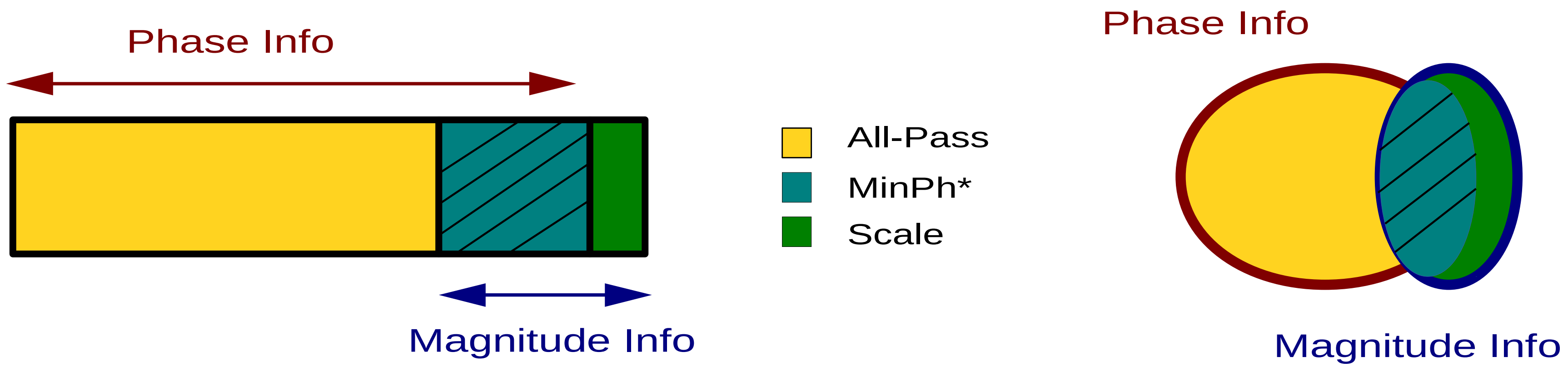
Two Important Questions ...

- Is using phase in short-term <u>along with</u> mag. useful?
  - Yes, it supplies extra info, namely All-pass part
- If YES, How much? Why?
  - The gain of using the *all-pass* part is <u>*limited*</u>

Two Important Questions ...

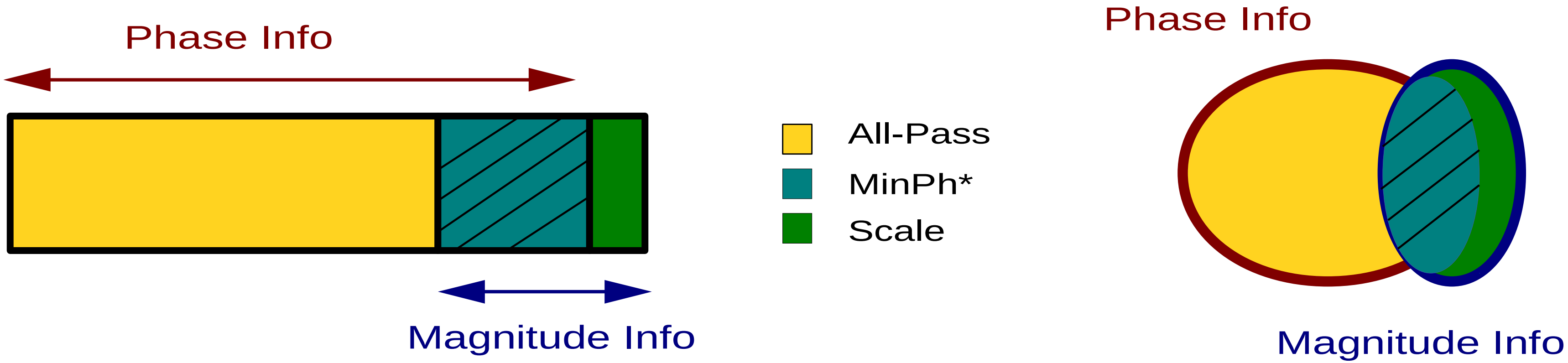
- Is using phase in short-term along with mag. useful?
  - Yes, it supplies extra info, namely All-pass part
- If YES, How much? Why?
  - The gain of using the *all-pass* part is ***limited***
  - In short-term temporal resolution is already high, therefore, no critical need to all-pass timing info
    - Unless stationarity is violated, e.g. for Stops/plosives

What is the relative importance of each region in long-term?



