

# Abstract of Doctoral Dissertation

## Title : Short-time Fourier Transform Phase Reconstruction Using Deep Neural Networks and Phase-based Features

Doctoral Program in Advanced Information Science and Engineering  
Graduate School of Information Science and Engineering  
Ritsumeikan University

グエン ビン ティエン  
NGUYEN Binh Thien

Recently, the use of speech processing in a time-frequency domain that handles the phase spectrum in addition to the amplitude one has been increasing because many studies have revealed the importance of phase. Motivated by this, the objective of this thesis is to address two key aspects of phase processing: exploring the hidden information within the phase and developing techniques for phase reconstruction.

To extract useful information from the phase, this thesis examines several conventionally used phase-based features. Additionally, two novel phase-based features, namely the derivative of instantaneous frequency and the inter-frequency phase difference, are introduced and their properties are investigated.

Regarding phase reconstruction, this thesis focuses on a two-stage approach that indirectly estimates the phase through the instantaneous frequency (IF) and the group delay (GD). Conventional methods for the second stage either overlook the significance of high-amplitude time-frequency bins, such as the least squares-based method, or lack a solid model, as in the average-based method. To address these limitations, this thesis proposes improvements to the second stage of two-stage algorithms by utilizing the von Mises distribution-based maximum likelihood and weighted least squares techniques.

In the online setting for phase reconstruction, this thesis concentrates on deep neural network (DNN)-based methods. Conventional DNN-based methods do not consider the distinct properties of the phase at different time-frequency bins, which may lead to limitations in training the DNNs. To overcome this, this thesis introduces loss functions for phase reconstruction that incorporate frequency-specific and amplitude weights to distinguish the importance of phase elements based on their properties. To enhance generalization, the thesis augments the data by randomly shifting the signals in the time domain for each epoch during training.

Both the subjective and objective experiments confirm the superior performance of the proposed phase reconstruction methods compared to conventional methods.

# 博士論文要旨

## 論文題名：深層ニューラルネットワークと位相ベース特徴を用いた短時間フーリエ変換位相再構成

立命館大学大学院情報理工学研究科  
情報理工学専攻博士課程後期課程

グエン ビン ティエン  
NGUYEN Binh Thien

最近、多くの研究によって位相の重要性が明らかにされたため、振幅だけでなく位相スペクトルを扱う時間周波数領域での音声処理の利用が増えています。この動機を受けて、この論文の目的は、位相処理の2つの重要な側面、つまり位相内に潜む情報の探索と位相再構築のための技術の開発に取り組むことです。

位相から有用な情報を抽出するために、この論文では従来から使われてきたいくつかの位相ベースの特徴を検討します。さらに、瞬時周波数の導関数と周波数間位相差という2つの新しい位相ベースの特徴を導入し、その特性を調査します。

位相再構築に関しては、この論文では瞬時周波数 (instantaneous frequency: IF) とグループ遅延 (group delay: GD) を介して位相を間接的に推定する2段階のアプローチに焦点を当てています。2段階のアルゴリズムの第2段階の従来の方法は、最小二乗法に基づく方法など、高振幅の時間周波数ビンが考慮されていないことや、平均ベースの方法のように頑健なモデルがないという制約があります。これらの制約に対処するため、この論文では、von Mises 分布に基づく最尤推定と重み付き最小二乗法の手法を用いて、2段階アルゴリズムの第2段階の改善案を提案します。

位相再構築のオンライン設定では、この論文ではディープニューラルネットワーク (deep neural network: DNN) を基にした手法に焦点を当てています。従来の DNN ベースの手法では、異なる時間周波数ビンにおける位相の特性を考慮しないため、DNN の訓練に制限が生じる可能性があります。これを克服するため、この論文では位相再構築のための損失関数を提案し、周波数固有の重みと振幅の重みを用いて位相要素の重要性を特性に基づいて区別します。さらに、汎化性能を向上させるため、訓練時に各エポックで信号をランダムに時間領域でシフトさせるデータの拡張も行います。

主観的および客観的な実験の両方により、従来の方法と比較して提案された位相再構築手法の有効性を確認できました。