氏　　名	RAHMAN MD. SAIFUR
博士の専攻分野の名称	博士（学術）
学位記号番号	博理工乙第 257 号
学位授与年月日	令和 2 年 9 月 23 日
学位授与の条件	学位規則第 4 条第 2 項該当
学位論文題目	Pitch Extraction for Speech Signals in Noisy Environments （雑音環境下での音声信号のピッチ抽出）
論文審査委員	委員長　教　授　島村　徹也
	委　員　教　授　小室　　孝
	委　員　教　授　小林　貴訓
	委　員　准教授　大久保　潤

論文の内容の要旨

The pitch period is defined as the inverse of the fundamental frequency of the excitation source from the voiced speech signal. The pitch period (in short, pitch) or fundamental frequency is a prominent parameter of speech and highly applicable for speech-related systems such as speech coding, speech recognition, speech enhancement, speech synthesis and so on. The pitch and fundamental frequency give the same meaning, while the pitch is inherently interpreted as the perception of the fundamental frequency. The pitch is generated from the vibration of the vocal cord causing periodicity in the speech signal.

Pitch extraction has proven to be a difficult task even for speech in a noise-free environment. The clean speech waveform is not really periodic; it is quasi-periodic and non-stationary. A large number of pitch extraction methods have been reported to deal with the noise-free environment. On the contrary, a small number of researchers attempt to extract the pitch in noisy environments. Under noisy environments, the periodic structure of the speech signal is destroyed so that the pitch extraction becomes an extremely complicated task. Therefore, the reliability and accuracy of the pitch extraction methods face real challenges in noisy environments.

From the above observations, the objective of this dissertation is to develop some approaches which are effective to handle the speech signals in the real application without any complicated post processing where speech signals are corrupted by noise. Some conventional state-of-the art approaches rely on a complicated post processing technique for pitch extraction. In this dissertation, we focus on simple and efficient approaches that are proposed and implemented to solve the factors that degrade the performance of pitch extraction methods.

In this dissertation, firstly, we propose the use of fourth-root spectrum instead of log spectrum for increasing the pitch extraction accuracy in noisy environments. To get clear harmonics, lifter and clipping operations are followed. When the resulting spectrum is transformed in the time domain by means of discrete Fourier transform, the pitch extraction is robust against narrow-band noise. When the above resulting spectrum is amplified by a power calculation and transformed in the time domain, the pitch extraction is robust against wide-band noise. These properties are investigated through exhaustive experiments in a variety of noise types. Computational time to be required is also studied. The

experimental results based on above properties demonstrate the effectiveness of the new approach for improving the performance of the pitch extraction. Also, the performance of this method sometimes deteriorates by the windowing effect. This method utilizes the Hanning window function which does not always better perform to extract pitch in noisy environments.

To improve the performance of the extraction accuracy, the second approach considers an advancing trend of recent techniques for pitch extraction of speech in noisy environments. Windowing effects are discussed analytically, and it is insisted that the Rectangular window should be proactively used instead of the popular Hanning or Hamming window. In a variety of noise environments, a performance comparison of the conventional pitch extraction methods is conducted, and as a result, we take a standpoint to support the autocorrelation (ACF) method. Incorporating accumulation techniques, three types of pitch extraction approaches are developed.

Through experiments, it is shown that the proposed approaches commonly have the potential to provide better performance for pitch extraction without relying on a complicated post processing technique.

論文の審査結果の要旨

当学位論文審査委員会は、令和２年８月７日に論文発表会を公開で開催した。その後、発表に対する質疑応答と学位論文の内容の審査を行った。以下に、審査結果の要約を示す。

本論文は、音声信号が有する周期性であるピッチ周期（基本周波数の逆数）を、観測波形から高精度に抽出する問題を取り上げている。ピッチ抽出は、音声合成、音声強調、話者認識などの多くの音声情報処理システムに利用できることから、従来多くの研究がなされてきたが、最近では、雑音環境の中においても、高精度な抽出結果が求められるようになってきた。しかしながら、本来音声信号は時変的なものであり、無雑音環境下においてさえも、100 パーセントの正確な抽出はできない。また、雑音環境下においては、雑音の特性の影響を受け、ピッチの抽出精度はさらに劣化してしまう。したがって、これまでに雑音環境下においてさえも確約して利用できる音声信号のピッチ抽出手法は見出されておらず、数多くの手法が提案されているのみである。このような研究背景において、本論文では２つの観点に着目している。１つは、考慮する雑音の種類である。従来においては、雑音環境を考慮していたとしても、２, ３種類の限られた場合のみが検討されていた。しかし、本論文では、８種のそれぞれ異なる雑音の種類を考察対象として取り上げ、より広範囲な雑音環境での音声信号のピッチ抽出問題を検討している。もう１つは、核となるアルゴリズム部分のみの検討を行っている点が、本論文の重要な着目点である。本来、ピッチ抽出は、短時間で切り取られた音声波形に対し、一括処理において、ピッチ抽出の結果を算出し、次の短時間に区切られた音声波形に対し、一括処理によりピッチ抽出を行うというように、フレーム処理が基本として施される。抽出結果が、短時間でのフレーム毎に得られることから、ピッチ抽出には、各フレームから得られた抽出結果を後処理で補正を施すというアプローチが取られることがある。この後処理に時間をかけることで、全体としてピッチ抽出精度が向上するが、この後処理をピッチ抽出手法の中に組み入れて考えるか、組み入れずに別に考えるかで、ピッチ抽出の評価が大きく変化してしまう問題があった。本論文では、この後処理の部分はピッチ抽出手法に含めず、核となるアルゴリズム部分のみを検討している。したがって、任意の有益な後処理手法を組み合わせることが可能となる方法を検討している。このような着目点から、本論文では、雑音環境下で高精度な抽出結果を与える、２つの異なる音声ピッチ抽出手法を提案し、それぞれの有効性を検証している。