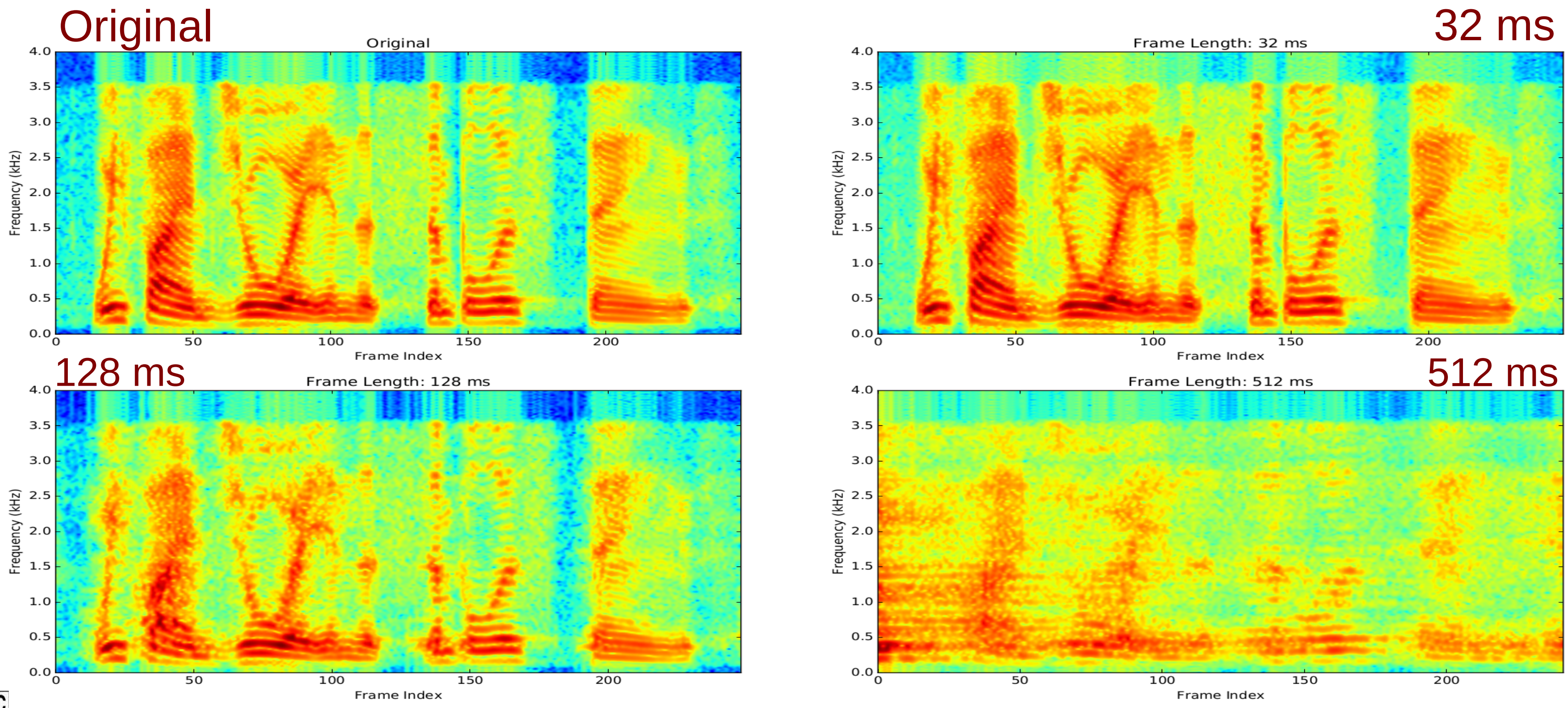
- The all-pass component is dominant

What is the relative importance of each region in long-term?



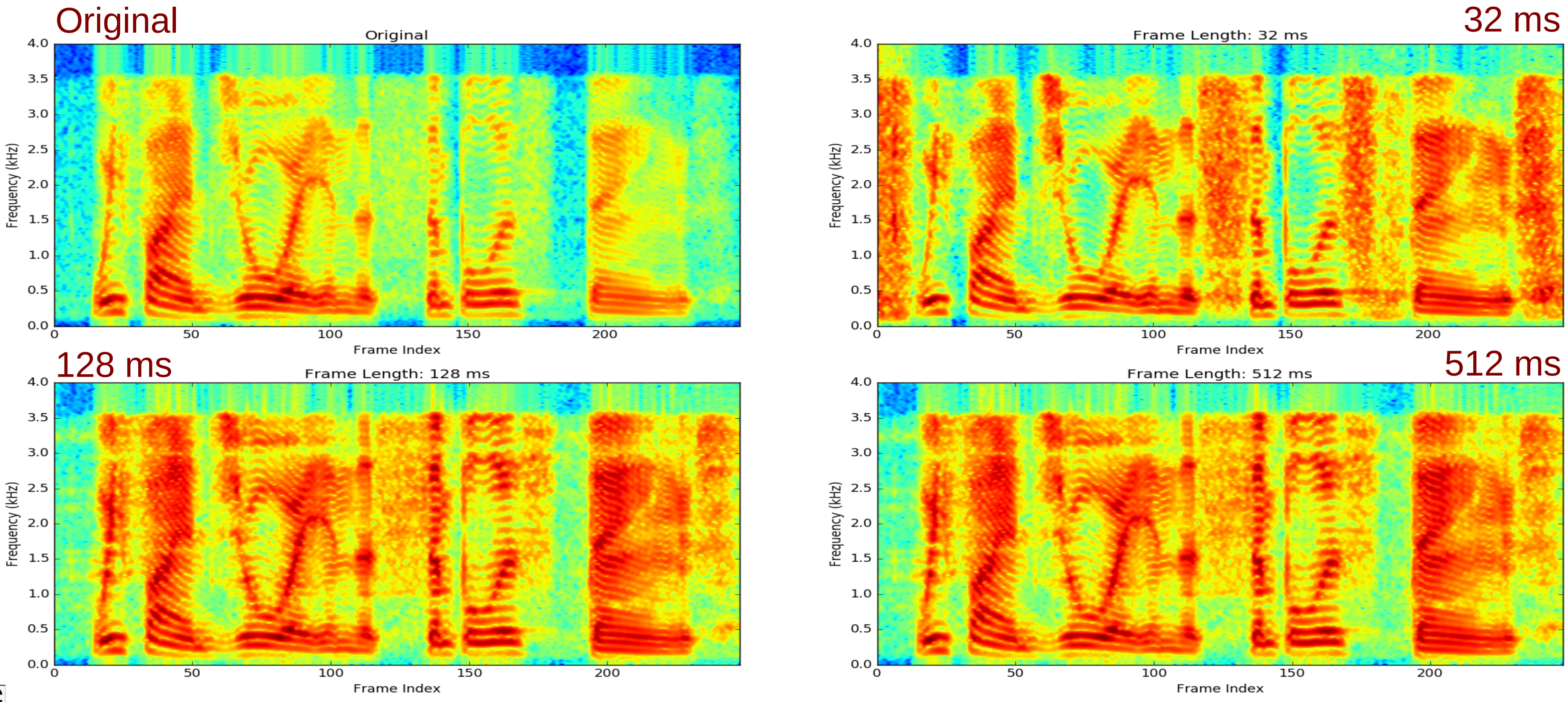
- The all-pass component is dominant
  - The phase spectrum is informative
  - **BUT** magnitude is not

