Estimating phonological awareness with interactive cognitive models: Feasibility study manipulating participants' auditory characteristics

Jumpei Nishikawa (nishikawa.jumpei.16@shizuoka.ac.jp)
Junya Morita (j-morita@inf.shizuoka.ac.jp)
Department of Information science and Technology, Graduate School of Science and Technology, Shizuoka University,
3-5-1, Johoku, Naka-ku, Hamamatsu, Shizuoka, 432-8011, Japan

Abstract

The difficulties encountered by children during language development varies among individuals. In particular, immaturity in phonological awareness, which supports speech perception, results in various speech defects. Accordingly, it is important to estimate the individual mechanism behind these problems to ensure proper support. In this study, we propose a method for estimating individual defects in the phonological process using cognitive models. As a preliminary step to targeting phonological processing difficulties in real world, we conducted an experiment with native adult speakers. Audio filters were applied to the output of the system to simulate phonological difficulties. This initial feasibility study revealed consistency in model preferences among participants when a particular audio filter was used. We consider that this study provides an important step toward the realizations of individualized cognitive modeling for mitigating various difficulties in language acquisition.

Keywords: ACT-R, Cognitive modeling, Phonological awareness, Individualized model

Introduction

Children (or second-language learners) face various difficulties during language acquisition. A prominent example is the segmentation of phonemes. In the early stages of language development, children perceive speech sounds as continuous but can gradually segment them into smaller units (Carroll, Snowling, Stevenson, & Hulme, 2003). In the process, sound can be segmented into various units (symbols), such as syllables and morae. As learners advance, they converge on a system of processing a series of units (e.g., mora in Japanese; Kubozono, 1989), as defined by their native language.

In the fields of developmental psychology and speech–language pathology, one of the abilities supporting this development is phonological awareness, which involves paying attention to phonological aspects of speech, such as phonemes and rhythm (Stahl & Murray, 1994). Some speech errors that occur during language development are attributed to a poorly formed phonological awareness of that particular language (Dynia, Bean, Justice, & Kaderavek, 2019; Kobayashi, 2018; Smith Gabig, 2010). In children with autism spectral disorder (ASD), an overall delay in phoneme acquisition and a partial inability to use some phonemes may occur (Grandin & Panek, 2013; Mugitani et al., 2019). To effectively support the formation of abilities that vary greatly among individuals, it is important to consider the cognitive characteristics of the individual child.

This study is a part of the studies aiming to develop a method of constructing a cognitive model adapted to the individual's phonological problem. Specifically, the current study is based on a model of Japanese phonological awareness (Nishikawa & Morita, 2022). The representation of phonological awareness is based on a general memory retrieval mechanism implemented in a cognitive architecture, Adaptive Control of Thought-Rational (ACT-R: Anderson, 2007), which is a framework for developing different models adapted to specific tasks and individuals.

The present study proposes a method of selecting a cognitive model and fitting it to individual phonological problems by varying the parameters in a previous model. The method is validated through an experiment in which participants' auditory traits are artificially manipulated. Finally, we discuss the feasibility of using the proposed method to estimate the users' state of phonological awareness in a summary of the experimental results.

Related Research

In this section, we introduce the literature relevant to the method used this study. We first present reports from clinical and experimental research on phonological awareness. Next, we introduce cognitive models of phonological awareness and previous research on tracing individual cognitive processes using cognitive models.

Research on Phonological Awareness

Several clinical reports and investigations have utilized experimental methods to investigate phonological awareness formation. Cases of children confusing certain morae have been reported in clinical speech–language pathology practice. Grandin reported that she could not distinguish silent consonants well in her childhood (cat, pat, and hat sounded like the same word) and stated that a child who can only utter vowels (consonant deletion) likely does not hear the consonants (Grandin & Panek, 2013).

In Japanese language, two- or three-year-old infants reportedly tend to confuse morae containing the consonants /r/ and /d/ (Kobayashi, 2018). This erroneous speech pattern diminishes when they reach four- or five-years. However, such phonological discrimination is sometimes delayed. A Japanese textbook (Oishi, 2016) for speech–language pathologists states that children with developmental disorders have difficulty distinguishing between vowels and vowel–consonant combinations with the same vowel (e.g., "a" and "ka"). These reports exhibit commonalities with the previ-

ously mentioned English reports (Grandin & Panek, 2013). Based on these reports, the current study focuses on phonological awareness in the development of Japanese language skills.

