

各 位

2021年8月12日
株式会社インプレス

「音声合成」の基礎から実装までを解説した、中級者以上向けの実践的な技術書
『Pythonで学ぶ音声合成』を8月12日（木）に発売

インプレスグループでIT関連メディア事業を展開する株式会社インプレス（本社：東京都千代田区、代表取締役社長：小川 亨）は、音声合成の基礎から実装までを解説した書籍『Pythonで学ぶ音声合成』を2021年8月12日（木）に発売いたします。



■音声合成の基礎から最新手法までカバー

「音声合成」とは、人間の音声を手動的に作り出す音声情報処理の一分野です。近年、深層学習（deep learning）の発展に伴い、機械学習による音声合成の技術は飛躍的に進歩しています。また、PyTorch、TensorFlowをはじめとした、深層学習のためのオープンソースソフトウェアとオープンソースコミュニティの発展により、研究者や技術者が公開しているソフトウェアやソースコードを無償で誰もが手に入れることができるようになったことで、専門家でなくても取り組みやすくなってきているといえます。

しかしその一方で、音声合成の最新手法について書かれた日本語の書籍は限られており（執筆当時）、初学者にはハードルが高い、という声もありました。本書はその問題を解決したいとの思いで誕生しました。全章を通して、従来の統計的音声合成システムの基礎から深層学習による近年の音声合成の発展まで詳説しています。

■日本語の音声合成システムの作り方を丁寧に解説

また、Pythonを使って深層学習に基づく日本語の音声合成システムを作る方法も丁寧に解説しています。ソースコードはすべてGitHub (<https://github.com/r9y9/ttslearn>) で公開しているので、初学者も実際に手を動かしながら学ぶことができます。

本書は、2020年8月24日刊行『Pythonで学ぶ音源分離』（戸上真人著）、2021年5月20日刊行『Pythonで学ぶ音声認識』（高島遼一著）に続く「機械学習実践シリーズ」の3冊目です。

