第12回：音声処理と認知科学・言語学

発達・神経多様性研究と音声技術

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。

この回の学習目標

1. 言語発達と音声知覚の主要理論（臨界期仮説・音韻発達）を説明できる
2. 認知科学的アプローチによる音声知覚研究の主要概念を理解できる
3. 神経多様性と音声処理の関係について多様な側面から説明できる
4. 人間と機械の音声認識の違いを批判的に比較・分析できる
5. 音声インタフェースが「バリア」にも「ツール」にもなりうることを議論できる

参加方法と配慮事項

- 発話によるディスカッションが難しい場合、Zoomチャット・共有ドキュメントへの書き込みを発言と同等に扱います
- カメラのオン/オフは自由です
- ディスカッションは「対話」を重視します。正解より「思ったことを共有すること」を大切にしています
- Zoomに参加できない場合は**メールで考えをまとめて送ってOK**です
- 自分の経験に関わる話題のため、気持ちの整理が必要な場合はいつでも退出してOKです

授業の流れ

時間	内容
0〜10分	前回の振り返り・アイスブレイク
10〜55分	講義（本文の6つのテーマ）
55〜85分	ディスカッション（Zoom参加またはメール提出でも可）
85〜90分	まとめ・次回予告・振り返りフォームの説明

この回の特別な文脈

神経多様性は、人々の神経・認知特性に多様性があることを指し、その多様性を尊重する視点にも結びつきます。

- 人の価値を能力差で序列化しないことと、**具体的な困難や支援の必要性を認めることは両立**する
- ディスレクシア・ADHD・ASD・聴覚処理上の困難は同一の状態ではない
- 診断名だけで個人の能力や好みを予測しない
- 当事者の視点を大切にし、本人が必要とする支援を考える

乳幼児の音声・言語発達

乳児は非母語の音の対比も弁別する場合があり、経験によって知覚が変化します。すべての言語音を区別できるという意味ではありません。

時期の目安	報告される変化
出生直後	母語の韻律などへの感受性。出生前の経験も関わる
乳児期前半	刺激や課題に応じて、音声の対比を弁別
6〜12か月頃	母語への感受性が高まり、一部の非母語対比の弁別が低下
1歳前後以降	初語が現れ、音と語の対応を学ぶ
2歳頃以降	語を組み合わせた発話が増える
幼児期〜学齢期	音節・モーラ・音素を意識する能力が発達

時期には個人差があり、**知覚の絞り込みが10〜12か月で完全に終了するわけではありません。**

出典：Werker & Tees (1984)、概観 Kuhl (2004)。詳しい条件・追加出典は講義本文を参照。

Kuhlの「知覚マグネット効果」

音韻カテゴリの**典型例（プロトタイプ）の周辺で、近い刺激どうしの違いを区別しにくくなる**という実験的現象。

- Kuhl（1991）は、母音の典型例と非典型例を用いて検討した
- 「磁石」は知覚空間の圧縮を説明する比喩
- 非母語の音が物理的に聞こえなくなることではない
- 日本語話者の英語 /l/・/r/ の弁別にも母語経験が影響するが、それだけでマグネット効果の実証にはならない

出典：Kuhl (1991)。詳しい条件・追加出典は講義本文を参照。

臨界期・感受性期と第二言語の学習

Lenneberg（1967）は、言語習得と発達段階の制約を論じました。

区別する観点	検討すること
第一言語／第二言語	初めて言語を獲得する場合と、既存の言語経験を持つ学習者
言語の側面	音声・語彙・文法で年齢との関係は同じか
年齢／経験	学習開始年齢、入力の量・質、使用頻度
学習／到達度	学べることと、母語話者と同じ成績に達すること

