第14回：まとめと最終発表

音声処理と認識

小野原彩香 ／ 北星学園大学

Zoom録画を後日公開します。課題はメール/Google Driveでも提出可。

この回の目標

1. 14回の授業を通じて学んだ内容を体系的に振り返る
2. 期末プロジェクトの要件と評価基準を正確に理解する
3. 発表・報告を通じて自分の学びと考察を他者と共有する
4. 音声処理・認識の技術と社会的意義を統合的に理解する

配慮事項

この回は発表セッションが含まれます。
口頭発表が困難な場合は、以下の代替形式を認めています：

- 事前録画動画の再生
- スライド＋テキストチャットでの説明
- 書面レポートのみでの代替

不安な場合はメールで小野原まで相談してください。

授業の流れ（90分の例）

口頭発表2件を、それぞれ20分＋質疑5分で扱う場合の例です。

時間	内容
00〜10分	授業全体の振り返り
10〜20分	期末プロジェクトの提出要件の最終確認
20〜70分	20分発表＋5分質疑を2件
70〜80分	フィードバック・全体質疑
80〜90分	アンケートの案内・クロージング

口頭発表の規定は20〜25分＋質疑5分。人数・形式・必要時間に合わせ、担当者が進行・視聴方法を事前に案内します。

14回の学習マップ — 信号と特徴量

回	実際の授業テーマ
1	ガイダンス／音声技術の応用紹介
2	発話・聴覚メカニズムの基礎
3	音声信号とデジタル化（サンプリング・量子化）
4	スペクトログラムの理解と可視化演習
5	音響特徴量（MFCC）の原理とPython実装
6	ピッチ・エネルギー・ゼロ交差などの前処理

発話・聴覚の知見を参照し、音声をデジタル信号として測定・表現する。

14回の学習マップ — モデルと評価

回	実際の授業テーマ	接続する概念
7	音声と統計モデル＋自己教師あり表現学習	HMM・GMM、wav2vec 2.0等
8	大規模End-to-End音声認識	Conformer・Whisper等
9	音声認識システムと評価指標	WER・CER・SLUE

モデルが何を入力し何を予測するか、どのデータと条件で学習・評価したかを結びつける。

文字列の正確さ、意味理解、話者照合は別の課題として測定する。

14回の学習マップ — 利用・認知・社会

回	実際の授業テーマ
10	音声インタフェースとUX設計
11	自然言語処理との接続（NLUと音声）
12	音声処理と認知科学・言語学
13	応用事例と社会的インパクト
14	まとめと最終発表

利用場面、言語発達・知覚、神経多様性、音声生成・偽造検出、バイアスと規制を接続する。

TTS・声質変換は、主に第13回で扱った関連技術。第8・9回の主題はASRと評価です。

振り返りたい10のキーワード

キーワード	説明
MFCC	メル帯域のエネルギー等に対数を取り、DCTで得る特徴
HMM	隠れ状態の遷移と観測を確率的に表すモデル
CTC	対応時刻が未指定の系列で、複数の整列を合算して学習
Attention／Transformer	重みづけて集約する仕組み／それを使うモデル構造
Whisper	大量の音声・テキスト対で学習した多言語ASRモデル群
TTS／VC	テキストから合成／内容を保って音声特性を変換
GAN・拡散・フローマッチング	音声生成にも使われる方式。TTS・VCという課題とは別
話者照合（ASV）	登録された対象話者の発話かを判定する課題
バイアス	集団・利用条件による偏りと、その要因を検証
Deepfake／ASVspoof	生成・加工音声等／検出・照合の共通評価枠組み

技術の仕組みと、利用への広がり

説明の粒度	例
信号	波形サンプル、周波数、振幅
表現	スペクトログラム、MFCC、学習された特徴
モデル	HMM、CNN、RNN、Transformer
システム	音声入力から文字列や応答を得る処理全体
サービス	誰が、どの場面で何をできるようにするか