まず第１章は、序論である。人間の発声の原理を説明し、音声生成モデルから、音声が有声音と無声音に大別できることを述べている。また、有声音の性質から、ピッチ周期の特性が述べられ、観測される音声波形からのピッチ周期の抽出が、いかにチャレンジングな問題であるかが述べられている。そして、本論文でピッチ抽出問題を取り上げる動機が示され、本論文の構成を説明している。

続く第２章では、従来の代表的なピッチ抽出手法の原理が説明されている。本論文では、フレーム処理からのピッチ抽出問題と取り上げるため、まずフレーム処理に用いられる窓関数が説明されている。続いて、ピッチ抽出のアプローチを、時間領域、周波数領域に分類し、それぞれの代表手法を述べている。時間領域では、自己相関関数（ACF)、平均振幅差分関数（AMDF)、重み付き自己相関関数（WACF)、YIN の方法が記述されている。周波数領域では、ケプストラム（CEP)、改良ケプストラム（MCEP)、窓なし ACF − CEP 法を取り上げている。そして、本論文で提案する２つの手法の１つは、ACF の発展手法であり、もう１つは、CEP および MCEP の発展手法であることが述べられている。

ここから続く２つの章において、本論文で提案する２つのピッチ抽出のための方式を論じている。

第３章では、べき乗スペクトルを利用するピッチ抽出手法を述べている。これは、CEP および MCEP の発展手法である。CEP は、無雑音環境下においては非常に高精度な抽出結果を与えることできる。しかしながら、雑音環境下においては、雑音の影響を大きく受けてしまう。これは、CEP に利用される対数スペクトルに起因する。対数スペクトルは、スペクトルの大きさの大小を圧縮し、ダイナミックレンジを狭めることができる。これにより、ピッチ誤り抽出を引き起こす、スペクトル中のホルマント特性の影響を抑えることができ、音声スペクトル中の調波構造を保つことができる。このため、無雑音環境下では CEP はたいへんに高精度に働く。しかしながら、雑音が混入すると、対数スペクトルの処理は、雑音特性が音声の調波構造をかき乱すように働く。したがって、雑音環境下では、CEP のピッチ抽出精度は大きく劣化する。そこで、スペクトルのダイナミックレンジは狭めつつ、雑音の影響も押さえるスペクトルの計算手法として、４分の１乗根スペクトルを見出している。他のべき乗根スペクトルとの比較から、４分の１乗根が最適であり、広帯域性を有する雑音、および狭帯域性を有する雑音など、８種類の雑音環境下で、実験的に検証を行っている。結果として、広帯域性を有する雑音環境下では、４分の１乗根スペクトル計算の後に、リフタリングとクリッピングを施すことで、高精度な抽出結果をもらし、狭帯域性を有する雑音環境下では、４分の１乗根スペクトル計算の後に、リフタリングとクリッピングを施し、さらに４乗のべき乗計算を付加することで、高精度な抽出結果をもたらすことが明らかにされている。それらの抽出結果は、ごく最近発表された従来手法よりも優れたものである。

第４章では、ACF の発展手法として、異なる帯域幅を有する帯域通過型フィルタを複数利用し、かつ、スペクトルの累積処理を施す手法を述べている。ACF は、比較的雑音の影響を受けにくい特性を有する。一方で、ホルマント特性は ACF のピッチ抽出結果に大きく影響を及ぼすことから、従来様々な改良がなされてきたが、複数種類の雑音の影響を考慮したものはほとんどない。そこで、人間の有する基本周波数の範囲を考慮して、50Hz から 900Hz までの範囲で異なる帯域幅を有する７つの帯域通過型フィルタを設定している。そして、それぞれのフィルタ出力から得られる信号のスペクトルを累積することで、多くの雑音特性の影響を抑えることを考えている。しかしながら、単なる帯域通過型フィルタの複数利用とスペクトルの累積処理の組み合わせでは、限られた雑音特性をカバーすることになるため、さらに、時間分割処理を加え、１つの帯域通過型フィルタを施す方法、および時間分割処理と帯域通過型フィルタの複数利用、スペクトルの累積処理を組み合わせる、トータル３つの手法を導出し、それぞれの特性解析を行っている。解析結果および実験結果より、強い時変性を有する雑音環境では、時間分割処理を加えることが有効であり、雑音の種類に応じて、３つの手法を使い分けることが最善の利用方法である結論を導き出している。

第５章は、本論文のまとめである。提案する２つのピッチ抽出手法の特性を整理し、実際の環境においてそれぞれどのように利用可能であるかの示唆を与えている。

本論文では、以上に述べたように、雑音環境下における音声信号のピッチ抽出手法を提案し、その実験的評価を行っている。本論文の結果は、２編のレフリー付学術雑誌に採択され、また国際学会での発表で公表されている。

以上のように、本論文は新しいピッチ抽出手法の提案と、その有効性を検証した論文であり、博士（学術）の学位にふさわしい内容を持つものと判断し、審査委員会として「合格」の判定を行った。