個別化認知モデルを用いた音韻意識推定手法評価のための音声フィルタの検討

西川純平 [1] 森田純哉 [1]
[1] 静岡大学
nishikawa.jumpei.16@shizuoka.ac.jp
j-morita@inf.shizuoka.ac.jp

概要

言語が発達する過程の一部は，音声言語における音素やリズムなど音韻的側面への注意に関係する音韻意識という能力に支えられるとされる．本研究は，計算機モデルを活用した音韻意識の形成を支援するシステムの開発を目指す．ユーザの音韻意識は，複数の計算機モデルを搭載したシステムにおける，モデルへの相対的選好として推定される．本稿では，成人を対象とした実験設定でこの推定手法を評価する．システム出力（単語の発声）に音声フィルタをかけることで，未成熟な言語学習者の状態を模擬する．実験では，音声フィルタ（に模擬される学習者個人の特性）がモデルの相対的選好に影響するという仮説を検証する．結果として仮説は支持されなかったが，参加者の行動のばらつきから音声フィルタ設計に関する示唆が得られた．

1 はじめに

言語学習者（子どもや第二言語学習者）は，その言語学習の過程においてさまざまな困難に直面する．言語が発達する過程の一部は，音声言語における音素やリズムなど音韻的側面への注意に関係する音韻意識という能力に支えられるとされる．音韻意識のような直接観察できない人の内部プロセスを理解・予測するためには計算機モデルの構築が有効である．このような考えから，著者らは，計算機モデルを活用した音韻意識の形成を支援するシステムの開発に取り組んできた．

本研究では，この長期的目標の実現に向け，著者らが提案してきた計算機モデルを利用した個人の音韻意識を推定する手法を評価する．言語学習者の特性を模擬した実験条件を用意することで，成人の参加者を用いて推定手法評価実験を実施した．本稿ではその実験の結果を報告する．

2 システム

先行研究 [1] において，認知アーキテクチャ ACT-R[2] を用いた音韻意識のモデルがすでに実装されている．このモデルは，ACT-R が持つ一般的な記憶検索のメカニズムを音韻意識と対応づけることで，未熟な音韻意識に対応づけられるしりとり中の誤りを表現した．具体的には，モデルが持つモーラ[1)]の記憶には類似度が設定される．「似た音を取り違える」ことによるしりとりの誤りは，モーラをキューとした単語の検索に際した，モーラ間で設定された類似度の影響を受けた誤検索として表現できる．さらに，このモデルは ACT-R のパラメータを調整することでバリエーションを持つ．たとえば，モーラ間の類似度計算方法や，類似度の影響の大きさに対応する係数 P の値が操作されている．このパラメータ調整により，一部のモデルは，子音脱落のような子どもに見られる特定の誤りに対応づけられている．

図 1 は音韻意識形成支援システムの概観である．このシステムには，先行研究に基づいた複数の音韻意識モデルのバリエーションが搭載される．システムは，単語選択制のしりとり（モデルの提示する選択肢から適切な単語を選び取ること）を提供する．ユーザの音韻意識が，システムとの単語選択制しりとりを通して推定されると想定する．たとえば，子音脱落のような音韻処理の困難を抱える学習者は，自身と同様の傾向を有するモデルの誤りに気づかず，そのモデルへの相対的選好が生じることが考えられる．

1) 音の単位のひとつ．継続時間によって定義される [3]．日本語はこれを単位として処理される．

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).