Several Japanese studies on phonological awareness have used the popular word game *Shiritori* as a task. *Shiritori* involves players taking turns uttering a word (noun); the word must begin with the mora that the previous word ended with. For example, after a player answers *"ri-n-go"* (meaning apple), the next player continues with *"go-ma"* (meaning sesame seeds). Takahashi (1997) examines the relationship between the stages of phonological awareness formation and the conditions necessary to playing *Shiritori* through a psychological experiment in cross-sectional development in children with typical development. Takahashi has shown that phonological awareness (especially the ability to segment sounds into morae and a mental lexicon indexed by morae) is a prerequisite for *Shiritori*. Takahashi also suggests that playing word games common to a specific culture, such as *Shiritori*, is important to the growth of phonological awareness in the mother tongue. Building upon this research, we utilize *Shiritori* as a task to be applied to the phonological awareness model.

Cognitive Modeling of Phonological Awareness

As noted in the first section, a cognitive model that focuses on human internal processes in the formation of phonological awareness exists (Nishikawa & Morita, 2022). This model assumes innate and experiential constraints of language acquisitions based on parameters implemented in ACT-R. In their model, from the viewpoint of generative phonology (Chomsky & Halle, 1968), innate factors are associated with sound similarities between Japanese morae. Based on this assumption, a partial match mechanism of ACT-R retrieves erroneous phonological knowledge, and it exhibits commonalities with the reports regarding the phonological awareness formation process (Kobayashi, 2018; Oishi, 2016). In addition, the model assumes that such errors derived from an innate factor can be mitigated by an experiential factor, with repetitive practice strengthening correct phonological knowledge.

Although the above study suggests the possibility of describing error patterns in a unified cognitive architecture, there are limitations in the number of error patterns and their practical correspondence to actual individuals. In contrast, many cognitive modeling researchers are trying to represent various individual differences (Smith, Chiu, Yang, Sibert, & Stocco, 2020; Somers, Oltramari, & Lebiere, 2020; Mätzig, Vasishth, Engelmann, Caplan, & Burchert, 2018). These studies constructed models with varying cognitive architecture parameters to fit target individuals.

Such models have been utilized in studies on support systems involving real humans as users. Model-based systems have been developed in the same studies to identify the current state of individual users to guide their activities (Anderson, Boyle, & Reiser, 1985; Klaproth et al., 2020; Morita, Hirayama, Mase, & Yamada, 2016; Morita et al., 2022). For example, the model-based reminiscence method by Morita et al. (2016) extends one of the existing mental health care methods for dementia patients. This study attempts to guide appropriate memory recollection by incorporating a cognitive model corresponding to individual users into a system of a photo slide show for reminiscence.

Table 1: Model declarative memory

(a) Word knowledge

word	sound
ringo	"rʲiŋgo"
goma	"goma"
riku	"rʲikɯᵝ"
...	...

(b) Phonological knowledge

mora	sound
/ri/	"rʲi"
/go/	"go"
/ku/	"kɯᵝ"
...	...

(c) Word–mora relationship

word	mora	position
ringo	/go/	tail
goma	/go/	head
goma	/ma/	tail
...	...	...

Individual Models of Phonological Awareness

Building upon previous studies, the current study develops a method of estimating users' internal state by utilizing their interactions with cognitive models. To achieve this goal, this section describes the cognitive model of phonological awareness proposed by Nishikawa and Morita (2022) as the base model for fitting individual users.

Phonological Awareness Model

Nishikawa and Morita (2022) targeted the phonological awareness observed during a *Shiritori* game. In this paper, we only present the basic model functions necessary to achieving an individualized cognitive model and the settings for individualization. For details, please refer to the original article.

Knowledge Representation Required for *Shiritori* This model realizes *Shiritori* based on the general implementation method of the model using ACT-R. That is, declarative knowledge is expressed in *chunks*, and the *Shiritori* procedure is represented by the application of *production rules* that manipulate the ACT-R modules. In the following, we show these two types of knowledge representation in the model.

Declarative chunks The model in this study retains three types of declarative chunks that relate to word (vocabulary), phonological (mora) knowledge, and the association between them (Table 1). These three types of chunks can be regarded as a network, consisting of the word chunk (chunk type (a) in Table 1) and the mora chunk (chunk type (b) in Table 1) nodes and the paths connecting them (Table 1 (c)).