成人にも学習が起こることだけでは、あらゆる年齢効果を否定できません。**SLM／SLM-r**は母語と第二言語のカテゴリの関係や経験を扱います。

出典：Mayberry & Kluender (2018)。詳しい条件・追加出典は講義本文を参照。

カクテルパーティ効果と選択的注意

重なった声の中から、特定の話者に注意を向ける。課題や音響条件によって聞き取りは変わります。

モデル	説明の要点
Broadbent（1958）のフィルターモデル	位置・ピッチなどの物理的特徴に基づいて入力を選択する
Treisman（1964）の減衰モデル	選ばれなかった入力も完全には遮断されず、弱く処理される

注意していない側の重要な語にも気づく場合があることを説明します。全員が必ず自分の名前に気づくわけではありません。

注意の制御には、聴覚皮質や前頭・頭頂領域を含むネットワークが関わります。

音声分離技術と人間の知覚の比較

技術	概要
Conv-TasNet（2019）	時間領域の表現を学習して音声を分離
SepFormer（2021）	Transformerを用いた混合音声の分離

人間で検討する機能	音声技術で考えられる対応
文脈・語彙を使う	言語モデル、文脈入力、デコーダ
重なった発話から話者を選ぶ	音源分離、対象話者抽出、ビームフォーミング

機能上の比較であり、機序が同じという意味ではありません。**ビームサーチ**は文字列候補の探索で、選択的注意やビームフォーミングと区別します。

出典：Conv-TasNet。詳しい条件・追加出典は講義本文を参照。

ディスレクシアと音韻処理

単語の読みや綴りの正確さ・速さに持続的な困難を示す学習障害。現れ方は文字体系にも影響されます。

- 音韻処理の困難はよくみられるが、全員に共通する唯一の原因ではない
- 知的能力全般が低いことを意味しない
- 音韻意識、音韻的な短期記憶、語の音へのアクセスを分けて調べる

日本語の課題例：「しんかんせん」を「し・ん・か・ん・せ・ん」の**6モーラ**に分ける。モーラ・音節・音素は別の単位で、「しん」は2モーラ・1音節です。

出典：International Dyslexia Association：2025年定義。詳しい条件・追加出典は講義本文を参照。

音声技術とディスレクシア

ツールとなりうる場面

- 音声入力（ASR）でテキスト入力の負担を軽減
- 読み上げ機能と組み合わせて文書作成を支援

新たな困難が生まれる場面

- 自分の発話が正確に認識されない → 逆にストレス
- 音声入力の誤りを「読んで確認する」必要がある

聴覚処理上の困難（APD）と音声技術

聴覚情報の神経的な処理に関する困難を指す概念。純音聴力検査だけでは説明しにくい、雑音下などの聞き取りの困難がみられることがあります。

- 「聴力が正常で聞き取りにくい」だけでAPDと判定はできない。言語や注意・記憶などとの区別が必要
- 字幕が補助になる場合があるが、正確さ・表示速度・読みの負荷も確かめる
- **CART**は専門の文字通訳者によるリアルタイムの文字化サービスで、自動ASR字幕と区別する
- ASRも雑音で誤ることは、APDと同じ機序を持つ証拠ではない

出典：ASHA：Central Auditory Processing Disorder。詳しい条件・追加出典は講義本文を参照。

ASDと音声処理の多様性

個人差、刺激、課題、言語能力、会話の文脈を考慮します。

観点	研究と設計で確認すること
音への感覚反応	強い負担を感じる人も、反応の弱さや特定の音への関心を示す人もいる
音声の社会的処理	感情や態度の判断で使う手がかりや課題成績の違いを、全員へ一般化しない
コミュニケーション手段	発話・文字・身振り・AACの組合せと、本人の選択を尊重する

発話が少ないことと、理解や伝えたい内容が少ないことを同一視しません。状況による発話量の変化だけで「場面緘黙」とは呼びません。

出典：ASHA：Autism and Autism Spectrum Disorder。詳しい条件・追加出典は講義本文を参照。

AAC（拡大・代替コミュニケーション）

発話などによるコミュニケーションを補ったり代替したりする手段。発話がまったくできない人だけを対象としません。

区分	例・音声技術との関係
道具を使わない	身振り・サイン・表情
非電子的な道具	絵・写真・文字のボード。絵カード一般とPECSは別
電子的な道具	アプリ、VOCA／SGD。合成音声や録音音声を出力
アクセス方法	視線・スイッチ・選択肢の順次提示。本人が選ぶ

