

各 位

2020年8月20日  
株式会社インプレス

「音源分離」の最新手法を解説した、中級者以上向けの実践的な技術書  
『Python で学ぶ音源分離』  
を8月24日に発売

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、「音源分離」技術の基礎から実装までを解説した書籍『Pythonで学ぶ音源分離 機械学習実践シリーズ』を2020年8月24日（月）に発売いたします。



### ■身近なものにも活用される技術「音源分離」の基礎と実装を解説する一冊

近年、AIスピーカをはじめとした、人が話した音声を理解する音声認識システムがさまざまな場面で使われています。一般的に音声認識システムは、1人の声を聞き取ることを想定しており、聞きたい声以外の音が入ってきたときには、聞きたい声を正確に聞き取ることが難しくなります。「音源分離」とはこのようにさまざまな音が混ざった中から、欲しい音だけを抽出する技術です。本書では、音源分離の基礎から、Pythonを用いた実装までを詳しく解説しています。

### ■音源分離に必要な不可欠な数学の基礎知識も網羅

音源分離を理解しコードを書くためには、プログラミングに関する知識に加え、線形代数、微分・積分、確率・統計といった数学的知識も必要不可欠です。本書では、線形代数や行列・ベクトルの微分の方法、確率統計の基礎について示しています。とくに音源分離では複素数の行列・ベクトルを用いるため、複素数の計算方法について重点的に解説しています。

### ■音源分離の実装に必要な準備段階の情報も丁寧にカバー

本書では第1章から第4章までを音源分離を扱うための準備の章としています。

第1章は、音源分離とはどういった技術なのか、具体例を提示しながら解説しています。第2章は、音声処理の基本として、音声ファイルの構造と、Pythonパッケージを用いて音声ファイルを開閉する方法を示しています。第3章は、音源分離に必要な、線形代数や行列、ベクトルの微分の方法、確率統計の基礎について示しています。第4章は、第3章の数学的知識をベースに、最適化に関する手法を示しています。

第5章からは音源分離の実装について解説します。第5章では、Pyroomacousticsというライブラリを用いて仮想的に設定した部屋で収録した音声をシミュレートする方法を紹介しています。第6章では、ビームフォーミングという手法を解説しています。第7章は、空間モデルのパラメータを、音源方向の推定結果を利用して求める方法を示しています。第8章は、ブラインド音源分離法について紹介しています。第9章は、残響除去の方法を示しています。第10章は、残響除去と音源分離を統合的に実行する方法を解説しています。第11章は、PyroomacousticsやNARA-WPEなどのツールや参考文献を紹介しています。

〈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
- Pythonで音源分離を実装したい人

## ■紙面イメージ

第1章 不要な音を除去する音源分離技術

近年、人が話した音声を理解する音声認識システムがいろいろな場面で使われています。例えば、AIスピーカー（図1-1）が、自宅に使われるようになってきていますし、音声で行き先を設定できるカーナビゲーションシステムの普及も進んでいます。

図1-1 音声で操作できるAIスピーカー

かしまりました  
●●の音楽が聴きたい

一般的に音声認識システムは、1人の人の声を聞き取れることを想定しており、聞きたい人の声以外の音が入ってくると、どうしても聞きたい人の声を正確に聞き取ることが難しくなってしまいます。例えば、複数の人が参加している会議シーンで、音声の議事録を音声認識する場合、白熱する会議ほど複数の人の声が重なることが多く、音声認識が困難となります（図1-2）。

図1-2 複数の人が参加する会議シーン

したがって、そのような不要な音の影響をいかに減らすかが、認識率向上の一つのポイントとな

ります。人の音声を機械が聞き取る音声認識システムだけでなく、例えば、人の音声を人が聞き取りやすくなるための補聴器でも、聞きたい人の音声だけをクリアに抽出したいというニーズが存在します。また、遠隔地と音声で会話するためのテレビ会議システムなどの遠隔通話装置でも、会議参加者以外の音声が入ると、どうしても聞きたい人の声が聞き取りにくくなるという問題が生じます。

本書で紹介する音源分離はこれようにさまざまな音が混ざった中から、欲しい音だけを抽出する技術になります。一概に欲しい音以外の音と書きましたが、このような不要な音は、以下のように大きく三つに分類できると思います。

- 1 聞きたい人以外の人が話している音声（以降、干渉音とします）
- 2 エアコンの音やコンピュータのファンの音のような雑音（以降、背景雑音とします）
- 3 音が壁や床や天井などで跳ね返る（反響する）ことにより生じる音（以降、残響音とします。音楽などのアプリケーションだと、残響音は聞きたい音の一部ということになると思いますが、音声認識装置など人の話している内容をクリアに聞くことが求められるアプリケーションでは、一般的には、残響音は聞き取りやすさを阻害する不要な音になります）