■本書の章立て

序章

第1章 音声合成とは？

第2章 音声の情報と物理

第3章 統計的音声合成

第4章 Pythonによる音声信号処理

第5章 深層学習に基づく統計的パラメトリック音声合成

第6章 日本語 DNN 音声合成システムの実装

第7章 WaveNet:深層学習に基づく音声波形の生成モデル

第8章 日本語 WaveNet 音声合成システムの実装

第9章 Tacotron 2:一貫学習を狙った音声合成

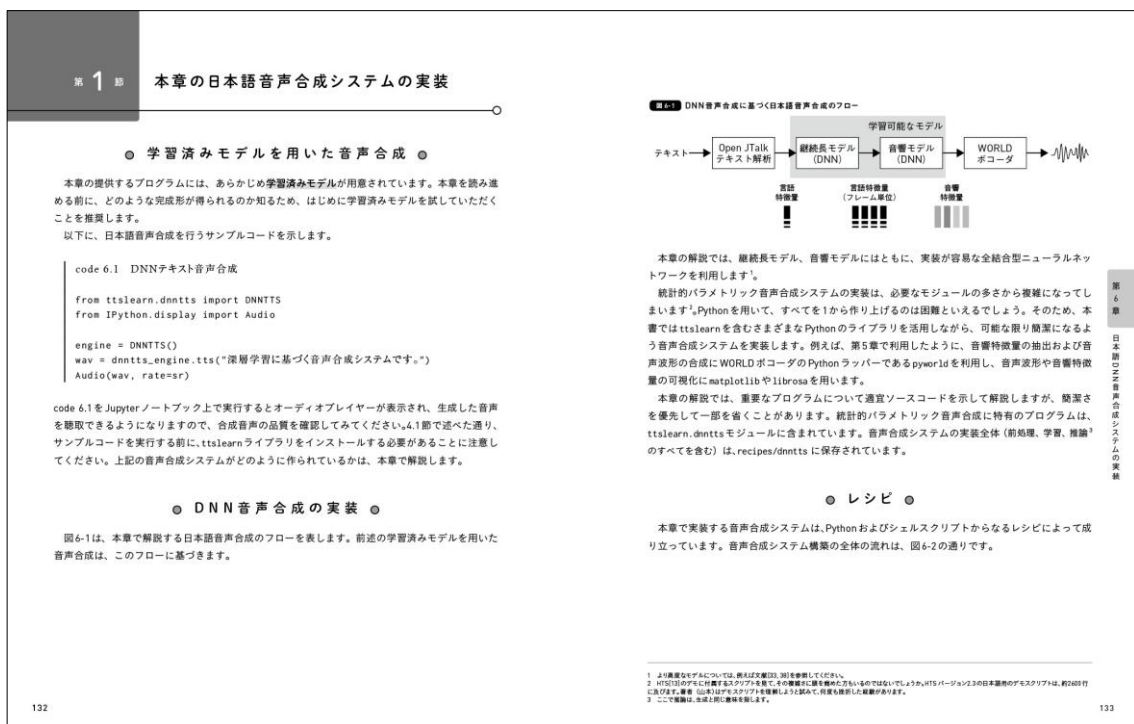
第10章 日本語 Tacotron 2に基づく音声合成システムの実装

第11章 音声合成システムを新たに作るときに

〈本書はこんな人におすすめです〉

- ・ 音声処理のエンジニア、研究者
- ・ 理系の大学生
- ・ 機械学習や音声処理に興味のある人

〈紙面イメージ〉



第6章では、深層学習に基づく統計的パラメトリック音声合成の枠組みに則り、日本語音声合成システムを実装する方法について解説しています

● 学習済みモデルを用いた音声合成 ●

本章の提供するプログラムには、あらかじめ学習済みモデルが用意されています。本章を読み進める前に、どのような完成形が得られるのかわかるため、はじめに学習済みモデルを試していただくことを推奨します。
以下に、Tacotron 2 を用いて日本語音声合成を行うサンプルコードを示します。

```
code 10.1 Tacotron 2を用いたテキスト音声合成
from ttslearn.tacotron import Tacotron2TTS
from tqdm.notebook import tqdm
from IPython.display import Audio

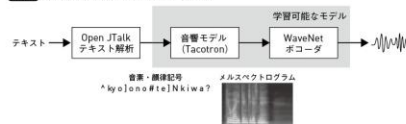
engine = Tacotron2TTS()
wav, sr = engine.tts("一貫学習にチャレンジしよう!", tqdm=tqdm)
Audio(wav, rate=sr)
```

code 10.1 を Jupyter ノートブック上で実行するとオーディオプレイヤーが表示され、ノートブック上で生成した音声を取ることができます。なお、第8章のWaveNet音声合成と同様に、Tacotron 2 は WaveNet を利用するため、音声波形の生成に時間を要します。音質よりも速度を優先する場合は、Tacotron2TTS クラスの `tts` メソッドの引数に `griffin_lim=True` と指定することで、WaveNet ボコーダの代わりに Griffin-Lim のアルゴリズム (4.7 節参照) を用いて音声波形を生成します。

● Tacotron 2 に基づく日本語音声合成の実装 ●

図 10-1 は、本章で解説する日本語音声合成システムのフローを表します。前述の学習済みモデルを用いた音声合成は、このフローに基づきます。

図 10-1 Tacotron 2 に基づく日本語音声合成のフロー



第9章では、文字列からメルスペクトログラム (メル周波数スペクトルの系列) を予測する注意機構付き sequence-to-sequence モデルに着目して Tacotron 2 を解説しました。もしかすると、第9章を読了したあとに「メルスペクトログラムは予測されたけど、音声波形が生成されていないのでは?」と疑問を持った読者の方がいるかもしれません。この箇所に対して本章では、第7章で解説した WaveNet¹ を組み合わせることで、音声合成システムを実装します。

本章では便宜上、波形生成の前段のメルスペクトログラムを予測するモデルを Tacotron² と表記し、WaveNet ボコーダを統合したモデル全体を指して Tacotron 2 と表記します。第9章では、Tacotron 2 を一貫学習に基づく音声合成として解説しましたが、本章の実装は Open JTalk によるテキスト解析を利用します。テキスト解析が必要な理由は、次節で解説します。

第6章と第8章の日本語音声合成システムの実装と同様に、本章はさまざまな Python のライブラリを活用して、可能な限り簡潔に音声合成システムを実装します。また、重要なプログラムについてそのソースコードと内容を解説しますが、そうでないプログラムについての解説を省略します。なお、すべてのソースコードは GitHub で公開されているため、併せて確認してください。Tacotron 2 に特有のプログラムは、`ttslearn.tacotron` モジュールに含まれています。音声合成システムの実装全体 (前処理、学習、推論のすべてを含む) は、`recipes/tacotron` に保存されています。