Production rules When the model receives a word as the partner's answer, it traverses the network of declarative chunks to search for a word that follows the rules of *Shiritori*. This process is performed by applying the model's production rules (procedural knowledge) in the following steps.

1. Using the word chunk (chunk type (a) in Table 1) acquired by the aural module, the model retrieves a chunk

that connects the word and the ending mora (chunk type (c) in Table 1).

2. Using the retrieved word–mora association knowledge, the phonological knowledge (chunk type (b) in Table 1) corresponding to the word ending is retrieved.
3. Using this phonological knowledge, the word that begins with the mora is then retrieved, and the selected word is held as a candidate answer in the goal module.
4. Afterward, the model checks that the stored answer candidate is valid according to the rules of *Shiritori*, such as not having been previously answered in the current *Shiritori* trial.

(a) If the current candidate violates these rules, the model again searches for a candidate answer.

(b) When the candidate word is confirmed as valid, the model stores it in the declarative module as an answered word and outputs the word through the speech module.

ACT-R Parameters for Knowledge Retrieval In the process described above, the phonological awareness involved in paying attention to word endings corresponds to the retrieval of phonological knowledge from word knowledge. Knowledge retrieval in ACT-R is controlled by a parameter called activation that is assigned to each chunk, and it affects the success or failure of the retrieval. The values are computed as the sum of several terms, such as learning effects, contextual effects, similarity between chunks, and a noise term that gives stochastic fluctuations to the activation values.

The similarity term (P_i) is noteworthy in the model's representation of phonological awareness. As mentioned earlier, Nishikawa and Morita (2022) incorporated innate bias into the knowledge similarity between mora chunks.

$$P_i = PM_{ki} \tag{1}$$

This value is computed as the summation of the weighted degree of similarity M_{ki} for each retrieval request k to chunk i. M_{ki} is typically negative, and P serves as a penalty during similarity retrieval. In addition, the partial matching following the introduction of similarity makes it possible to reproduce flexible choices and certain types of errors.

Diversified Models

We extend the phonological awareness model constructed in the previous study (Nishikawa & Morita, 2022) to account for the different problems in phonological awareness. One of the elements to be manipulated to construct an individualized model is the method of computing similarity between the morae. We prepared several *similarity tables* in this study for defining a method of calculating the similarities between morae.

It also manipulates the coefficient corresponding to the magnitude of the influence in the similarity table (P in Eq. 1). It is expected that the similarity table and coefficient P will result in a high level of similarity between certain mora pairs, which will allow us to address the real-world phenomena (i.e., confusion of /r/ and /d/, consonant deletion, etc.).

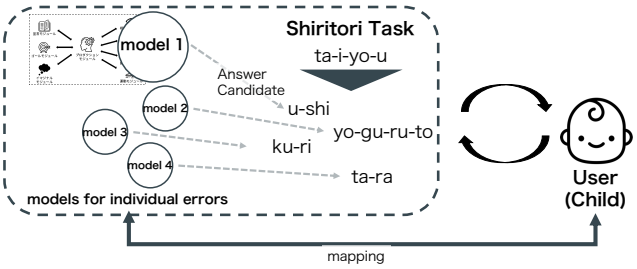


Figure 1: Concept of Individual Models and Tasks

Shiritori Game System for Model Selection

Figure 1 conceptualizes multiple models and *Shiritori* tasks. To confirm that the models constructed in the previous section can capture an individualized phonological process, we set up a task in which the participants play *Shiritori* with the models. This section describes the system developed to perform the task.

User Interface

Figure 2 shows the user interface of the system. In this system, a word-choice-based Shiritori game is set as a task. The user responds by selecting the appropriate word from the candidate answers proposed by multiple models. This response format is based on the Shiritori used in phonological awareness studies (e.g. Takahashi, 1997).

Procedure of the Model Selection

Figure 3 shows the flow of system usage comprising the following six procedures.

1. First, the system presents the starting word (In Figure 2, "せみぷろ") to a set of individualized models and users by playing audio from the experimental window (Figure 3①).
2. Each model recognizes the starting word (Figure 3②) and its ending, and it answers according to *Shiritori* rules (Figure 3③).
3. The words answered by the model are displayed in the experiment window (Figure 3④) and serve as choices for the user to select as an answer.
4. The user answers (chooses) a word they deem appropriate based on the starting word and candidate words (Figure 3⑤). Here, selecting a word is equivalent to selecting the model that proposed the word.
5. When the system receives the user's response, it records the model that proposed the chosen answer (Figure 3⑥).
6. After a series of processes, the user's answer is used as the next starting word, and the game is repeated (Figure 3⑦).