認識・生成・対話、個人・社会、性能・倫理は同じシステムを考える複数の観点です。対立する一つの軸に限りません。

音声処理と認識の「大きな問い」

技術的な問い

- どうすれば音声を正確に認識・生成できるのか？
- 少ないデータでも頑健に機能するモデルをどう作るか？

社会的な問い

- 音声AIは誰のために作られているのか？
- 誰が恩恵を受け、誰が排除されているのか？

個人的な問い

- あなたはどんな技術を、誰のために作りたいか？

> 期末プロジェクトでは、これらの問いのどれかに対するあなた自身の答えを示してください。

期末プロジェクト：テーマと提出物

テーマ	タイトル
A	音声特徴量分析レポート
B	Whisper等による音声認識の精度比較
C	音声インタフェースのUX提案書
D	音声Deepfake検出の調査報告（技術＋倫理）
E	自由テーマ（事前確認）

全形式でレポート本文を提出します。発表時のスライド、実験時のコード・データ等は説明書の条件を確認。

成績評価：**期末プロジェクト70%・小課題と演習30%**。締め切り・提出先は担当者が案内します。

テーマA：音声特徴量分析レポート

自分で収録、または適切な同意・利用条件を確認した音声を分析します。

- MFCCを課題要件として最低13次元で計算・可視化。0次係数を含むかも明記
- 比較対象を2種類以上設定し、計算方法と条件を記録
- 話者・発話・感情・環境などの違いについて、観察と原因の仮説を分けて考察
- 雑音の影響では、同じ音声の処理前後など、比較条件をそろえる

分量の目安：レポート4,000〜6,000字、発表スライド10〜15枚。コードとグラフを添える。

テーマB：音声認識の精度比較

言語・方言・話者・音声品質など、最低2条件を設定し、正解テキストと比較します。

WERの百分率表示（参照文が1語以上の場合）

$$\mathrm{WER} = \frac{S + D + I}{N} \times 100\%$$

S：単語の置換、D：削除、I：挿入、**N：参照文の単語数**。CERは文字単位。

- 正規化・分かち書き、モデルの版・設定、話者数・音声量を記録
- 挿入が多いと100%を超える。空の参照文は別の扱いを明示
- 精度差だけで、非公開の学習データの内訳や原因を確定しない

テーマC：音声インタフェースのUX提案書

- 利用者の目標・ニーズ・制約を整理。ペルソナの仮定と調査結果を区別
- 主要なユースケースを**3つ以上**記述
- 最低1シナリオの対話フロー図と誤認識時の対応
- ASR・TTS等の技術を説明し、実現可能性を検討
- アクセシビリティ、プライバシー等への配慮を明記

分量の目安：提案書5,000〜8,000字、発表スライド12〜18枚。プログラミングは任意。

テーマD：音声Deepfake検出の調査報告

- 生成方式を最低2種類説明。TTS・VCという課題の分類と区別
- 検出方法を最低2種類、表現・分類器・評価条件から比較
- 具体的な社会的事例を最低2件、発生年・通貨・確認内容と共に分析
- 法規制・指針の対象、法的な位置づけ、適用時点を確認
- 自分の見解を示し、出典を確認した参考文献を最低5件引用

分量の目安：レポート5,000〜8,000字、発表スライド12〜18枚。ASVspoofは版・タスク・指標も明記。

テーマE：自由テーマ

音声処理・音声認識に関連するテーマを自由に設定して報告する。

自由テーマの条件

1. 音声処理・音声認識の技術的内容が含まれていること
2. 自分の分析・考察・見解があること
3. テーマを事前に担当者に報告し、承認を得ること

アイデア例

- 音楽情報処理（楽器音の認識・楽曲生成）
- 吃音者向けの音声インタフェース改善提案
- 日本語の方言認識の課題と展望
- ポッドキャスト自動文字起こしの精度評価

評価ルーブリック（期末内100点）

項目	配点	評価すること
内容の正確さ	30点	正確な説明、適切な用語、出典と条件
分析・考察の深さ	30点	根拠に基づく考察、技術と社会・倫理への視点
発表／報告の明確さ	20点	論理的な構成、主張と方法の伝わりやすさ
独自性・創造性	20点	自分の分析・設計・解釈・視点

