Abstract: Background: Understanding is a crucial element for accurate communication during foreign language learning. The learning results of conventional teaching strategies are often poor due to their very low preparation and response. There are more practicabilities and advantages to speech recognition technology, which provides immediate responses, but its incorporation requires cautious study to safeguard its practicality and advantages. Method: In this paper, we gather speech signal datasets from 120 learners and divide them equally into two groups such as baseline and study group. In baseline group we provide a traditional technique but in study group, we proposed a new technique named artificial rabbit optimized hidden markov model (AR-HMM). Additionally, we made an analysis with proposed technique to improve the listening training effect in foreign English language. Result: As a result, first we used nine questionnaires to examine the participant's experience. We evaluate the performance of the proposed technique and compare it with an existing technique based on the parameters such as accuracy (95%), involvement (low (15%), medium (59%) and high (93%)), efficiency (90%) and time (14hr). Conclusion: When compared to the existing technique, our proposed technique has superior performance than others to improve listening training effect in foreign English language.

Keywords: Listening, training, foreign language, artificial rabbit optimized hidden markov model (AR-HMM)

I. INTRODUCTION

The efficiency of listening instruction in foreign languages can be greatly increased by incorporating recognition of speech algorithms [1]. Speech popularity software efficaciously converts conversations into text by using herbal language processing and device studying. These algorithms have numerous benefits when used in language learning environments [2]. During listening lessons, learners receive real-time feedback from speech recognition algorithms. Students can also quickly detect and fasten pronunciation mistakes these immediate comments, which is vital for boosting spoken proficiency and listening comprehension [3]. Learners can more effectively reinforce proper pronunciation styles by way of getting immediate corrections.

The use of speech recognition technology allows for customized educational opportunities [4]. Considering the capability to adjust to the unique knowledge of techniques and ability stages, these algorithms can provide every student with duties and feedback that are specifically desirable to them. Because they feel encouraged in their language learning process, pupils are more engaged and motivated while using this individualized approach [5]. Methods for speech recognition make learning on your own easier.

Despite educator's constant oversight, students can improve their listening skills at their own pace [6]. The flexibility enables students to take charge of their training and manage their time in accordance with their schedules, which promotes more effective ability development. The speech reputation application is capable of producing comprehensive performance reports [7]. When speech recognition algorithms are used in foreign language listening instruction, inclusivity and accessibility are promoted. Learners with varying learning requirements, inclusive of people with language or listening to issues, can advantage of the visible comments that language translation affords [8].

The graphical tool makes language acquisition possibilities available to all students, irrespective of their unique difficulties, by helping them take an active role in listening exercises and better understand spoken information [9]. Instructors can learn a great deal about each student's development and areas for growth by examining how they engage with the system. The efficacy of language training is increased by these analytics, which allow for targeted interventions like more practice especially listening skill regions or individualized coaching sessions [10].

There are many advantages to the usage of speech-recognizing technology in foreign language listening instruction, such as overall performance information analysis, self-reliant exercise, individualized coaching, real-time comments, and absolutely immersive classrooms [11]. Teachers can grow the efficacy of their listening
coaching and provide students with the tools they need to become more proficient in the language by utilizing technology.

A. Motivation

Recognition of the imperfections in traditional methods of passive listening, lack of context, low reliability in resources, difficulty individualizing, evaluation difficulties, low interaction, and few feedback mechanisms is the driving force behind enhancing listening training's impact on foreign language instruction. These shortcomings motivate the development of a novel approach artificial rabbit optimized hidden markov model (AR-HMM) to improve listening skills in foreign English language, ultimately aiming to optimize students' language proficiency.

B. Aim

The purpose of the paper is to utilize an artificial rabbit-optimized hidden markov model (AR-HMM) to improve listening skills in foreign English languages.

C. Organization

Part 2 discusses the related work, Part 3 provides comprehensive information about listening, Part 4 elaborates on the process of methodology, Part 5 represents the result and discussion and Part 6 discusses the conclusion of the paper.

II. RELATED WORK

In comparison to conventional teacher-led feedback and training, was to determine whether the method could be a useful tool for improving second language (L2) pronouncing and speaking abilities described by Sun (2023) [12]. Sixty-one intermediate-level Chinese EFL students were split into two groups randomly: the experimental group (EG) and a control group (CG). The findings of the experiment indicate that one effective way to help EFL learners improve their L2 pronouncing and ability to speak was to use ASR technology in conjunction with peer correction.

Dillon and Wells (2023) explored the impact of automatic speech recognition (ASR) technology-based pronouncing instruction on typical pronunciation problems made by Korean students of English [13]. The individuals were split up into two distinct categories. One group received training and instruction on how to practice pronouncing words correctly using automated voice recognition and the other group did not. However, there were still some reservations about the test's technical components, respondents reacted well to the use of automated speech recognition for practicing pronunciation and evaluation.