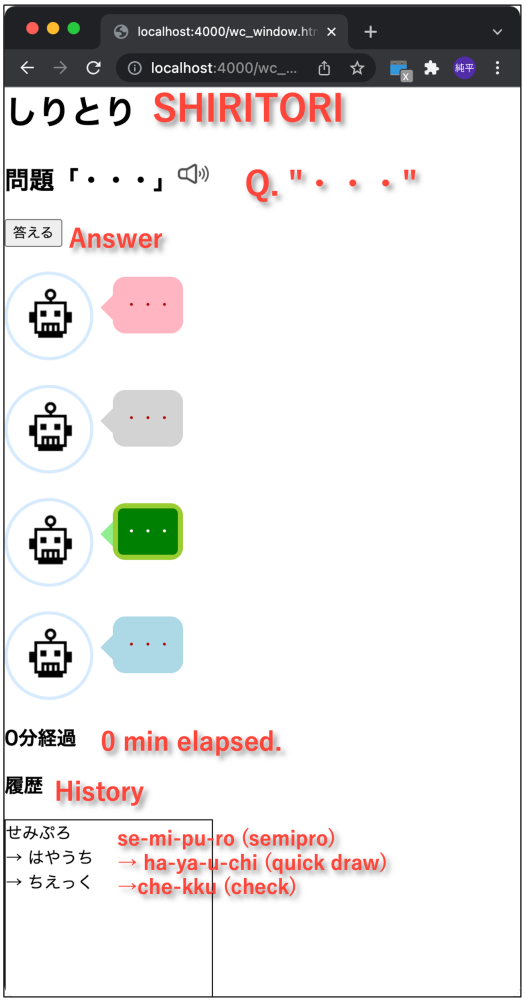


Figure 2: User interface of the system. The upper, middle, and bottom part of the screen show the question (a speaker icon), choices (robot icons and balloons), and game history, respectively. This screen shows that the user has selected a word from the green model for the fourth *Shiritori* word. The red strings are shows for explanatory purpose.

The above process is reiterated, and the best-matched model is ultimately selected according to the frequency of choice.

Experiment

This section describes the experiment for testing the proposed method of estimating individual phonological processes by selecting cognitive models. The concept behind the experiment is shown in Figure 4. Because this experiment is an initial feasibility study, adult Japanese native speakers were placed in situations where the phonological process was artificially generated. In each turn, an audio filter receives the model (a word) to generate phonological processing difficulties for participants. In this setting, we test the following hypothesis: different audio filters produce different model preferences in a word-choice-based *Shiritori* task.

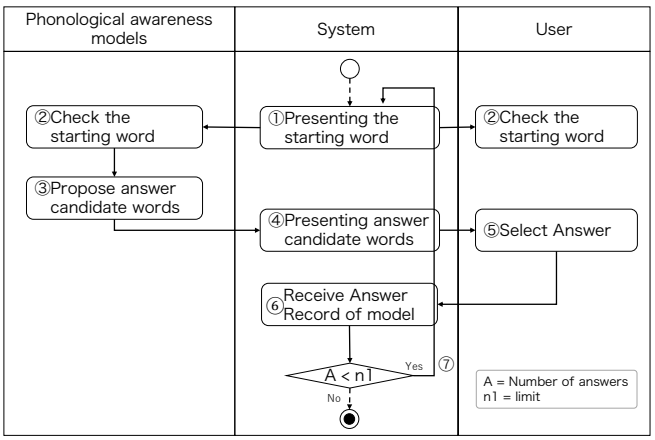


Figure 3: System activity and task progression

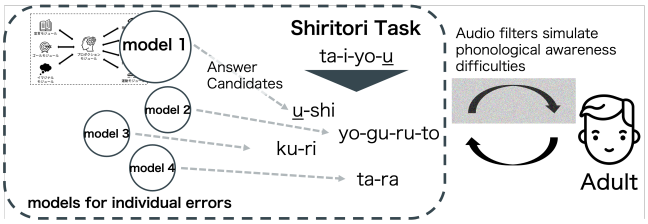


Figure 4: Concept behind the Experiment

Method

Experimental Design The participants' behaviors were compared by manipulating model parameters and audio filters. The specifics of the manipulation are as follows:

Model Settings Four models were prepared, as indicated in Table 2, and the following two factors were considered.