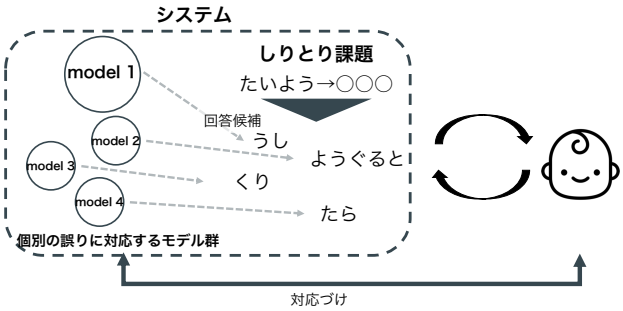

図 1: システムのイメージ．ユーザが単語選択制しりとりをプレイするなかで音韻意識が推定される．単語は例．

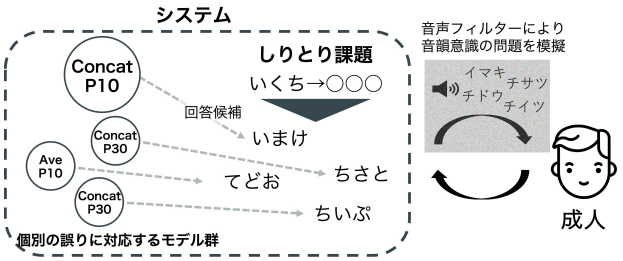

図 2: 実験コンセプト．無意味単語による単語選択制しりとり．システムの出力にフィルタをかけることで，個人の音韻意識の特性を模擬する．音声フィルタの例は Concat ベースフィルタを適用したもの．

3 実験デザイン

図 2 に実験のコンセプトを示す．この実験において，モデルのパラメータはモーラ類似度の計算方法で 2 水準 (Concat vs. Ave) [2] と類似度の影響度合 P で 2 水準 (10 vs. 30) で操作され，計 4 体のモデルがシステムに搭載された．

あるユーザの音韻意識の状態を推定することは，そのユーザが持つ音声に対する聴覚的なフィルタのようなもの同定する課題として表現できる．この課題設定の妥当性を検討するにあたり，実ユーザ（つまり子どもや第二言語学習者）を対象とした実験ではユーザの持つフィルタを統制できないという困難がある．このことを踏まえて，本研究では，実験者が設定した音声フィルタの同定をテストする．つまり，ユーザの持つフィルタとしてシステムの出力音声にかかる音声フィルタを設定し，言語学習者（子どもや第二言語学習者）ではなく，日本語を母語とする成人を参加者とする．

モデルは，100 種類のモーラを 3 つ結合することで生成された無意味単語 2,000 個を語彙として有すると仮定する[3]．単語選択制しりとりにおいてモデルが提案する回答候補単語は，音声ファイルを再生することでシステムから出力される．モデルの語彙 2,000 単語についてテキスト読み上げサービス Amazon Polly[4] を利用して音声ファイル（mp3）を用意した．音声ファイルの作成で指定する SSML (Speech Synthesis Markup Language) では，alphabet タグに”x-amazon-pron-kana”を，ph タグにカタカナ（たとえば単語「てどお」に対して発音「チドウ」）を指定した．

システムの出力（モデルの回答候補単語の読み上げ）に音声フィルタが適用されることは，誤った発声の音声ファイルを用いることで実現できる．音声ファイルを作成する際，正しく語彙を発声する音声ファイルに調整を加え，語彙を誤って発声する音声ファイルを用意する．今回の実験では，複数の学習者個人の特性におけるモデルの選好を調査するため，2 種類の発声の誤りパターンを用意した．2 種類の誤発声は 2 種類のモーラ類似度の計算方法 (Concat, Ave) に基づく．システム出力（単語の読み上げ）の都度，正誤いずれかの音声が確率的に使用（発声）され，実験時間を通して聞き取りづらい状態を模擬する．

誤って発声される音声ファイルは，誤発声に対応する文字列を Polly に入力することで生成される．モデルの語彙 2,000 単語のそれぞれについて，語頭 1 モーラおよび語尾 1 モーラを置換することで誤発声を作成する．モーラ置換規則は以下の手順で作成される．手順に関連する例をそれぞれ上位 5 つずつ表 1 に示す．

1. 2 つのモーラの全組み合わせについて類似度を計算し類似度ランキングを作成する（表 1a）．この類似度計算はモデルが持つ類似度テーブルに等しく 2 種類作成される (Concat/Ave).
2. 日本語辞書 [4] に収録された全単語から音素の出現頻度を数え，モーラの頻度ランキングを作成する（表 1b）．
3. モーラ類似度ランキングの上位から順にモーラのペアを取り出す．取り出されたペアに含まれるモーラのうち，より頻度が低順位のものを高

2） 詳細は先行研究 [1] を参照．類似度計算のためモーラをベクトル化する際，モーラに含まれる音素を結合 (Concat) するか，平均 (Ave) するかが異なる．

3） 濁音・半濁音・拗音を含む．語頭に存在し得ない「ん」は，生成に利用したモーラから除外した．モーラの出現頻度は，日本語の辞書 [4] に登場する頻度に基づく．

4） https://aws.amazon.com/de/polly/

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).