この100点を**科目全体の70%**として扱い、小課題・演習30%と合わせます。発表形式による基準の差はありません。

評価ルーブリック — 詳細（考察の深さ）

評価段階	点数	基準
優秀	27〜30点	事実の記述にとどまらず、複数の視点から深く考察している。技術と社会・倫理の両面を統合した分析がある
良好	21〜26点	考察があり、自分の見解が示されている。ただし一部に浅い部分がある
基準達成	15〜20点	考察が表面的。既存情報をまとめることが中心
基準未達	0〜14点	考察がほとんどなく、情報の羅列にとどまる

出典と独自の貢献：他者の成果を自己のものとして提出する剽窃は、説明書の規定に従い0点となります。出典・利用条件と自分の分析を区別してください。

発表形式の選択肢

どの形式でも評価基準は同一で、レポート本文は提出します。

形式	内容・時間
A：口頭発表	発表20〜25分＋質疑5分
B：事前録画	15〜25分の動画。チャットなどで質疑に参加できる
C：スライド＋テキスト	テキストでのやり取りを含め20〜25分
D：書面レポートのみ	発表を免除し、詳細なレポートで代替

形式の選択理由を説明する必要はありません。希望は1週間前までに連絡し、当日の変更も可能です。

発表セッションの進め方

口頭発表1件の内容	時間
発表	20〜25分
質疑	5分
合計	25〜30分

担当者のまとめのフィードバックは全体質疑などで行います。冒頭の90分の例は、20分発表＋5分質疑を2件扱う場合です。

- Zoomチャットでの質問も受け付ける
- 主張・方法・根拠へ具体的に質問する。建設的な批判も歓迎
- その場で返答できない点は、確認して後の改善につなげる

担当者から伝えたかったこと

音声処理・音声認識は「面白い技術」であると同時に、「社会と深く結びついた技術」です。

どんな素晴らしい技術も、それが誰かを傷つけたり、不平等を拡大したりするために使われるなら、技術者はそのことを考え続ける責任があります。

反対に、技術は誰かの声を届け、誰かのアクセシビリティを高め、誰かの生活を豊かにすることができます。

この授業がその両方を考え続けるための出発点になれば、担当者としてとても嬉しいです。

実現した機能と、利用時に検証すること

機能の例	確認する条件
音声認識	未知の話者・雑音・方言、文字列の正確さと利用上の負担
自然さを高めた音声合成	読み・韻律・話者類似性、利用目的・同意・権利
短い参照音声を使う模倣	事前学習、対象言語・音声条件、検出と本人確認
リアルタイム処理	遅延、計算資源、誤りへの対応、性能差

具体的な利用条件と測定結果を示し、説明できる範囲と限界を明確にします。

まとめ：第14回・そしてこの授業全体

- 14回を通じて「音声の物理」から「AIの倫理」まで縦断しました
- 技術の進化は止まらない。学び続ける姿勢が大切です
- **「誰のための技術か」を問い続けることが、良い技術者の第一歩です**

授業アンケート

匿名・任意のアンケートです。**成績評価には用いません。**

- 授業内容の難易度
- 動画・CBT・同期授業の組合せ
- 神経多様性への配慮
- 印象に残ったこと、今後学びたいこと
- 改善してほしい点・担当者へのメッセージ

回答先は担当者が別途案内します。

期末プロジェクトの提出

- 全テーマ・全形式：レポート本文（PDFまたはWord）
- 発表する場合：スライド（PDFまたはPowerPoint）
- プログラミングを含む場合：コード
- 実験を含む場合：データ（利用条件を守って共有。大容量はURLも可）

提出期限・提出先は担当者が別途案内します。提出方法や事情の相談は、メール等で連絡してください。

期末プロジェクト説明書の最終確認をお願いします。

ありがとうございました。