Magnitude-only Reconstructed Speech



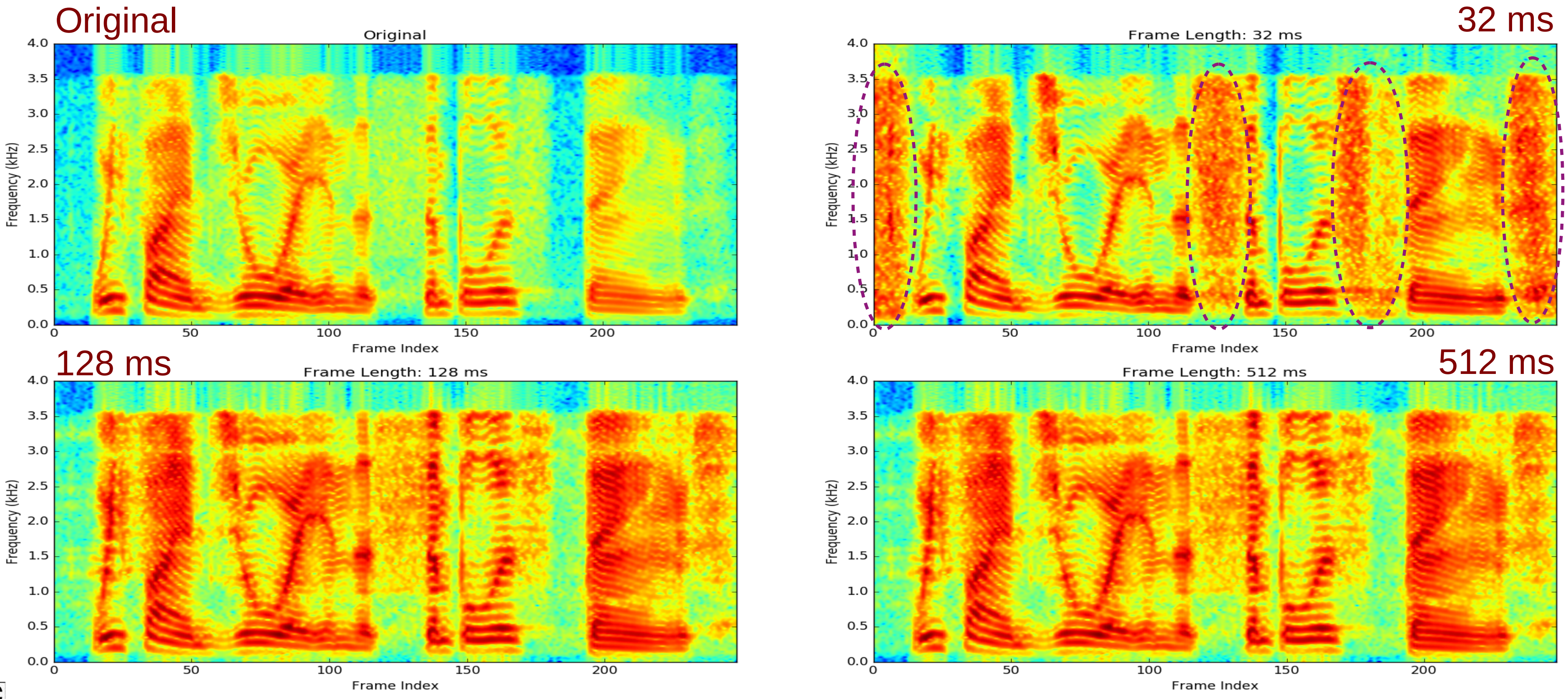
#iterations: 100

Phase-only Reconstructed Speech



#iterations: 100

Phase-only Reconstructed Speech



#iterations: 100

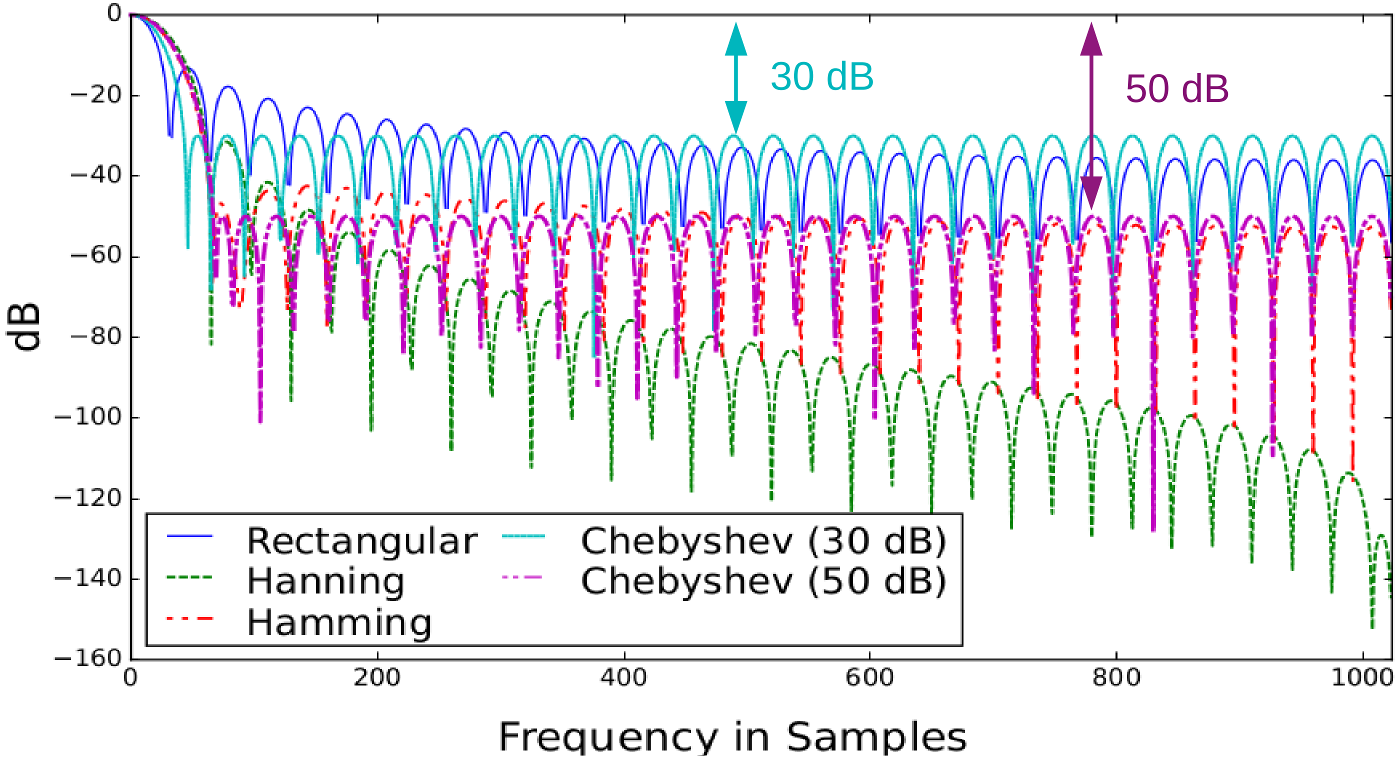
PESQ Scores for Magnitude-only Reconstructed Signal

Window Shape

Frame length (ms)	32	64	128	256	512	1024
Rectangular	4.00	3.91	3.61	3.05	2.36	1.90
Triangular	4.18	3.99	3.54	2.73	1.95	1.18
Hanning	3.34	3.37	3.19	3.18	0.81	0.50
Hamming	4.25	4.09	3.83	3.28	2.66	1.43
Cheb-20	2.42	2.44	2.29	2.18	1.91	1.95
Cheb-25	2.54	2.54	2.40	2.29	2.10	1.88
Cheb-30	2.86	2.65	2.53	2.43	2.27	1.77
Cheb-35	3.40	2.90	2.61	2.57	2.45	1.87
Cheb-40	3.78	3.36	2.88	2.72	2.37	1.53
Cheb-45	4.08	3.60	3.30	2.93	2.58	1.73
Cheb-50	4.18	3.86	3.50	3.09	2.63	1.48
Cheb-80	4.19	4.10	3.95	3.53	1.64	0.77
Cheb-110	3.61	3.54	3.44	3.13	0.75	0.53