Similarity table This factor indicates the difference in computing similarity between morae. We used the Consonant–Vowel concatenation table (C–V concatenation table) and the Consonant–Vowel average table (C–V average table), both were presented by Nishikawa and Morita (2022).

Similarity coefficient This indicates the degree of error suppression caused by morae similarity ($P = 10,\ 30$ in Eq. 1).

Filter settings We prepared two audio filter settings [+10 and −10 filter conditions]. These filters indicate the formant setting of Voice Transformer, which is a plug-in effect of Apple GrageBand (MacOSX 10.x). Negative and positive formants transform input voice into deep/muffled and high-pitched/thin tones, respectively.

Participants One male graduate student and one female undergraduate student participated in the experiments. Both were native Japanese speakers majoring in informatics. Henceforth, the two participants will be referred to as Participant A and Participant B.

Table 2: Individual model settings and their notaions

	C–V concatenation table	C–V average table
similarity coefficient $P = 10$	P10-SIM1	P10-SIM2
similarity coefficient $P = 30$	P30-SIM1	P30-SIM2

Table 3: Experimental procedures and conditions for each participant

		Participant A	Participant B
Instruction	(5 min)		
Shiritori 1	(25 min)	$+10$ condition	-10 condition
Break	(5 min)		
Shiritori 2	(25 min)	-10 condition	$+10$ condition
Questionnaire	(5 min)		

Apparatus The system was displayed on a external monitor connected to the laptop (Apple MacBookPro M1 2020, macOS Big Sur) and operated with a built-in touchpad. Google Cloud Text-to-Speech API was used for the audio output of the system. This was then input to GarageBand through a virtual audio driver called BlackHole. The audio input to GarageBand was distorted by a large formant shift, as described above. The affected audio was output from the speakers through BlackHole from the GarageBand monitor function.

Procedure The flow of the experiment is presented in Table 3. First, the participants were seated in front of the display showing the system, and the experiment objectives and how to use the system were orally explained. After, the participants confirmed that they had no questions, they performed the *Shiritori* task. Each condition was allocated 25 min for one *Shiritori* task, and the participants answered a questionnaire after completing two *Shiritori* tasks. The two *Shiritori* tasks were performed using different audio filter settings. Participant A was subjected to the $+10$ condition, followed by the -10 condition. Participant B was subjected to the -10 and $+10$ conditions in that order.

Results

In this section, we analyze the effect of the audio filter by tabulating the model whose answers were selected by the participants.

Number of model selections. Figure 5 (a) is a stacked bar graph showing the model chosen by the participant. The four bars denote the different combinations of participants and filter conditions. The vertical axis corresponds to the total number of selections without considering the correctness of the model answers. Although the bars have different heights, we can find commonalities in all conditions. Regardless of the difference between participants/filters, the P30-SIM2 (thick green regions) was the most frequently selected, whereas the P10-SIM1 (light blue regions) was chosen the least often.

To confirm such commonalities across the conditions, we constructed a correlation matrix for all combinations of the conditions (Figure 5 (b)). The Spearman correlation coefficients were computed, treating each model as a unit of analysis ($n = 4$). The figure indicates high correlations in all cells in the matrix, suggesting that frequently selected models were common across participants and filters.

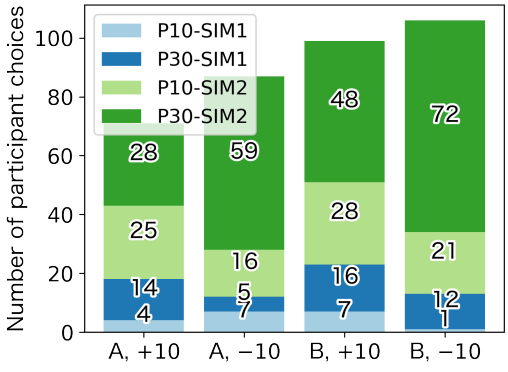
This result agrees with the performance of the original models. As in Eq. 1, the large P suppresses errors derived from the similarities among morae. Regarding the similarity table, Nishikawa and Morita (2022) suggested that a high error rate in the C–V concatenation table is representative of consonant deletions that are similar to those that occur in children with ASD (Grandin & Panek, 2013; Mugitani et al., 2019).

