

## 区間分割した音声信号の周波数処理による特定人物への音声変換

松田 実紗<sup>†</sup> 西出 俊<sup>‡</sup> 任 福継<sup>‡</sup><sup>†</sup>徳島大学 工学部 <sup>‡</sup>徳島大学 大学院ソシオテクノサイエンス研究部

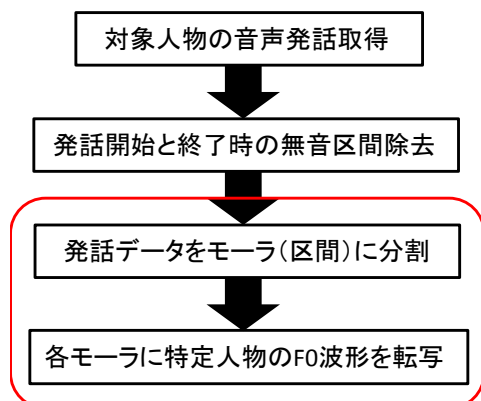
## 1. はじめに

本研究では音声信号の周波数を操作することで発話音声特定の人物の音声信号に変換することを目標とする。特に音声信号に含まれる個人性を転写することで特定人物の音声を模倣することを目指している。

Fujisakiは音声信号に含まれる情報を、言語情報、パラ言語情報、非言語情報)の三つに分類した[1]。人物の音声特徴の中でパラ言語情報の一つであるF0波形は多くの情報を持っており、音声信号に個人性や感情を付与する際に用いられる[2]。歌声や音声の分析においてもF0の情報はよく用いられる[3]。本研究でも同様にF0波形を転写することで特定人物のイントネーションを他の音声信号に付与する。本稿ではこれまで構築した基礎システムと評価実験の結果について報告する。

## 2. 提案手法の概要

本研究で目標とするシステムの流れを図1に示す。対象人物とは本手法によって特定人物のF0



本稿では基礎実験のため、モーラに分割せず、発話区間全てのF0波形を転写した

図1. 目標とするシステムの流れ

波形を転写する対象を表す。目標システムではまず、対象人物の音声波形を取得し、開始と終了時の無音区間を除去する。音声波形をモーラごとに区間分割し、各区間に対し、特定人物のF0波形を転写する。

本稿では目標とするシステムの実現可能性を検証するための基礎実験として、下記の制約を入れてシステムを構築した。

1. 音声波形をモーラで分割せず、音声全体のF0波形を転写する。(図1の赤枠部分)
2. 対象人物と同じ発話内容の特定人物の音声発話も取得する。

1.について、本研究では基礎実験を行うため、区間分割する前段階として、F0変換による特定人物の音声模倣の有効性について検証した。2.について、本システムではF0以外の要因を極力減らすため、対象人物と特定人物の同じ発話内容間でのF0転写を行った。

本システムを構築するため、フリーの音声分析ソフトであるPraat [4]を用いた。Praatはアムステルダム大学で開発されたオープンソフトなソフトウェアであり、音声分析ソフトとして、様々な研究で活用されている。まず、取得した音声信号の先端と終端のパワーが微少な区間を除去した。次に、特定人物の音声信号をPraatで読み込み、F0波形を求めた。計算したF0波形を対象人物のF0波形に上書きすることで、F0波形の変換を行った。

1. こんにちは
2. 今日はとても寒いですね。
3. 今日は何をしに来たのですか？
4. それは楽しそうですね！
5. ところで昨日のテストはどうでしたか？
6. すごいですね！
7. 私はよく分かりませんでした。
8. 今度勉強を教えてくださいませんか？
9. ありがとうございます！
10. また明日、さようなら。

図2. 読み上げ文

### 3. 評価実験

F0波形の転写による特定人物の音声模倣の有効性を検証するため、5人の被験者の音声を録音し、それぞれのF0波形を他の音声に転写する実験を行った。まず、図2に示す10文の読み上げを5人の被験者に読んでもらい、録音した。それぞれの読み上げ音声について無音区間除去し、F0波形を文ごとに算出した。その後、各被験者の音声のF0波形に、他の4人のF0波形を上書きしたものを作成した。

本実験の目的は F0 波形の上書きによる音声模倣の性能を評価することである。そこで、評価実験は以下の手順で行った。

1. 被験者の5人とは異なる評価者2人に対し、ある特定の被験者の1. から3. までの読み上げ文の録音音声を聞かせる。
2. 4. から10. について、特定の被験者に加え、F0 波形を上書きした他の4人の音声も聞かせる。
3. 4. から10. の各文に対し、どの音声が特定の被験者のものであるか紙に記入する。

本実験の結果を表1に示す。表1において、特定の被験者をA、他の4人の被験者をB、C、D、Eとしている。4. から10. の文について、二人の評価者が選択した音声を示している。

本実験の結果より、F0 波形の上書きのみでもある程度は特定人物の音声を模倣することが可能であることが示された。さらに、モーラによって区間分割し、区間ごとに F0 波形を処理することでより高性能な音声模倣が実現できる可能性を示唆する結果が得られた。

表1 評価実験の結果

	評価者1	評価者2
文4.	A	A
文5.	A	A
文6.	A	A
文7.	A	A
文8.	A	B
文9.	A	B
文10.	A	A

### 4. 考察と今後の課題

本実験を行う際、音声データの収集効率を上げるため、各音声データは別々の環境で録音されたものを用いた。そのため、実験においては

ノイズなどの環境音によって特定人物の音声を判別しやすくなったという意見もあった。このような意見から同一の環境で録音した音声を用いた場合はさらに高性能な音声模倣が可能になると考えられる。

本稿は図1のシステムを構築するための基礎システムについて報告したものであり、全体システムを構築する上で、下記に示すように様々な課題を解決することが必要である。

1. 分割した区間の関連性を考慮した F0 処理
2. F0 波形以外の音声情報の転写

1.について、これまでは音声全体の F0 波形を用いたが、区間に分割する際は、区間同士のつながりを考慮した上で F0 を処理する必要がある。2.については音声情報に含まれる他のパラ言語情報も転写することでより高性能な音声模倣が可能になると考えられる。

### 5. おわりに

本稿では特定人物の音声模倣を実現するため、音声信号のF0波形を他の音声信号に転写することによる模倣性能を評価した。音声信号をモーラに区間分割してF0波形を転写するシステムの基礎実験として構築した。評価実験の結果、特定人物のF0波形を転写した音声を特定人物の音声と混同することがあることが示された。今後はモーラ単位で区間に分割し、各区間のF0を処理することでより精度の高い音声模倣を実現することを目指す。

### 謝辞

本研究は JSPS 科研費 15H01712 の助成を受けたものである。

### 参考文献

- [1] H. Fujisaki, "Prosody, information, and modeling - With emphasis on tonal features of speech," Proc. Speech Prosody, pp. 1-10, 2004.
- [2] 劉丹, 堂元健太郎, 井上祐輔, 宇津呂武仁: "単語アクセント型ごとのF0波形と時間長を利用した単語音声の感情変換," 情報処理学会研究報告, Vol. 2014-MUS-103, No. 33, pp. 1-6, 2014.
- [3] 中野倫靖, 後藤真孝, : "歌声・音声分析合成のためのF0適応多重フレーム統合分析に基づくスペクトル包絡と群遅延の推定法," 情報処理学会音楽情報科学研究会 研究報告, Vol. 2012-MUS-96, No. 7, pp. 1-9, 2012.
- [4] Praat Software:  
<http://www.fon.hum.uva.nl/praat/>