– Griffin-Lim
– 100 iterations
– 75% overlap

Windows Spectral Content



PESQ Scores for Magnitude-only Reconstructed Signal

Frame length (ms)	32	64	128	256	512	1024
Rectangular	4.00	3.91	3.61	3.05	2.36	1.90
Triangular	4.18	3.99	3.54	2.73	1.95	1.18
Hanning	3.34	3.37	3.19	3.18	0.81	0.50
Hamming	4.25	4.09	3.83	3.28	2.66	1.43
Cheb-20	2.42	2.44	2.29	2.18	1.91	1.95
Cheb-25	2.54	2.54	2.40	2.29	2.10	1.88
Cheb-30	2.86	2.65	2.53	2.43	2.27	1.77
Cheb-35	3.40	2.90	2.61	2.57	2.45	1.87
Cheb-40	3.78	3.36	2.88	2.72	2.37	1.53
Cheb-45	4.08	3.60	3.30	2.93	2.58	1.73
Cheb-50	4.18	3.86	3.50	3.09	2.63	1.48
Cheb-80	4.19	4.10	3.95	3.53	1.64	0.77
Cheb-110	3.61	3.54	3.44	3.13	0.75	0.53

Best Window:
Hamming

PESQ Scores for Phase-only Reconstructed Signal

Window Shape

Frame length (ms)	32	64	128	256	512	1024
Rectangular	3.51	3.55	3.66	3.78	3.83	3.59
Triangular	1.85	1.96	2.03	2.33	2.58	2.53
Hanning	1.37	2.25	1.92	2.48	0.89	0.77
Hamming	2.41	2.47	2.58	2.80	2.99	2.88
Cheb-20	3.78	3.98	4.04	4.20	4.20	4.04
Cheb-25	3.78	3.98	4.09	4.24	4.17	4.21
Cheb-30	3.73	3.96	4.05	4.20	4.16	4.24
Cheb-35	3.63	3.88	3.97	4.12	4.05	4.08
Cheb-40	3.36	3.65	3.81	3.92	3.86	3.80
Cheb-45	2.92	3.26	3.54	3.64	3.60	3.44
Cheb-50	2.57	2.83	3.14	3.32	3.33	3.08
Cheb-80	1.88	1.85	2.02	2.36	2.72	2.63
Cheb-110	1.88	2.10	2.29	2.53	2.31	1.54

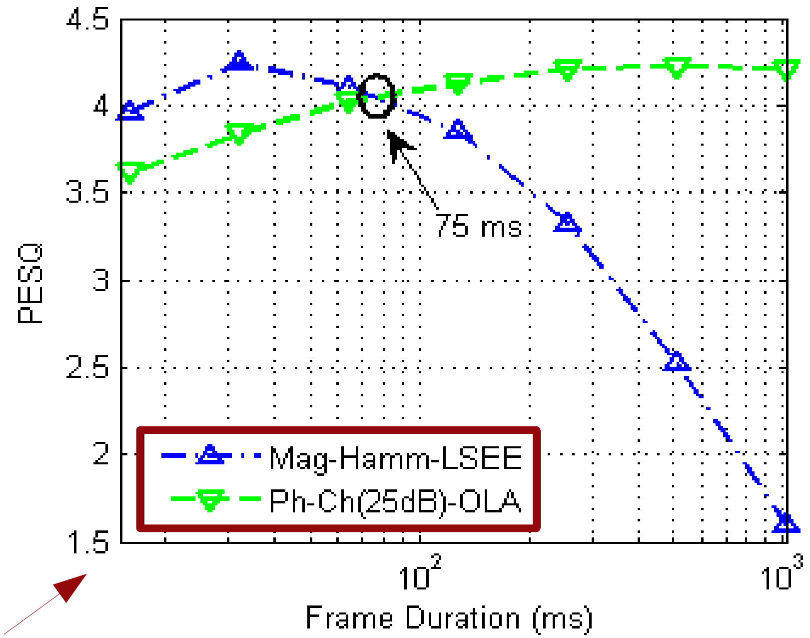
– Griffin-Lim
– 100 iterations
– 75% overlap

PESQ Scores for Phase-only Reconstructed Signal

Best Window:
Chebyshev (25 dB)

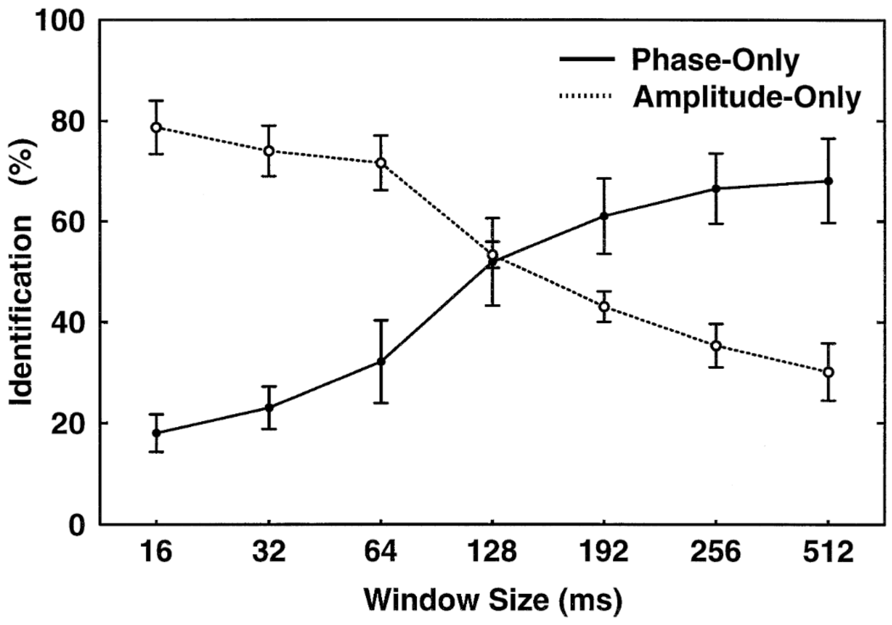
Frame length (ms)	32	64	128	256	512	1024
Rectangular	3.51	3.55	3.66	3.78	3.83	3.59
Triangular	1.85	1.96	2.03	2.33	2.58	2.53
Hanning	1.37	2.25	1.92	2.48	0.89	0.77
Hamming	2.41	2.47	2.58	2.80	2.99	2.88
Cheb-20	3.78	3.98	4.04	4.20	4.20	4.04
Cheb-25	3.78	3.98	4.09	4.24	4.17	4.21
Cheb-30	3.73	3.96	4.05	4.20	4.16	4.24
Cheb-35	3.63	3.88	3.97	4.12	4.05	4.08
Cheb-40	3.36	3.65	3.81	3.92	3.86	3.80
Cheb-45	2.92	3.26	3.54	3.64	3.60	3.44
Cheb-50	2.57	2.83	3.14	3.32	3.33	3.08
Cheb-80	1.88	1.85	2.02	2.36	2.72	2.63
Cheb-110	1.88	2.10	2.29	2.53	2.31	1.54