表 1: 誤発声ファイルを作成するためのモーラ置換規則の作成手順

(a) 類似度ランキング上位のモーラのペア.

順位	Concat	Ave
1	しゅ, じゅ	は, ひゃ
2	じゅ, じょ	きゅ, ぎゅ
3	にゅにょ	く, きゅ
4	にゅみゅ	ぐ, ぎゅ
5	りゅ, りょ	きゃ, ぎゃ
...	...	...

(b) 日本語の出現頻度ランキング上位モーラ.

順位	モーラ
1	う
2	い
3	く
4	か
5	し
...	...

(c) 置換規則の例. 2 種類のモーラ類似度に基づくふたつの置換規則について，それぞれ上位 5 つを例として示す.

順位	Concat	Ave
1	じゅ → しゅ	ひゃ → は
2	にょ → にゅ	ぎゃ → きゃ
3	みゅ → にゅ	ぎゅ → きゅ
4	りょ → りゅ	ぎょ → きょ
5	ず → ず	ひょ → ほ
...	...	...

表 2: 音声ファイルの例

フィルタタイプ	ひらがな	発音
オリジナル	てどお	＼
Concat ベースフィルタ	てどお	チドウ
Ave ベースフィルタ	てどお	テドウ

順位のものへ置き換える規則を追加する（表 1c）. ただし，連鎖する変換を防ぐため，現在の高順位モーラから変換する規則がすでに存在する場合は，現在のモーラペアを無視する.

表 2 は，この規則によって作成された誤発声の例である. 正しく発声される音声ファイルとふたつの誤って発声される音声ファイルについて，ひらがなで単語をとカタカナで発音を示した.

4 実験

これまでに述べた設定に基づいてオンライン実験を実施した. 音声フィルタ（に模擬される学習者個人の特性）がモデルの選好に影響するという仮説を検証するために，2 種類の音声フィルタについてそれぞれ 30 分間の単語選択制しりとりを実施し，システム中の各モデルが選択された回数を調査した.

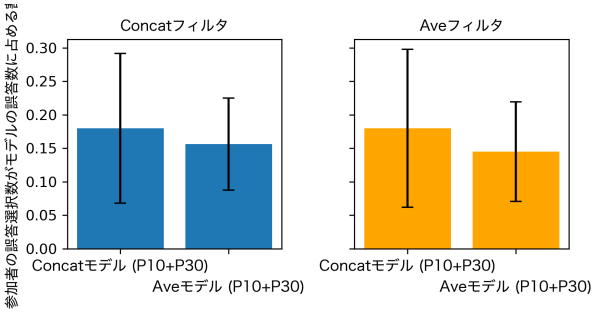

図 3: 参加者の誤答選択数がモデルの誤答数に占める割合. エラーバーは SE

4.1 方法

参加者はクラウドソーシングプラットフォーム Lancers[5] で募集された. 実験は，2 種類の音声フィルタについてのカウンターバランスおよびサーバ負荷分散の観点から，4 回に分割して実施された. 4 回の募集を通して計 81 名（うち女性 26 名，無回答 2 名，平均年齢 43.7 歳，SD=8.5）が参加した.

4.2 結果

結果を図 3 に示す. 横軸はモデルに対応する. モデルが持つ類似度テーブルごとに値を集約（類似度の影響度合 P について合算）した結果を示している. 縦軸は参加者の誤答選択数がモデルの誤答数に占める割合を示す. この指標は以下の式 1 で計算される.

$$\text{参加者の誤答選択数がモデルの誤答数に占める割合} = \frac{\text{参加者がモデルを誤提案を選択した回数}}{\text{モデルの誤提案の回数}} \quad (1)$$

図の最左にある青バーは Concat ベースフィルタ条件の 30 分間のしりとりにおいて，Concat テーブルのモデルに関する結果を示すものである. 式 1 に当てはめれば，分子は参加者が Concat テーブルのモデルを選択しつつ誤答した回数である. 同様に分母は，同条件で Concat テーブルのモデルが誤答した回数となる. この指標は，参加者があるモデルを選んで誤答した回数を，そのモデルの誤答のしやすさによって正規化した値であると言える. この図において，フィルタ間でモデルの選択されやすさが異なるとはいえない. つまり，音声フィルタがモデルの相対的選好に影響するという仮説は支持されなかった.

5） https://www.lancers.jp

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).