音源分離では、このような不要な音と欲しい音が混ざった音（混合音）がマイクホンで観測されて記憶されていることを前提とします。コンピュータ上でのファイル形式としては、例えば、音楽データなどでよく使われるようなwav形式のものになります。このような混合音をプログラムの入力として取り取り、それをプログラム内で処理することにより、不要音が除去されたきれいな音声を出力します（図1-3）。

図1-3 音源分離のプログラム構成

混合音 → 入力: wav形式など → 音源分離 (プログラム) → 出力: wav形式など → きれいな音

本書は、このような不要音を除去するための音源分離技術について理解し、音源分離のプログラムをPythonを用いて実装できるところまで、読者の方がとり着くことを目的としています。本書の残りのパートでは、簡単な数式のみを用いて、できるだけ直感的に、音源分離の基本理をお伝えしたいと思います。

「音源分離とは何か？」から丁寧に解説しています

さて、時間周波数領域の音声波形を何もしないで時間領域に戻すのも芸がないので、加工してから時間領域に戻すことにしましょう。

- 1 特定の周波数を消す
- 2 背景の雑音を消す (スペクトルサブトラクション、ウィナーフィルタ)

◎ 特定の周波数を消す ◎

短時間フーリエ変換後の信号  $x(i, k)$  について、思い切って特定の周波数よりも高い周波数成分を全て消してみましょ。そして、これを逆フーリエ変換し、その音を聞いてみます (code2.13)。

```
code2.13 特定の周波数を消した音を再生する

#特定の周波数成分を消す(100番目の周波数よりも高い周波数成分を全て消す)
stft_data[100:,:]=0

#時間領域の波形に戻す
t,data_post=sp.istft(stft_data,fs=wav.getframerate(),window="hann",np
erseg=512,noverlap=256)

#2バイトのデータに変換
data_post=data_post.astype(np.int16)

#dataを再生する
sd.play(data_post,wav.getframerate())

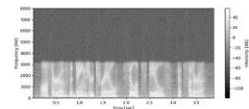
print("再生開始")

#再生が終わるまで待つ
status = sd.wait()
```

```
#wavファイルを開く
wav.close()
```

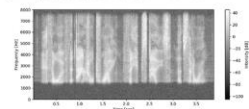
stft\_data[100:,:]=0でNumPyのndarrayのスライスの機能を使い、100番目の周波数成分より上の成分について全時間のデータに0を代入しています。こもったような音になっていませんか？この音のスペクトログラムを図2-11に示します。

図2-11 高周波数成分をカットした音のスペクトログラム



おおむね、3,000ヘルツ以上の成分が0になっている様子がわかるかと思いますが、100の部分のいろいろと変化させると、消す周波数を変えることが可能です。次は、stft\_data[50,:]=0として、低い周波数を消してみましょ。今度は、軽い乾いた音になっているかと思えます。また、ボリュームも落ちているように感じられるでしょう。この音のスペクトログラムを図2-12に示します。

図2-12 低周波数成分をカットした音のスペクトログラム



NumPy/SciPyを用いた方法や、スペクトログラムの見方も詳しく解説しています

■書籍の詳細



書名：Pythonで学ぶ音源分離 機械学習実践シリーズ  
 著者：戸上真人  
 発売日：2020年8月24日（月）  
 ページ数：352ページ  
 サイズ：B5変型判  
 定価：本体3,500円＋税  
 電子版価格：本体3,500円＋税 ※インプレス直販価格  
 ISBN：978-4-295-00984-9  
 ◇Amazonの書籍情報ページ：  
<https://www.amazon.co.jp/dp/4295009849/>  
 ◇インプレスの書籍情報ページ：  
<https://book.impress.co.jp/books/1119101154>

■著者プロフィール

戸上真人 (とがみ・まさひと)

2017年1月～2018年5月、スタンフォード大学のStanford Data Science Initiative (SDSI) Visiting Scholarとして在籍。2018年6月にLINE株式会社入社。現在、同社のAI研究組織であるData LabsのSpeech TeamのManager兼Principal Researcher。(一社)人工知能学会理事。16年以上に渡り、音声処理の研究開発に従事。東京大学工学系研究科より2011年に博士(工学)授与。登録特許20件以上、査読あり論文9件、査読あり国際会議での採択論文数60件以上。IEEE Senior Member。

■「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」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

TEL：03-6837-5034 E-mail：[pr-info@impress.co.jp](mailto:pr-info@impress.co.jp) URL：<https://www.impress.co.jp/>