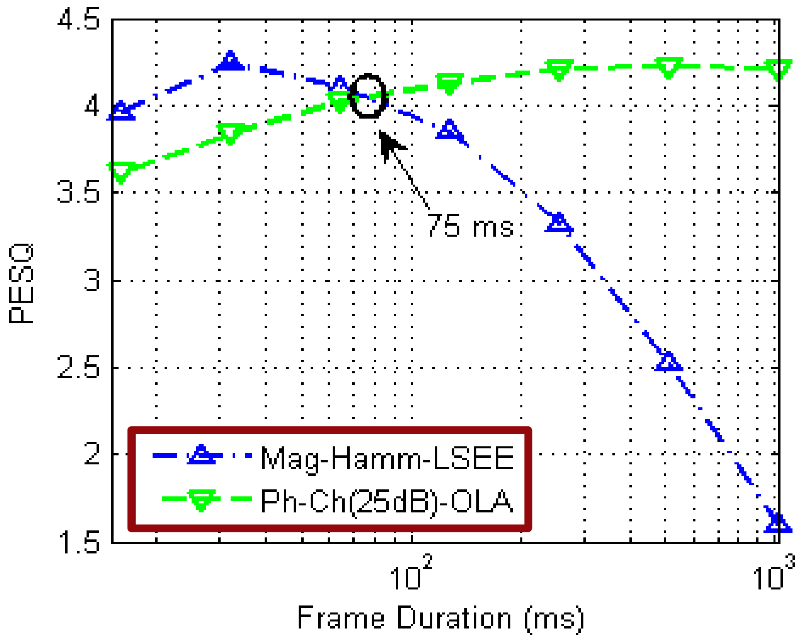
Phase Vs Magnitude



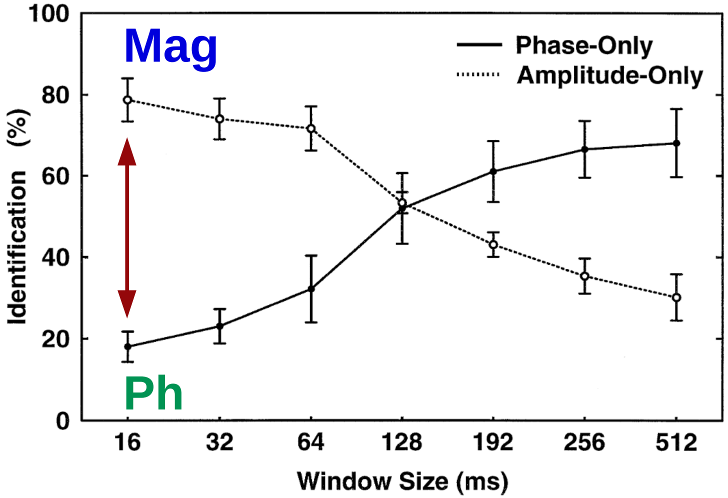
Optimal
Setting

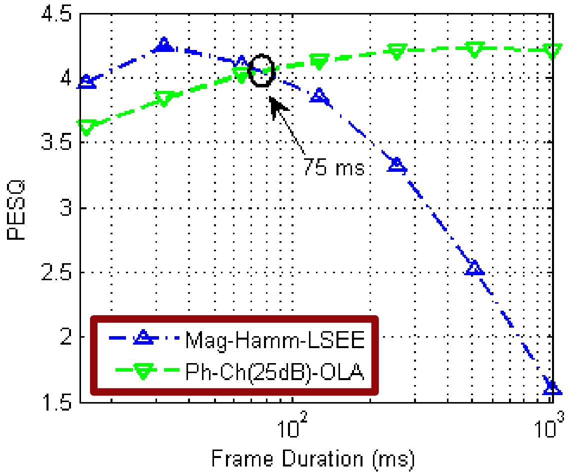
Phase Vs Magnitude





Rather Paradoxical!

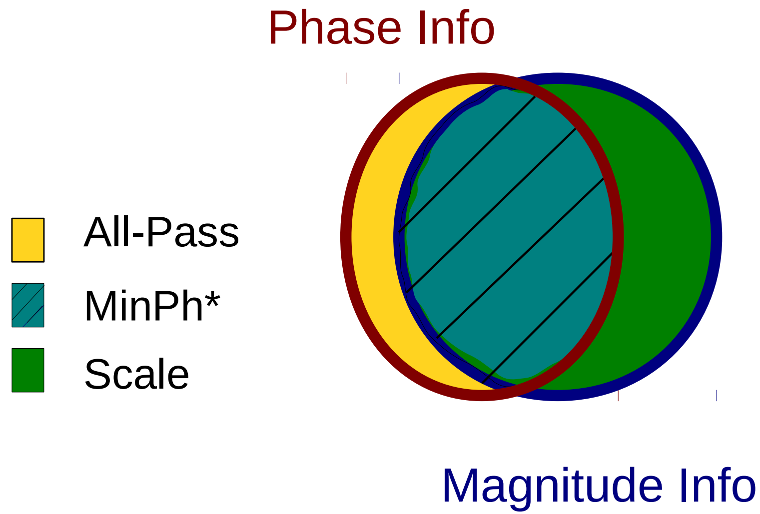




Theorem 3

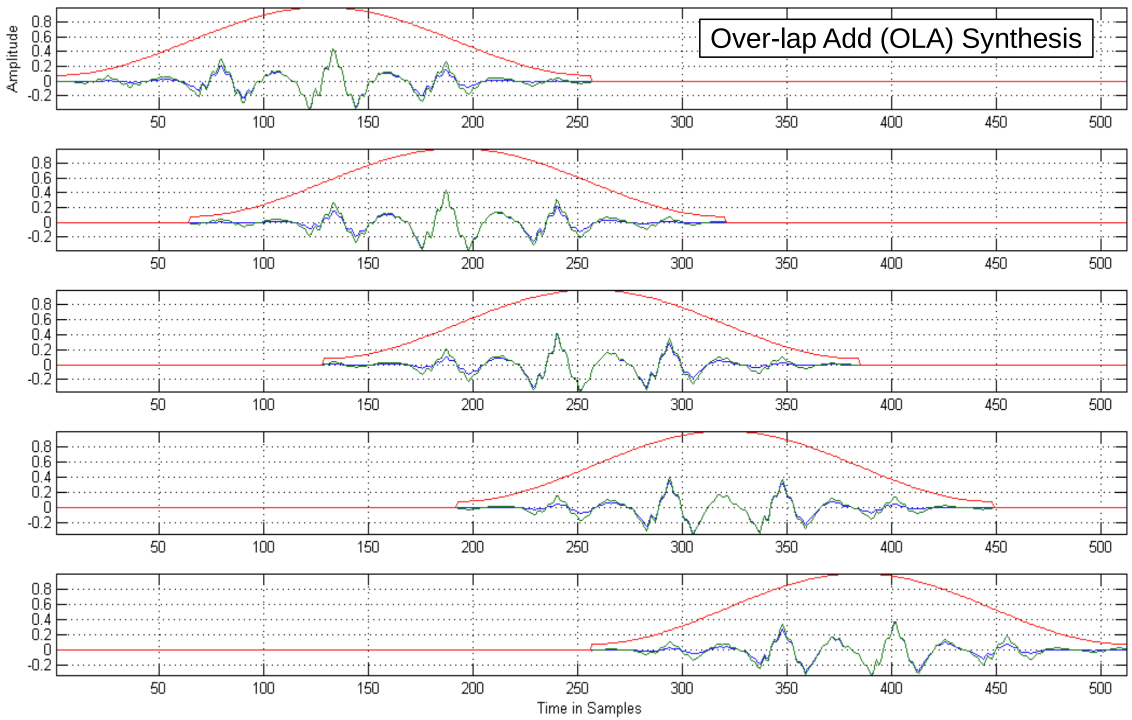
"*A sequence which is known to be zero outside the interval $0 \leq n \leq N - 1$ is uniquely specified to within a scale factor by $N - 1$ distinct samples of its phase spectrum in the interval $0 \leq \omega \leq \pi$ if it has a z-transform with no zeros on the unit circle or in conjugate reciprocal pairs*"

