音声処理と認識

第1回：ガイダンス／音声技術の応用紹介

担当：小野原彩香
3年後期｜CBT・動画回

📌 この回は**CBT・動画回**です。自分のペースで視聴し、最後のCBTテスト（5問）を受けてください。目安：40〜50分

この回で学ぶこと

1. この授業の全体構成・評価方法・受講ルールを**説明できる**
2. 音声技術がどのような製品・サービスに使われているかを**具体的に挙げられる**
3. 音声処理の大まかなパイプライン（流れ）を**ステップ順に説明できる**
4. この授業で「学ぶこと」と「学ばないこと」を**区別できる**

> 🧩 **神経多様性への配慮**
> この授業の資料は、視覚的説明（図の文字説明）とテキスト説明を常に両方用意しています。
> 動画は字幕付きで提供します。**自分に合ったペースで取り組んでください。**

3つの大きな問い

このコースを貫く軸

① 人間はどのように声を出し、聴き取るのか？

② コンピュータはどのように音声を「理解」するのか？

③ 私たちはその技術を使って何ができるのか？

> 人間の声　→　デジタル処理　→　応用・設計

授業の全体マップ（14回）

回	形式	テーマ
1	**CBT**	ガイダンス・音声技術の応用
2	同期	発話・聴覚メカニズム
3	**CBT**	音声信号のデジタル化
4	同期	スペクトログラム・周波数分析
5	**CBT**	音響特徴量（MFCC）の原理
6	同期	ピッチ・エネルギー・前処理演習
7	**CBT**	HMM・GMM・自己教師あり学習
8〜14	交互	ASR・UX・NLU・認知科学・倫理

評価方法

評価項目	割合	内容
小課題・演習	**30%**	動画回のCBTテスト（各回5問・Google Forms・何度でも回答可・最高点を採用）＋同期回の演習課題
期末プロジェクト/レポート＋発表	**70%**	Google Drive に提出・複数の形式から選択可能・第14回に最終発表

- CBTテストの締め切り：**各回の翌週末 23:59**
- 期末評価の形式：プレゼン録画・レポート・スライドメモ等から選択可

CBT動画回の進め方

```
[1] 動画を視聴する（YouTube 限定公開）
        ↓
[2] 補足資料を読む（Google Drive）
        ↓
[3] CBTテスト 5問を解く（Google Forms）
        ↓
[4] 次回の準備（やれる範囲で）
```

- 動画は **1.0〜1.5倍速** で視聴可（字幕付き）
- テストは締め切りまで **何度でも回答可能**（最高点を採用）
- 質問は **メール** で受け付けます

オンライン同期回の進め方

時間帯	内容
前半（約30分）	講義（Zoom・録画を YouTube で後日公開）
中盤（約20分）	演習・ディスカッション
後半（約10分）	まとめ・課題説明

参加の仕方は自分で選べます：

- リアルタイム参加（Zoom）
- 録画を後から視聴
- ディスカッション課題 → メールで提出でも可
- 演習 → 締め切りまでに Google Drive へ提出でも可

私たちの周りの音声技術

応用分野	代表例
🎙️ 音声アシスタント	Siri, Alexa, Google Assistant
📝 文字起こし	Whisper, Notta, Google音声入力
💬 自動字幕	YouTube字幕, Zoom ライブ字幕
🔊 音声合成（TTS）	カーナビ, VOICEVOX, スクリーンリーダー
🏥 医療・臨床	失語症支援, 認知機能低下の早期発見支援
🎮 エンタメ	VOCALOID, AIボイスチェンジャー

ケーススタディ：Alexa はどう動くか

「アレクサ、明日の天気は？」
↓
マイクで収音し、音声信号をデジタル化
↓
端末内でウェイクワード検出（"Alexa" を認識）
↓
検出後の音声をクラウドへ送信
↓
自動音声認識（ASR）：音声 → テキスト
↓
自然言語理解（NLU）：意図・エンティティを抽出
↓
バックエンド処理（天気 API へ問い合わせ）
↓
音声合成（TTS）：返答を音声に変換
↓
スピーカーから音声で返答

※ クラウドを使う代表的な構成を簡略化した例。端末内で認識まで行う構成もある。

音声認識を使うシステムの 5 ステップ

```
[1. 収音] → [2. デジタル化] → [3. 特徴抽出] → [4. 認識・理解] → [5. 応用]
  マイク       サンプリング        MFCC 等          HMM・深層学習      TTS・対話
               量子化                                                 システム
```

ステップ	詳しく学ぶ回
デジタル化	第3回
特徴抽出（MFCC）	第5回
認識（ASR）・統計モデル	第7〜9回
意図の理解（NLU）	第11回
応用・UX・社会的インパクト	第10・13・14回

