Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 81 (2016) 1–29

www.elsevier.com/locate/specom

# Advances in phase-aware signal processing in speech communication

Pejman Mowlaee[a,*], Rahim Saeidi[b], Yannis Stylianou[c]

[a] *Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria*
[b] *Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[c] *Computer Science Department, University of Crete, Crete, Greece*

Received 1 July 2015; received in revised form 8 April 2016; accepted 11 April 2016
Available online 9 May 2016

**Abstract**

During the past three decades, the issue of processing spectral phase has been largely neglected in speech applications. There is no doubt that the interest of speech processing community towards the use of phase information in a big spectrum of speech technologies, from automatic speech and speaker recognition to speech synthesis, from speech enhancement and source separation to speech coding, is constantly increasing. In this paper, we elaborate on why phase was believed to be unimportant in each application. We provide an overview of advancements in *phase-aware signal processing* with applications to speech, showing that considering phase-aware speech processing can be beneficial in many cases, while it can complement the possible solutions that magnitude-only methods suggest. Our goal is to show that phase-aware signal processing is an important emerging field with high potential in the current speech communication applications. The paper provides an extended and up-to-date bibliography on the topic of *phase aware speech processing* aiming at providing the necessary background to the interested readers for following the recent advancements in the area. Our review expands the step initiated by our organized special session and exemplifies the usefulness of spectral phase information in a wide range of speech processing applications. Finally, the overview will provide some future work directions.


*Keywords:* Phase-aware speech processing; Phase-based features; Signal enhancement; Automatic speech recognition; Speaker recognition; Speech synthesis; Speech coding; Speech analysis.

## 1. Introduction

In many everyday-life applications a reliable speech communication system is in demand whose performance is expected to be robust enough to deliver a certain high quality of service to establish a reliable speech communication medium. For this purpose, it is highly important to verify the robustness of the underlying speech communication interface against impairments that occur due to some acoustic interference, background noise or those introduced by the failure in the communication channel, which can be modeled, for example, as the additive background noise or reverberations in the room, respectively. In a speech communication application, the design goal is to deliver enough flexibility and reliability such that it provides a certain quality of service that may differ depending on the specified application. A full end-to-end speech communication chain from microphone to receiver end involves several blocks including: speech analysis, multi-channel processing (beamformer), single-channel signal restoration entailing speech enhancement, source separation and artificial bandwidth extension. Depending on the desired application, one might be interested in playing back the reconstructed signal at the receiver end and, as a consequence, a speech synthesis block is required. Alternatively, in biometric and dictation applications, the goal is to recognize, or verify the identity of the speaker (speaker identification/verification) or to recognize the phonemes and words spoken by the transmitting party (automatic speech recognition). Another application is speech watermarking, for which certain cover data for copyright issues are required to be embedded in speech without modifying the speech quality of the watermarked signal still be robust against spoofing attacks.

* Corresponding author. Tel.: +43 316 873 4433; fax: +43 316 873 7941.
*E-mail addresses:* pejman.mowlaee@tugraz.at (P. Mowlaee), rahim.saeidi@aalto.fi (R. Saeidi), yannis@csd.uoc.gr (Y. Stylianou).

http://dx.doi.org/10.1016/j.specom.2016.04.002
0167-6393/

Finally, in speech coding the phase information is important to fulfill the bits budget while delivering a transparent speech quality at the receiver end. This is feasible by keeping the phase quantization errors below a perceptual threshold, determined by just noticeable difference, not perceptible by the listener.

The common strategies described in the aforementioned speech processing applications mainly focus on spectral amplitude modification and signal processing methods that are used to filter the spectral amplitude of the speech signal. More recently, the research attention towards incorporating the spectral phase information has increased. In particular, individual steps have been taken by researchers through different applications from different sub-communities in speech signal processing. In order to unify the attempts made by the researchers in such diverse communities and to gather them together to share their findings, a special session at INTERSPEECH 2014 entitled *Phase Importance in Speech Processing Applications* (Mowlaee et al., 2014) has recently been organized.