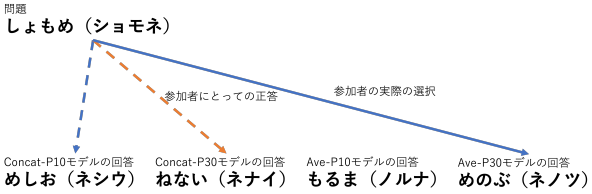

図 4: ある参加者による課題中の単語選択の事例．各単語のひらがながシステム内部で処理される単語で，カタカナはフィルタ適用後の発音．実線矢印は参加者による実際の選択．点線矢印は参加者にとっての正答．とくに橙矢印は，音声フィルタによって同じ音に聞こえるが実際には異なるモーラであるという，実験デザインから想定される誤りの例である．

4.3 参加者の選択に関する考察

単語選択制しりとり課題中の回答の事例をもとに，成人を参加者としてシステムの実現性を評価する実験デザインのための音声フィルタについて検討する．図 4 に，参加者による単語選択の事例を示す．参加者に聴こえる音声フィルタ適用後の発音は「ショモネ → ネノツ」であり適切である．また，しりとりとしても「しょもめ → めのぶ」であり正答である．

この事例で，音声フィルタの影響を受けた誤答として想定されるのは，「ねない（ネナイ）」のように問題の語尾と異なる語頭をもつ単語の選択である．参加者は「ネ」で始まる 3 つの選択肢を区別することはできないため，誤答に気づくことなく「ねない（ネナイ）」を選択する可能性があった．ただし，4 つの選択肢のなかで「ねない」の他にふたつが「ネ」音で始まっており，これらの語頭音は問題の語尾と同じ「め」である．参加者が選択肢を選ぶ際には UI 上の選択肢表示（色で区別されるロボットアイコン）に対する選好なども影響すると考えられるものの，このような多くの選択肢が同じように聞こえるしりとりが，参加者が判断の基準を定めることに対してのノイズとなった可能性がある．このような無作為な選択が，フィルタによる影響ではなく，参加者間のばらつきとして表れ，音韻意識の推定につながるモデルの相対的選好が観察できなかったことが考えられる．

表 3: フィルタ適用前後のモーラの個数

	モーラ個数
フィルタ適用前	100
Concat ベースフィルタ	46
Ave ベースフィルタ	52
Concat/Ave 間の重複	34

5 まとめ

本稿では，個別化認知モデルを用いた音韻意識形成支援システムの実現に向け，音声フィルタの適用によって成人を参加者としてシステムの実現性を評価する実験を実施した．クラウドソーシングを用いた実験の結果として，音声フィルタがモデルの相対的選好に影響するという仮説は支持されなかった．ただし，参加者ごとの結果においては，選択割合にばらつきが見られた．今回設定した音声フィルタによるモーラの置換では，しりとり中の多くの選択肢が参加者にとって区別できないもの（たとえば 4 択のうち 3 つが正答に聞こえるなど）となっていることで，参加者の選択が無作為に近いものになっている可能性がある．

今後は，音声フィルタの設定を吟味が必要である．困難を抱える子どもの特性（自閉スペクトラム症の特異な知覚）を再現する研究 [5] を参考に，知見に基づくフィルタを作成することで，より現実の困難に対応づいた検証が可能になる．これらの実験の結果を踏まえてシステムのブラッシュアップをした後，システムが本来対象とする言語学習者を対象としたシステム評価実験を実施する．

参考文献

[1] Jumpei Nishikawa and Junya Morita. Cognitive model of phonological awareness focusing on errors and formation process through shiritori. **Advanced Robotics**, Vol. 36, No. 5-6, pp. 318–331, 2022.

[2] John R Anderson. **How can the human mind occur in the physical universe?** Oxford University Press, 2007.

[3] Robert F. Port, Jonathan Dalby, and Michael O'Dell. Evidence for mora timing in japanese. **The Journal of the Acoustical Society of America**, Vol. 81, No. 5, pp. 1574–1585, 1987.

[4] 天野成昭, 小林哲生. 基本語データベース: 語義別単語親密度. 学習研究社, 2008.

[5] 長井志江. 自閉スペクトラム症の特異な視覚世界を再現する知覚体験シミュレータ. 精神看護, Vol. 19, No. 1, pp. 59–63, 2016.

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).