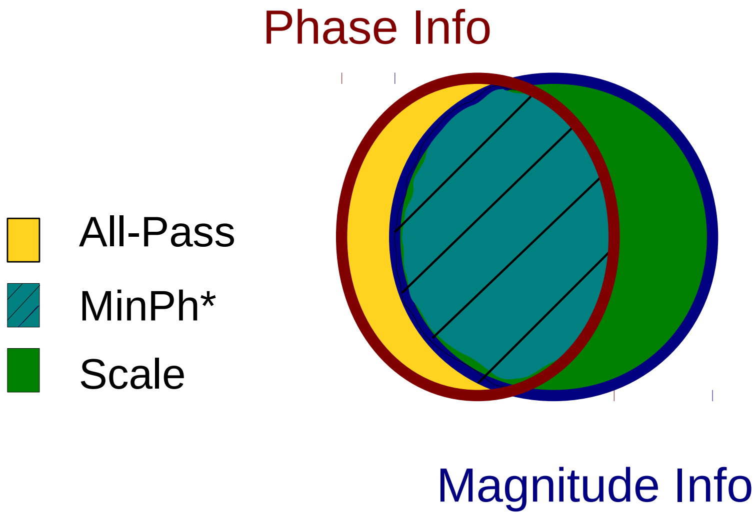
Scale Incompatibility Error (SIE)



$$\text{scale} = \exp(\underbrace{mean(\log|X(\omega)|}_{\hat{x}[0]})$$

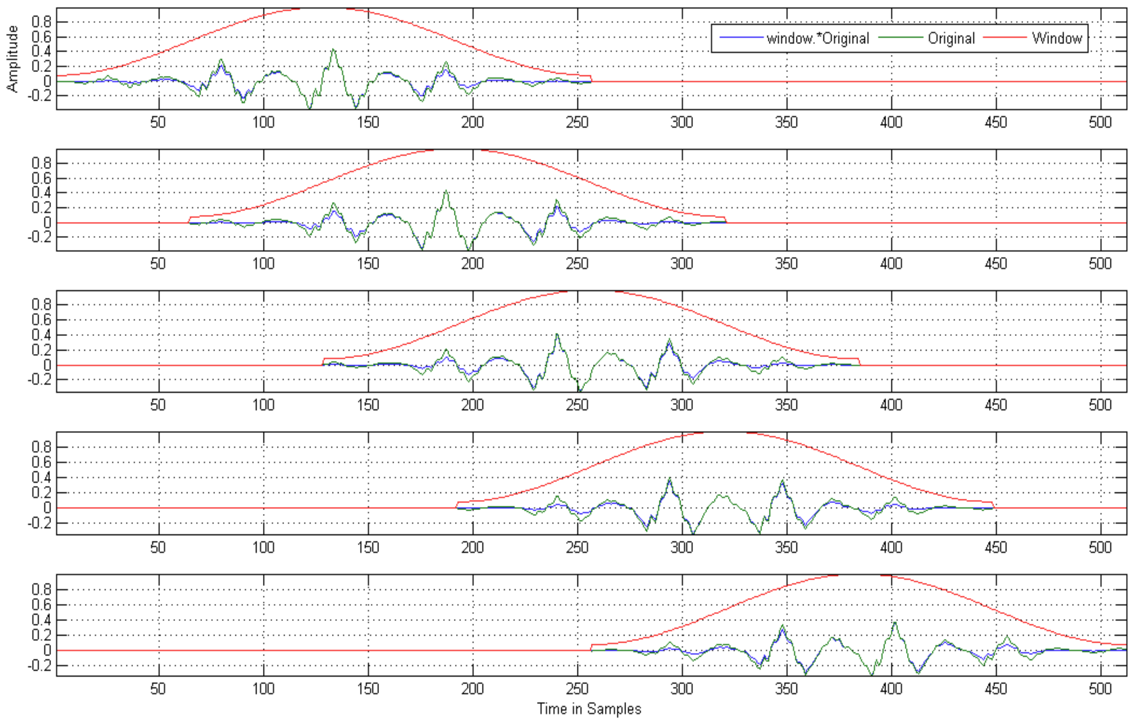
Scale Incompatibility Error (SIE)





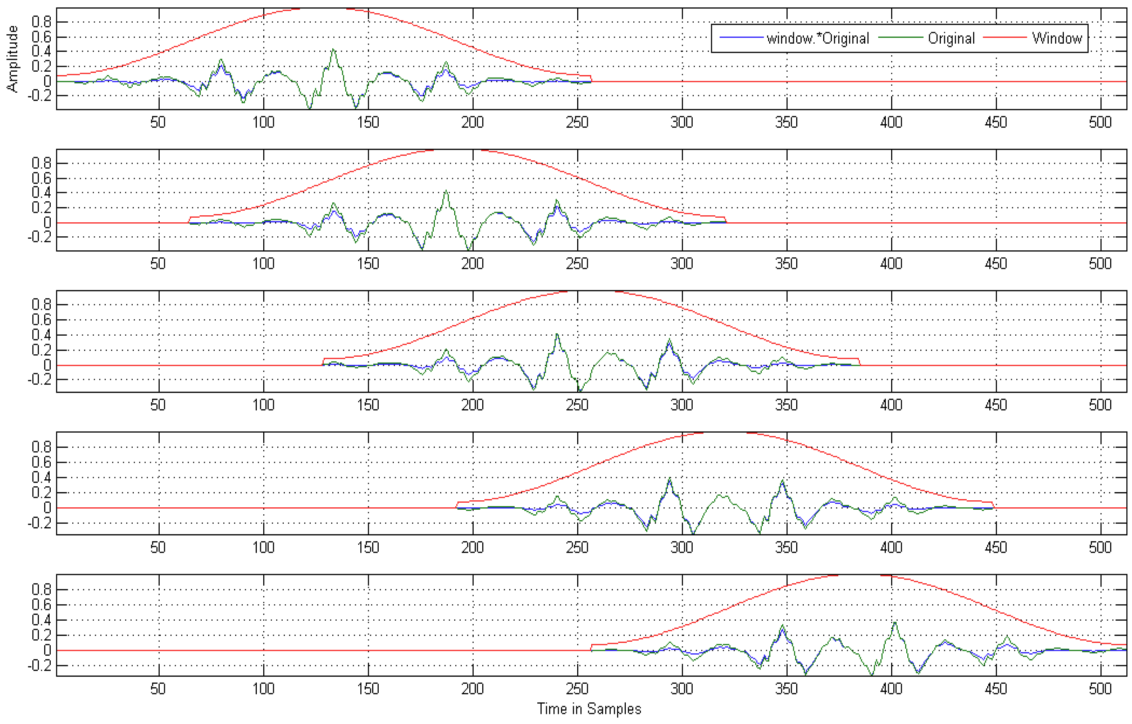
$$\text{scale} = \exp(\underbrace{mean(\log|X(\omega)|)}_{\hat{x}[0]})$$