BCIには埋込み電極から発話の試みをデコードする研究があります。頭皮脳波から自由な思考をそのまま読むという説明ではありません。

出典：ASHA AAC、BCIの原研究は本文参照。詳しい条件・追加出典は講義本文を参照。

人間と機械を、同じ課題・条件で比較する

比較軸	確認すること
文字起こし	同じ音声、同じ正規化・分かち書きでWER／CERを測る
雑音・重なり	音響条件と対象話者の指定をそろえる
感情・意図	正解ラベルの定義と、判断の一致度を示す
多言語	対応言語数と各言語での習熟度・性能を分ける
再現性	人の注意・疲労、機械の版・乱数・設定を考慮
エネルギー	処理量・時間・測定範囲をそろえて測る

WERは誤り率で、小さいほどよい。「95%正解」と「95% WER」は別です。会話理解全体と文字起こしの能力も区別します。

音声AIの公平性：研究の条件を読む

Koeneckeほか（2020）：当時の5社の英語ASRを、研究対象の黒人話者と白人話者の発話で比較し、WERに差を報告。

- 特定のデータ・話者・当時のシステムでの結果であり、すべてのASRや話者への結論ではない
- 性別・年齢・方言・アクセントの差を調べるときも、録音条件・内容・話者ごとの分布を確認する
- 観察された差を、訓練データの偏りという一つの原因だけに自動的に帰属させない

設計の問い：誰の利用状況で誤りが多いか。どんな負担が生じ、どう検証・改善できるか。

出典：Koenecke et al. (2020)。詳しい条件・追加出典は講義本文を参照。

ディスカッション（30分）

テーマ：「多様な神経特性を持つ人々にとって音声インタフェースはバリアかツールか？」

参加方法（どれでもOK）

- Zoom で発話による発表
- Zoom チャットへの書き込み
- 共有ドキュメントへの記入
- **Zoom に参加できない場合はメールで考えをまとめて送ってOK**

ディスカッション：考えるヒント

以下は議論用の仮想事例です。利用者全員の経験を表すものではありません。

- **視覚障害のある人とスマートスピーカー**：定型操作が便利な一方、誤認識を画面以外で修正できるか
- **吃音のある人と音声アシスタント**：発話途中で入力が終了しないか。利用者が待ち時間を調整できるか
- **APDのある人と字幕AI**：文字が聞き取りを補助するか。表示速度や誤りが負担にならないか

設計者の視点：技術的要因、操作・修正の方法、使う場所や相手、本人の選好を分けて考えましょう。

まとめ：第12回

- **音声知覚の発達**：経験で変化する。知覚マグネット効果は典型例周辺の弁別を扱う
- **臨界期・感受性期**：第一／第二言語、言語の側面、年齢と経験を区別する
- **神経多様性**：個人差を尊重し、困難と必要な支援を具体的に考える
- **音声AI**：同じ人にも、状況・設計によって支援にも負担にもなる
- **比較・公平性**：同じ課題・条件で評価し、観察と原因の推定を分ける

振り返りフォーム

授業後、Google Formsまたはメールで振り返りを提出できます（任意）。

- 印象に残ったこと・気づいたこと
- 考え方が変わった点・深まった視点
- 授業の進め方や配慮についてのコメント（匿名可）

提出期限は授業で案内します。採点する場合は小課題・演習30%の範囲で扱い、匿名の授業改善コメントは採点と分けます。

第13回（動画回）：応用事例と社会的インパクト

- 生成AI音声偽造（Deepfake）の仕組みと社会的影響
- 音声AIにおけるバイアスの問題
- 検出技術の現状（ASVspoof）
- 倫理的ガイドラインと規制動向