

|              |                                                                                   |
|--------------|-----------------------------------------------------------------------------------|
| Title        | 振幅変調特性に着目した雑音残響に頑健な基本周波数推定法                                                       |
| Author(s)    | 三輪, 賢一郎                                                                           |
| Citation     |                                                                                   |
| Issue Date   | 2018-12                                                                           |
| Type         | Thesis or Dissertation                                                            |
| Text version | ETD                                                                               |
| URL          | <a href="http://hdl.handle.net/10119/15758">http://hdl.handle.net/10119/15758</a> |
| Rights       |                                                                                   |
| Description  | Supervisor: 鷗木 祐史, 情報科学研究科, 博士                                                    |

|               |                             |
|---------------|-----------------------------|
| 氏 名           | 三輪 賢一郎                      |
| 学 位 の 種 類     | 博士(情報科学)                    |
| 学 位 記 番 号     | 博情第 406 号                   |
| 学 位 授 与 年 月 日 | 平成 30 年 12 月 21 日           |
| 論 文 題 目       | 振幅変調特性に着目した雑音残響に頑健な基本周波数推定法 |
| 論 文 審 査 委 員   | 主査 鵜木 祐史 北陸先端科学技術大学院大学 教授   |
|               | 赤木 正人 同 教授                  |
|               | 党 建武 同 教授                   |
|               | 長谷川 忍 同 准教授                 |
|               | 森勢 将雅 山梨大学 准教授              |

## 論文の内容の要旨

Fundamental frequency (F0) has a very important role in various speech/music information processing scenes including the speech communications in which a machine intervenes, and depending on whether or not the fundamental frequency can be accurately grasped, the processing accuracy of these various sound/music information processing is influenced.

Estimation of the fundamental frequency, F0, of a target sound signal has been widely studied in the literature on speech signal processing, and many methods have been proposed over the last half century. The conventional methods can be divided into processing in the time domain or frequency domain, or both. Most of these methods use periodic features in the time domain (e.g., autocorrelation) or harmonic features in the frequency domain (e.g., comb filtering). The others focused the information of sound source by the vocal tract filter effect based on the source-filter model. Some methods that precisely estimate the F0 of target noiseless speech have been established. However, they cannot precisely estimate F0 under noisy conditions, reverberant conditions, or both conditions. Other methods have been proposed that robustly estimate the F0 of observed speech under noisy conditions. On the other hand, a few methods to estimate the F0 of a target signal robustly and accurately have been studied in reverberant environments. However, it has not yet been clarified whether all these methods can precisely estimate F0 in very noisy reverberant environments. In other words, there is no method that can accurately estimate F0 in an environment with the both of noise and reverberation.

Therefore, in this research, we got a hint from the behavior of the pitch perception for AM signal, and focused on the modulation component of the sound signal. We proposed a novel F0 estimation method robust to noise and reverberation by using an unprecedented approach to simulate human pitch perception by amplitude modulation demodulation technology. Focusing on the modulation component of the sound signal has several merits. Considering the amplitude modulation signal from the viewpoint of the modulation transfer function, the influence of disturbance such as noise and reverberation can all be put in a simple diagram of a decrease in modulation index. Therefore, it is possible to grasp the influence of the disturbance by the

observing of the modulation index without going through complicated signal processing such as examining the autocorrelation of the waveform or examining the frequency spectrum information. Although the amplitude of the modulation component decreases due to disturbance, the modulation component period is retained, so the modulation component itself of the sound signal is a signal component robust to disturbance. In addition, it is also possible to add the waveform restoration function with modulation index as a parameter, if necessary.

Simulation results have shown that the proposed method can reliably cope with time-variant signals and has robustness for noise and reverberation. Furthermore, we consider the F0 estimation of musical instrument sound as an example of an application that matches the characteristics of the proposed method, and shown that the proposed method can be applied sufficiently to a signal such as instrument sound.

Throughout this research, we have proposed F0 estimation method based on pitch perception for AM signal concerning F0 estimation in noisy reverberant environment, we show that F0 estimation method robust to noisy reverberation can be constructed by using amplitude modulation demodulation technique.

There should certainly be a field where the proposed method can contribute. For example, by implementing the proposed method as a part of other F0 estimation methods, we can expect roles that complement other F0 estimation methods robustly. Thus, the essence of the proposed method will contribute greatly to speech/music information processing technology in the future.

**Keywords:** F0 estimation, pitch perception, amplitude modulation/demodulation, noisy reverberant environments

## 論文審査の結果の要旨

音声や楽器音といった周期信号の基本周波数（F0）は、ピッチ知覚や音高推定といった様々な音信号処理において重要な特徴として利用されている。F0推定の研究は既に半世紀前から行われており、様々な方法が提案されている。代表的なものとしては、時間領域あるいは周波数領域にて相関処理による周期性・調波性を利用したもの、ケプストラムや線形予測を利用した声道特性の影響を考慮したものなどがあげられる。その推定精度を向上させるために、STRAIGHTやYIN、最近ではSWIPE'といった方法も提案されている。これらの推定法の正確性は非常に高く、実用的な音信号処理で要求される精度に十分到達している状況にある。しかし、背景雑音や残響といった外乱に対する頑健性に関しては未だ問題が残る、決定的な解決法がないところである。

本論文では、ヒトの振幅変調音のピッチ知覚の研究からヒントを得て、音信号の振幅変調成分に着目した上で振幅変調の復調技術を用いることで、高い耐雑音性・耐残響性を有するF0推定法を提案した。ここでは、音信号の変調成分に着目することで、次のようなメリットを活用した。まず、任意の三つの調波成分を振幅復調すること（2番目の調波をキャリア、1番目と3番目調波を側波成分として復調）で基本周波数を変調信号として取り出せ

ることである。これは周期信号のどの調波成分を利用しても同様に **F0** を推定できることから、推定の耐性を大幅に向上させることができる。次に、振幅変調信号を変調伝達関数 (Modulation Transfer Function : MTF) の観点から考えると、雑音や残響などの外乱による影響は、全て変調度 (Modulation Index) の低下という単純な図式に落とし込めることである。これは、時間領域や周波数領域における信号の相関処理を利用せずとも、変調度の観測から外乱の影響を把握することができる点で優れる。最後に、MTF の概念に基づき、雑音や残響による変調成分の減少分を見積もって変調成分を回復させることで、頑健な **F0** 推定を可能とする。提案法は、原理的にステップ状に変動する周期信号に対して頑健で正確な **F0** 推定を可能とするため、AM 音とみなせる信号や楽器音、単母音といった周期信号に関して適用可能であり、既存の優れた方法と活用することで、正確性と頑健性が高い **F0** 推定を実現させることも可能である。

以上、本論文は、ヒトの AM 音のピッチ知覚の観点から、まったく新しいアプローチから雑音・残響に耐性の高い **F0** 推定法を提案したものであり、学術的に貢献するところが非常に大きい。よって博士 (情報科学) の学位論文として十分価値あるものと認めた。