In this paper, we present an extended literature review of earlier viewpoints about the importance of phase information in speech signal processing. We exemplify the importance of phase information in several speech processing applications and further demonstrate the positive impact brought about by phase-aware signal processing in each of the aforementioned applications. Finally, recent and future directions towards applying phase-aware signal processing to tackle different speech applications are pointed out. Our work extends the initial step made at special session (Mowlaee et al., 2014). In contrast, here we present a broader picture of phase-aware signal processing and provide a detailed overview of the advances made by researchers in the field. Following the literature review, we present several useful representations for phase spectrum that illustrate certain structures in the phase domain. The common short-time Fourier transform (STFT) spectral phase representation exhibits no useful structure or distribution in the phase domain due to its cyclic wrapping nature. However, new presentations provide means to explore various properties and details of speech signals (e.g., continuity, formant, harmonicity, or statistical explanation captured by mean, variance and probability density function). We then describe the phase-based derived features appearing in the literature for different speech applications. To highlight the importance of phase-aware signal processing, we give examples in several speech applications: single-channel and multi-channel speech enhancement, source separation, automatic speech recognition, speaker recognition, speech coding, speech analysis/synthesis, digital speech watermarking, and speech quality estimation. For each of these applications, we present the impact of phase information and explain the state-of-the-art methods that neglect the phase information, as well as more recent ones that incorporate phase information into their proposed solutions.

Several recent attempts have presented partial overviews of some aspects of phase processing for speech applications (Alsteris and Paliwal, 2007b; Gerkmann et al., 2015; Mowlaee and Kulmer, 2015a; 2015b):

- Alsteris and Paliwal (2007b) reviewed experimental results for short-time phase spectrum usage in speech processing. This work demonstrated the usefulness of phase spectrum for automatic speech recognition, human listening and speech intelligibility. Paliwal et al. (2011) studied the importance of phase information for speech enhancement and proposed a phase spectrum compensation (PSC) method using the conjugate symmetric property of the discrete Fourier transform (DFT). These subjects are covered in Section 4.1.3.
- Gerkmann et al. (2015) reviewed phase processing for single-channel speech enhancement. In contrast to this narrow focus on phase-aware speech enhancement, the current review presents a unified approach to phase-aware speech communication including other applications such as speech analysis, speech synthesis, speaker/speech recognition, speech coding, speech watermarking, source separation and multi-channel speech processing. These subjects are covered in Section 4.
- Mowlaee and Kulmer (2015a); 2015b) reviewed phase estimation in speech enhancement and demonstrated the potential and limits of the phase estimation methods demonstrated with a comparative study. The main contribution in Mowlaee and Kulmer (2015a); 2015b) was on the phase estimation in noise and incorporate it for improved signal reconstruction. Efficient techniques for estimating phase in noisy environments are reviewed in Section 4.1.3.

In this paper, throughout the extended up-to-date bibliography we provide a guidance for our readers to a summary of what has been done with regard to phase processing of speech signal. For readers more interested in the specific solution, we refer to bibliography where a list of early to latest contributions can be found. Our goal is to help the interested readers to have a quick access to the novel papers, and get to know the earlier and current innovations and efforts made to explore *phase-aware signal processing*. We also elaborate on why phase was believed to be unimportant in each application and why phase processing is recently attracting increasing interest in speech processing community. The usefulness of spectral phase information will be exemplified in a wide range of speech processing applications (speech analysis/synthesis, speech enhancement, source separation, speech coding, speech quality estimation, speech watermarking, automatic speech recognition, and speaker recognition), providing a reference for the researchers who are just starting with their research topic related to the phase-aware signal processing. The current paper is an extensive overview to date on the topic of phase-aware speech processing and provides a unique picture that depicts the importance of phase assessment in many aspects of speech technology.

The paper is organized as follows: Section 2 presents the controversial standpoints in the literature with regard to the importance of phase information in speech signal processing; Section 3 presents useful phase representations derived from the short-time Fourier transform phase; It also focuses on phase-based features used in speech processing applications;