

博士論文要旨

論文題名：光レーザーマイクロホンを用いた音声強調システムに関する研究

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

サイ セイカイ
CAI Chengkai

音声収集は日常生活でよく使用される情報獲得の方法であり、録音するために様々なデバイスが開発されていた。しかし、災害救助や防犯セキュリティなどの場合において、遠方から音声を取得することが必要となる。このような場合において、通常の音声収集技術を適用すると、遠方からの音声を観測することが困難だけでなく、マイクロホン周囲の雑音も混入する。ここで、光レーザーマイクロホンを用いて、音源の近くにある物体を照射することで、音源により発生した振動を計測し、その振動を音に復元する技術が提案された。一般的なマイクロホンと比べて、レーザー光はよい直進性を持つため、遠方からの音を計測することが可能し、マイクロホン周囲の雑音に影響されない特徴を持っている。

しかし、この技術は幾つの問題点が存在する。光レーザーマイクロホンは音源付近の振動物体を介して音声を収録するため、観測音声の音質は被照射物体の振動特性に依存する。例えば、振動物体の振幅応答は小さい周波数帯域において、音声成分は欠落する。また、表面の粗い物体を照射する際に、反射光の強度は小さくなり、雑音は混入する。高音質の音声を取得するために、光レーザーマイクロホンの観測音声に対して、音声強調などの処理を適用することが必要となる。

本論文では、深層学習を利用した光レーザーマイクロホンのための音声強調手法幾つを提案する。被照射物体は既知の場合において、スペクトルと時間波形それぞれに基づいて、二段階処理深層学習を利用した両手法を提案する。スペクトルに基づいた手法は、パワースペクトルを復元する上、位相スペクトルの復元も考慮し、2つのネットワークを利用することで、パワースペクトルと位相スペクトルそれぞれを処理する。時間波形に基づいた手法は、観測音声の低域と高域における歪みの違いに着目する。まず、低域成分の雑音を除去する、そして強調後の低域成分を利用して高域成分を復元する。被照射物体は未知の場合において、まず、パワースペクトルのフレーム方向の差分を利用して、パワースペクトルの包絡を推定する。そして、観測音声の基本周波数情報を抽出し、推定したパワースペクトルの包絡を用いて、強調音声のパワースペクトルを復元する。最後に、パワースペクトルから位相スペクトルを推定し、強調音声を生成する。提案した3つの手法において、客観的評価実験を行った結果、提案手法の有効性を確認した。

Abstract of Doctoral Dissertation

Title : Speech Enhancement System with Optical Laser Microphone

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

サイ セイカイ
CAI Chengkai

Speech measurement is a common method of information acquisition in daily life. For speech acquisition various devices have been developed. However, in the case of disaster relief and crime prevention security, it is necessary to acquire speech from a long distance. In such case, the ordinary speech acquisition method is not useful. Here, using a laser Doppler vibrometer (LDV) to acquire speech in distance by measuring the vibration cause by sound source was proposed. Since the laser beam has straightness, it can measure the sound from very far away.

However, there are several problems in this technology. Since the sound quality of the observed speech depends on the vibration characteristics of the object, speech component is lacking in the small amplitude response frequency bands. Also, when irradiating an object with a rough surface, the intensity of the reflected light is reduced that noise mixed. Therefore, to obtain high-quality speech, it is necessary to apply speech enhancement processing for the observed speech.

In this paper, we proposed 3 speech enhancement methods for LDV using deep learning. In the case that the irradiated object is known, we propose 2 methods based on the spectral and waveform, respectively. The spectrum-based method first restores the power spectrum and then restores the phase spectrum. Waveform-based method first suppressing noise in the low frequency bands and then estimate the high-frequency component. In case of irradiated object is unknown, the envelope of the power spectrum of enhanced speech is first estimated. Then, the power spectrum is reconstructed using extracted fundamental frequency and reconstructed envelope. Finally, the phase spectrum is estimated from the power spectrum and the enhanced speech is generated.

Objective evaluations are carried out to evaluate the three proposed methods. As the results of the experiments, the effectiveness of the proposed method was confirmed.