Despite the above overall commonalities among conditions, slight differences exist in the correlation between condition pairs. In particular, Participants A's $+10$ filter condition has relatively weak correlations with the other filter (-10) conditions (Participant A's -10 filter condition: $r = .73$, $p > .10$; Participant B's -10 filter condition: $r = .80$, $p > .10$). These results validate our assumption, revealing that different audio traits lead to different preferences for phonological models.

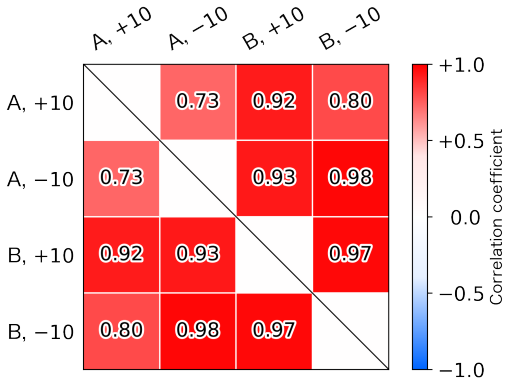
Number of incorrect selections. Figure 6 (a) is the stacked bar graph showing the number of chosen models limited to the incorrect answer. Unlike in the previous graph, large differences exist between participants/filters in the graph. The same is confirmed in Figure 6 (b). The figure reveals a significant correlation ($r = 0.97$, $p < .05$) between participants in the $+10$ condition. No significant correlation exists among the other conditions. These results suggest that under the $+10$ filter, the preference for the models was consistent across the participants.

Discussion

In this experiment, we tested the hypothesis that different audio filters (artificially generated individual differences in auditory traits) produce different model preferences. To this end, we compared the selection of the models under the different participants and audio filters. As shown in Figure 6 (b) there was a significant correlation for only the $+10$ condition among the participants, whereas no significant correlation was observed for the other combinations. In other words, when the $+10$ filter is applied, there is some match among participants regarding the ease of model selection. Therefore, the tested hypothesis has some validity.

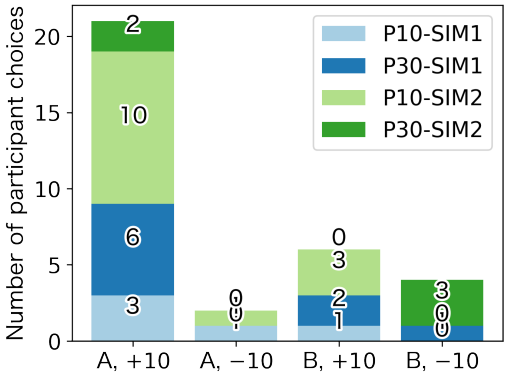


(a) Number of participant choices for each model.

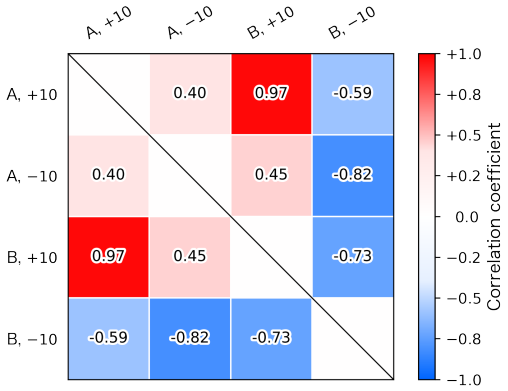


(b) Correlation matrix of the number of model selections (each bar in Figure 5a) across conditions.

Figure 5: Model selection by participants



(a) The number of choices of the model when a participant incorrectly selected.



(b) Correlation matrix of the number of incorrect model selections (each bar in Figure 6a) across conditions.

Figure 6: Incorrect model selection by participants.

Summary and Future Work

We proposed a method for fitting a cognitive model to individual phonological problems. We prepared cognitive models for a *Shiritori* game and assigned participants to play a word-choice-based *Shiritori* game. By the participants repeatedly selecting the words proposed by the model, we could estimate a model that structurally represents the participants' phonological awareness. We evaluated the feasibility of this method in an experiment with adult native speakers. The feasibility experiment was designed to simulate phonological processing difficulties by applying an audio filter to the words proposed by the model.

In this experiment, we tested the hypothesis that different audio filters produce different model preferences in a word-choice-based *Shiritori* task. The experiment involving two participants revealed a significant correlation in model selection among the participants under certain audio filter conditions. This means that there is consistency among the participants in the model that is more or less likely to be selected as the incorrect choice (not appropriate as a *Shiritori* answer). Accordingly, the experiment hypothesis had some validity.