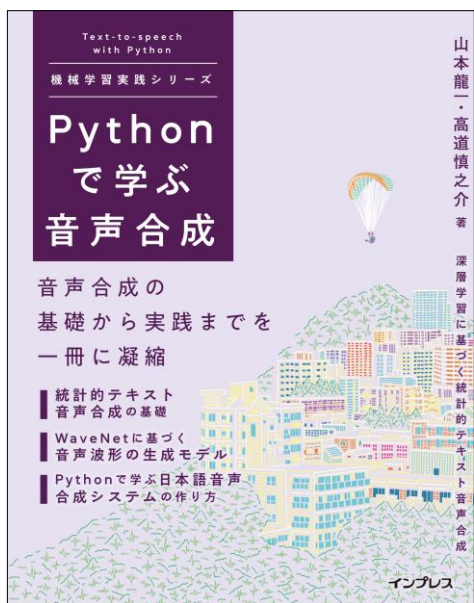
● レシピ ●

本章で実装する音声合成システムは、Python およびシェルスクリプトからなるレシピによって成り立っています。レシピとは何かの説明は、6.1 節の「レシピ」の項を参照してください。音声合成システム構築の全体の流れは、図 10-2 の通りです。

1 Tacotron 2 の論文の WaveNet ボコーダは、電子書で解説した WaveNet をそのまま適用したもので、5.1 節の原理を忠実に再現するには、改造された WaveNet を実装する必要があります。本書では簡便さを優先し、電子書の内容をそのまま適用しています。
2 文脈依存モデルおよび音質改善のために Tacotron を実装しているわけではないことは留意してください。

第 10 章では、前章までで解説した Tacotron 2 に基づき、日本語音声合成システムの実装について解説しています

■書籍の詳細



書名：Python で学ぶ音声合成 機械学習実践シリーズ

著者：山本龍一・高道慎之介

発売日：2021年8月12日(木)

ページ数：352ページ

サイズ：B5変型版

定価：3,850円(本体3,500円+税10%)

電子版価格：3,850円(本体3,500円+税10%) ※インプレス

直販価格

ISBN：978-4-295-01227-6

◇Amazonの書籍情報ページ：

<https://www.amazon.co.jp/dp/4295012270/>

◇インプレスの書籍情報ページ：

<https://book.impress.co.jp/books/1120101073/>

■著者プロフィール

山本龍一 (やまもと・りゅういち)

LINE株式会社 Voice チーム、音声処理開発者・研究者。2013年に名古屋工業大学大学院博士前期課程修了。チームラボ株式会社を経て、2018年2月にLINE株式会社に入社(現職)。2018年9月から2019年7月までNAVER Corp. Clova Voice チームにて音声研究を行う。音声合成の研究開発に従事。WaveNet や Tacotron に代表される音声合成に関するオープンソースソフトウェアを多数公開。

高道慎之介 (たかみち・しんのすけ)

東京大学大学院情報理工学系研究科助教。2011年に長岡技術科学大学を卒業。2013年・2016年それぞれに奈良先端科学技術大学院大学博士前期・後期課程を修了。2018年より東京大学助教(現職)。博士(工学)。音声合成変換、音声信号処理の研究に従事。

■「機械学習実践シリーズ」とは

本書を通して「実際に動くものが作れる」ことを目指して、特定の技術のアルゴリズムと、それを実装するためのコードを豊富に紹介するシリーズです。機械学習の基本から実装までを学ぶことができます。

以上

【株式会社インプレス】 <https://www.impress.co.jp/>

シリーズ累計 7,500 万部突破のパソコン解説書「できる」シリーズ、「デジタルカメラマガジン」等の定期雑誌、IT 関連の専門メディアとして国内最大級のアクセスを誇るデジタル総合ニュースサービス「Impress Watch シリーズ」等のコンシューマ向けメディア、「IT Leaders」、「SmartGrid ニュースレター」、「Web 担当者 Forum」等の企業向け IT 関連メディアブランドを総合的に展開、運営する事業会社です。IT 関連出版メディア事業、及びデジタルメディア&サービス事業を幅広く展開しています。

【インプレスグループ】 <https://www.impressholdings.com/>

株式会社インプレスホールディングス（本社：東京都千代田区、代表取締役：松本大輔、証券コード：東証 1 部 9479）を持株会社とするメディアグループ。「IT」「音楽」「デザイン」「山岳・自然」「モバイルサービス」「学術・理工学」「旅・鉄道」を主要テーマに専門性の高いメディア&サービスおよびソリューション事業を展開しています。さらに、コンテンツビジネスのプラットフォーム開発・運営も手がけています。

【本件に関するお問合せ先】

株式会社インプレス 広報担当：丸山

E-mail: pr-info@impress.co.jp URL: <https://www.impress.co.jp/>

※弊社はテレワーク推奨中のため電話でのお問い合わせを停止しております。メールまたは Web サイトからお問い合わせください。