💡 **この図は授業全体の地図です。迷子になったらここに戻ってきてください。**

音声認識の精度はどこまで来たか

評価指標：単語誤り率（WER）

> WER = （置換 S ＋ 削除 D ＋ 挿入 I）÷ 正解の単語数 N × 100%
> 例：正解「今日／は／いい／天気／です」（5語）で1語が置換 → 1÷5 = **20%**

同じモデルでも、評価する音声で値は変わる（Whisper large-v2・2022年）

評価データ	音声の種類	WER
LibriSpeech test-clean	朗読（雑音が少ない）	2.7%
Switchboard	電話での会話	13.8%
CallHome	家族・友人との電話会話	17.6%

- 一部の条件では人間の文字起こしに近いが、**言語・話者・録音環境・評価方法で大きく変わる**
- 比較するときは**同じ条件で評価した数値**を見る（第9回で詳しく）

技術の恩恵と課題

🟢 恩恵・可能性

- 視覚障害者・肢体不自由者へのアクセシビリティ向上
- 高齢者の情報アクセス支援
- 多言語コミュニケーションの促進

🔴 課題・リスク

- プライバシーとデータ収集（常時録音）
- **バイアス**：方言・アクセント・非標準的な声への認識精度の偏り
- **Deepfake音声**による詐欺・なりすまし
- 少数言語への対応格差

この授業で学ぶこと ✅

- 音声のデジタル化（サンプリング・量子化）
- 周波数分析（フーリエ変換の直感的理解）
- 音響特徴量の計算（**MFCC**）
- 古典的音声認識モデル（**HMM**）の基礎概念
- 深層学習を使った音声認識の概要（Transformer系）
- 現代の音声認識システム（**Whisper**等）の使い方
- PythonとGoogle Colabによる実装演習
- 音声技術のUX・社会的応用

> 🎯 「すべてを使いこなせる技術者」ではなく、**「仕組みを説明でき、適切に応用できる素養」**を身につけることが目標です。

この授業で学ばないこと ❌

- 数学的な厳密な証明（フーリエ変換の完全な導出など）
- 最新の研究論文レベルの内容
- 大規模モデルの学習（GPUクラスタの使い方）
- 音楽信号処理（楽器認識・ビート検出など）
- 自然言語処理（NLP）の深い内容

> 💡 これらは「この授業の範囲外」ですが、**基礎を理解すれば自分で学べます。**
> 参考リンクは授業ページ本文の末尾を参照してください。

受講にあたっての準備

前提とする知識：

- Pythonの基礎（ループ・関数・リスト操作）
- 高校レベルの数学（三角関数・対数の概念）

Python実習環境：

- **Google Colab** を使用（ブラウザで動く、インストール不要）
- Googleアカウントがあれば今すぐ使える → colab.research.google.com

Pythonが不安な場合：

- Paiza / Progate などの無料入門サービスを活用
- 第4回の演習前までに基礎を確認しておくと安心

質問・連絡の方法

手段	用途	回答目安
メール	授業内容の質問・個人的な相談	48時間以内
オフィスアワー	対面相談（要予約）	要相談

> 質問はメールで気軽に送ってください。
> 同じ質問が複数来た場合は、次回の動画や補足資料で共有します。

次回（第2回）の予告

第2回：発話と聴覚のしくみ【オンライン同期】

- テーマ：ヒトはどのように声を出し、聴くのか
- 内容：発話器官・調音音声学・聴覚メカニズム・音声知覚の不思議

今すぐできる事前学習（15〜20分）：

1. 📱 自分の声を録音して聴いてみる
   →「録音した声は普段の声と違う？」と感じる理由を考える
2. 🎬 「マクガーク効果」を YouTube で検索・視聴
3. 📖 「カクテルパーティ効果」とは何かを調べる

> 事前学習は提出不要です。やれた範囲で参加してください。

第1回のまとめ

1. 音声技術は **Siri・字幕・医療AI** など幅広く社会に浸透している
2. 音声処理には **「収音→デジタル化→特徴抽出→認識→応用」** の5ステップがある
3. この授業は **仕組みの理解** と **Pythonでの実装** を重視する
4. **CBT動画回** と **オンライン同期回** が交互に配置される

次のステップ

- **CBTテスト 5問** を解く（この動画の後、Google Forms で）
- 事前学習（録音体験・マクガーク効果・カクテルパーティ効果）

CBTテスト

続けてGoogle Formsのテストを受けてください

全 **5問**・4択形式
繰り返し回答可・最高点が採用されます

締め切り：**翌週末 23:59**