This research ultimately aimed to develop a phonological awareness support system that utilizes cognitive models that are adapted to individual error patterns and the individuals themselves to estimate the users' phonological awareness. Accordingly, the results should be analyzed further. In this paper, we only analyzed the number of answers and the frequency of model selection by focusing on wrong answers in *Shiritori*. However, it is necessary to rigorously confirm the corresponding effects by performing tests with statistical methods.

Moreover, the method should be expanded. The extent of parameter exploration in the system presented in this study is limited. Only four models were prepared and selected by the participants. In the future, we intend to construct a method of automatically generating models by combining parameters related to phonological awareness. Experiments under such dynamic conditions are also necessary. After sufficient feasibility studies have been conducted, we will conduct an experiment involving learners, who are the original target of the method, such as children with phonological awareness problems and second-language learners.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*(4698), 456–462.

Carroll, J., Snowling, M., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, *39*(5), 913–923.

Chomsky, N., & Halle, M. (1968). *The sound pattern of english*. Harper & Row New York.

Dynia, J. M., Bean, A., Justice, L. M., & Kaderavek, J. N. (2019). Phonological awareness emergence in preschool children with autism spectrum disorder. *Autism & Developmental Language Impairments*, *4*, 2396941518822453.

Grandin, T., & Panek, R. (2013). *The autistic brain: Thinking across the spectrum*. Houghton Mifflin Harcourt.

Klaproth, O. W., Halbrügge, M., Krol, L. R., Vernaleken, C., Zander, T. O., & Russwinkel, N. (2020). A neuroadaptive cognitive model for dealing with uncertainty in tracing pilots' cognitive state. *Topics in cognitive science*, *12*(3), 1012–1029.

Kobayashi, H. (2018). *Oninishiki no keisei to kotoba no hattatsu -"kotoba ga osoi" wo kangaeru-*. Kodama shuppan.

Kubozono, H. (1989). The mora and syllable structure in japanese: Evidence from speech errors. *Language and Speech*, *32*(3), 249–278.

Mätzig, P., Vasishth, S., Engelmann, F., Caplan, D., & Burchert, F. (2018). A computational investigation of sources of variability in sentence comprehension difficulty in aphasia. *Topics in cognitive science*, *10*(1), 161–174.

Morita, J., Hirayama, T., Mase, K., & Yamada, K. (2016). Model-based reminiscence: Guiding mental time travel by cognitive modeling. In *Proceedings of the fourth international conference on human agent interaction* (pp. 341–344).

Morita, J., Pitakchokchai, T., Raj, G. B., Yamamoto, Y., Yuhashi, H., & Koguchi, T. (2022). Regulating ruminative web browsing based on the counterbalance modeling approach. *Frontiers in Artificial Intelligence*, *5*.

Mugitani, R., Homae, F., Hiroya, S., Satou, Y., Shirose, A., Tanaka, A., … Tachiiri, H. (2019). *Kodomo no onsei* (No. 21). Corona Publishing Co., Ltd.

Nishikawa, J., & Morita, J. (2022). Cognitive model of phonological awareness focusing on errors and formation process through shiritori. *Advanced Robotics*, *36*(5-6), 318-331.

Oishi, N. (2016). Evaluation. In H. Ishida & I. Ishizaka (Eds.), *Gengochokakushi no tameno gengohattatsushogaigaku* (2nd ed., pp. 77–117). Ishiyaku shuppan.

Smith, B., Chiu, M., Yang, Y., Sibert, C., & Stocco, A. (2020, 07). Modeling the effects of post-traumatic stress on hippocampal volume..

Smith Gabig, C. (2010). Phonological awareness and word recognition in reading by children with autism. *Communication Disorders Quarterly*, *31*(2), 67–85.

Somers, S., Oltramari, A., & Lebiere, C. (2020). Cognitive twin: A personal assistant embedded in a cognitive architecture. In *Proceedings of ICCM 2020, 18th international conference on cognitive modelling*.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of educational Psychology*, *86*(2), 221–234.

Takahashi, N. (1997). A developmental study of wordplay in preschool children: The japanese game of "shiritori". *The Japanese Journal of Developmental Psychology*, *8*(1), 42–52. (In Japanese)