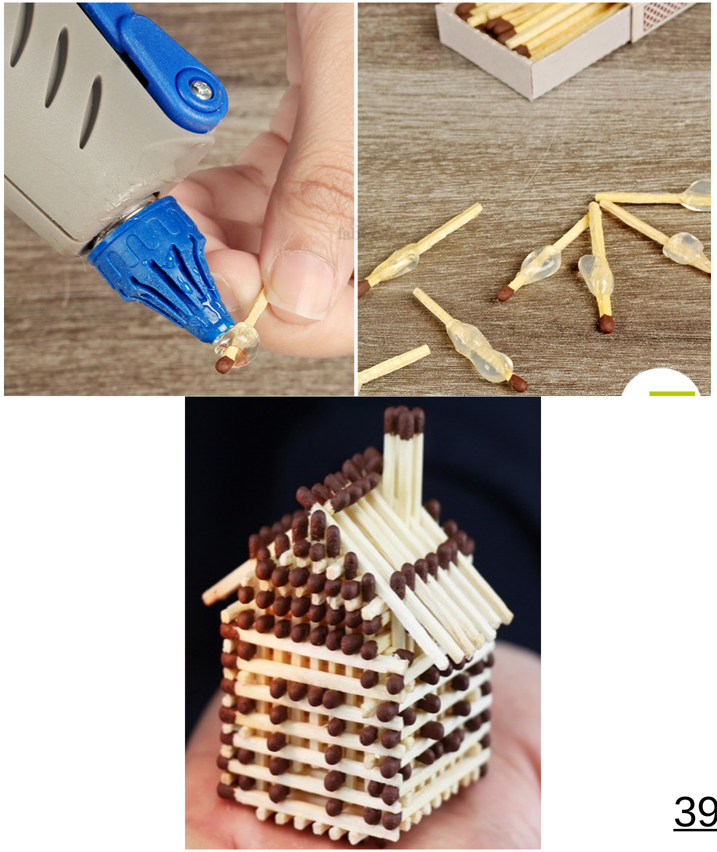
Scale Incompatibility Error (SIE)



Scale error is different for each frame!

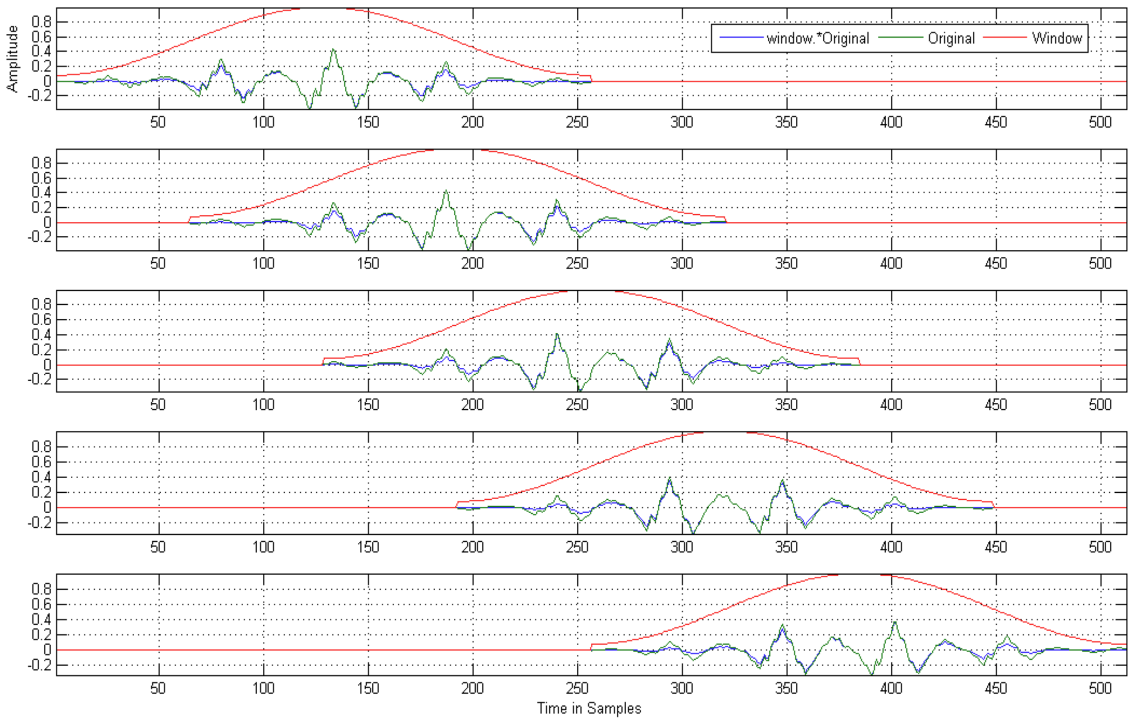
Scale Incompatibility Error (SIE)

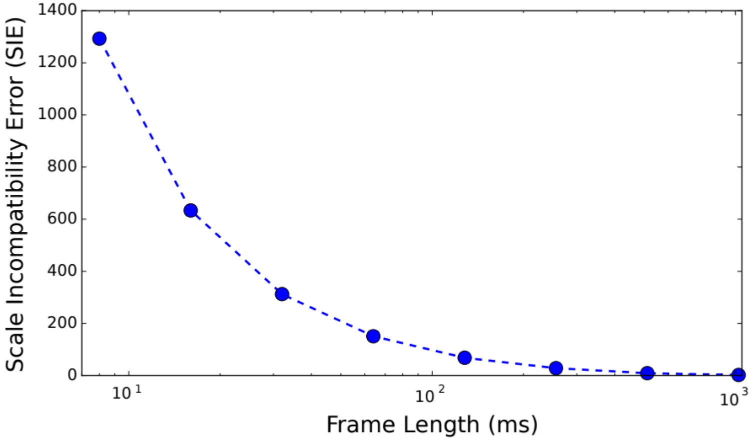




E. Loweimi

Scale Incompatibility Error (SIE)





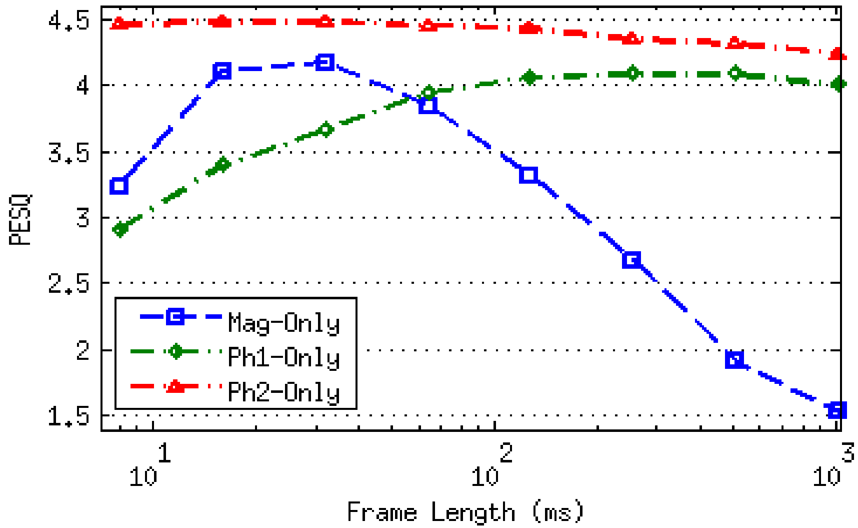
$$SIE = \sum_{n} (1 - \exp(\hat{x}[n, 0]))^2$$

Frame index

Removing SIE

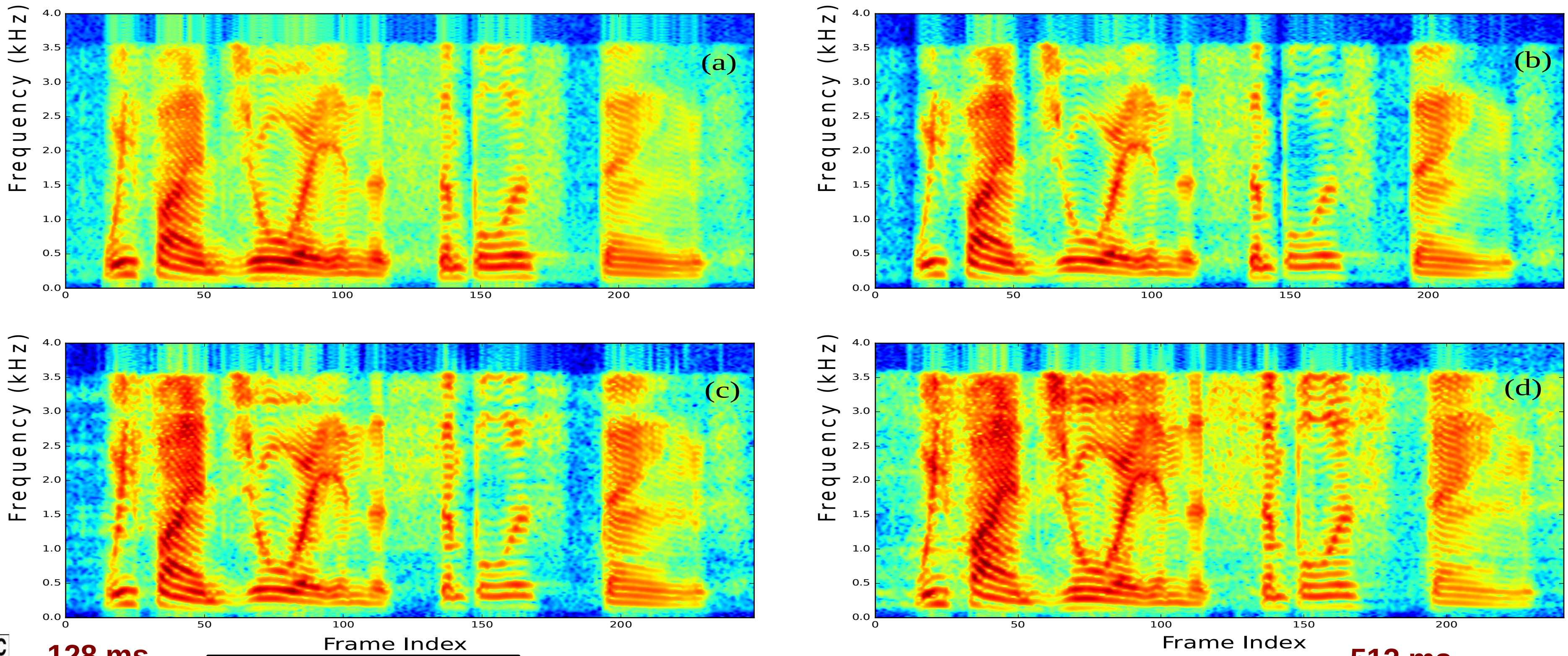
- Initialise the magnitude spectrum with $\exp(\hat{x}[n, 0])$

Mag: Magnitude-only
Ph1: phase-only
Ph2: Phase-only + scale Info



Phase-only Reconstructed Speech + Scale Info

Original **32 ms**



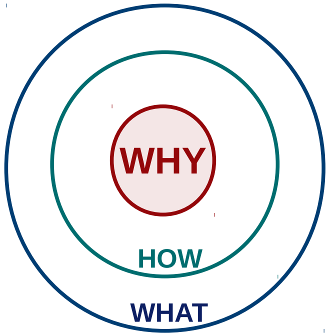
128 ms **512 ms**

#iterations: 100

Summary for Part I

- Signal Information regions
  - All-pass → unique to phase → dominant in long-term
  - Min-Ph* → shared → dominant in short-term
  - Scale → unique to magnitude → important in short-term/synthesis
- Phase spectrum is informative in both short and long-term
- Using Phase along with magnitude in short-term has limited advantage due to minor role of the all-pass part, unless ...
- Phase spectrum includes both source and filter information

Outlines of Part II



- Problems with Phase Spectrum
- Phase Information Content
- Phase-only Speech Signal Reconstruction

Part I

- Source-filter Separation in the Phase Domain
- Applying Phase Filter Component for ASR
- Applying Phase Source Component for F_0 Extraction

Part II

Why phase is more important than magnitude for image reconstruction?

E. Loweimi

That’s it!

- Thanks for your attention
- Q & A