The pronunciation abilities of the participants were evaluated using a pretest-posttest controlled experimental designed by Saadia (2023) [14]. In order to assess participants' beliefs and perspectives on the usage of both instruments, an assessment was also administered. The findings indicate that text-to-speech (TTS) and ASR can help learners acquire desired pronunciation qualities and can support various phases of pronouncing improvement. The experiment highlights the educational benefits of using TTS and ASR technologies for pronouncing training, improving students' speech abilities, and extending the learning experience outside of traditional classrooms.

Jiang et al., (2023) defined to show the relevant paths to development and investigate how ASR-based technologies affect the oral proficiency and precision of English language learners [15]. Employing a long-term experimental design, a 14-week quasi-experiment involving one hundred sixty first-year college learners was conducted. To arrive at the within- and between-subjects effects and the associated interaction impacts, a two-way multiple measure ANCOVA was performed. The group receiving treatment achieved better than the control group in terms of phonetic precision, speed comprehension, and restoration fluency, as evidenced by the between-subjects effect.

The assessment methods for conventional spelling, effortlessness, and accent, respectively, build a framework for the assessment of spoken English pronouncing quality based on a dynamic temporal warping (DTW) algorithm proposed by Duan and He (2023) [16]. The efficient fusion of several dimension assessment features is achieved by the application of a support vector regression (SVR) technique. Additionally, an evaluation measure that assessed students' oral pronouncing quality from various angles was correlated between computer and human rating. The model performs better than average when pronouncing words that are either syllable or polysyllabic, which helps pupils learn English by word.

John et al., (2023) investigated Google Translate's (GT) automated speech recognition (ASR) capacity to render suggestions for pronunciation problems in second languages (L2) [17]. Four hundred eighty records of male and female students realizing and vowel-initial targets in sentence-final location were used to test the software. The
transcribed accuracy rates of the initial objectives, whether realized properly or erroneously, were lower than those of h-initial or vowel-initial items, indicating a hierarchy of accuracy among pronunciation faults. The consequences of these findings extend to both L2 teachers and learners.

In order to obtain learners’ voice signals using many sensor nodes, a speech recognition method based on portable sensor networks was employed by Jingning (2024) [18]. The accuracy to be acknowledged was increased by applying an algorithm for machine learning to the obtained information after preprocessing and feature extraction. The findings from the experiments demonstrate that a system for speech recognition based on a mobile network of sensors performs better in the presence of noise from the surroundings and operator change when compared to traditional speech recognition methods.

The possible use of speech-to-text (STT) programs, a variant of automatic speech recognition (ASR) technology, to assess adult EFL students’ accents of distinct first languages (L1) was examined by Hirai and Kovalyova (2023) [19]. The experiment concluded with the observation that speech recognition for EFL learners continues to require work, with an accuracy rate of between 50 and 70 percent. On the other hand, EFL learners were able to recognize their pronunciation problems because they did not have flawless voice recognition. Teachers can even more confidently utilize STT applications because their assessment of pronouncing was proven to be comparable with that of human raters.

Geng (2023) developed a feature extraction and long short-term memory (LSTM) network-based speech recognition tool for English language instruction [20]. To analyze voice data, identify language attributes, and categorize them, an LSTM network was employed. The set of speech data was tested and assessed using a set of various English-speaking words. According to the experimental results, speech recognition systems based on the LSTM network can distinguish between words with effectiveness and obtain the greater classification accuracy in a variety of electromagnetic interference and noise environments.

The way adults performed differently when using an automated speech recognition (ASR) system and receiving feedback from peers or practicing on their own was explored by Evers and Chen (2022) [21]. During twelve weeks, the trial group and the control group both utilized the identical ASR software. Peer feedback appears to be a more successful method for correcting speech than individual practice, as seen by the large difference in both groups’ post-test results. Compared to the comparison group, the experimental group expressed more satisfaction with the program and the results, difficulties in implementing the ASR software, and effects on adult learning languages.

Ran et al., (2021) described effective techniques as the system’s methods of machine learning models, and they examine and improve voice recognition methods [22]. In the meantime, the fundamentals of language cutting including front-end speech analysis and feature parameter extraction were expounded upon, with phoneme-level language correction of errors as the foundation. It was done after presenting the fundamental concepts, building, and instruction of acoustical models. A control experiment was also created for this study in order to evaluate and validate the model for recognition of voices and rectification using machine learning. The results of the experiments show that the strategy recommended in the study has a specific effect.

The challenge of assessing English recognition of speech and pronouncing performance using deep learning was investigated by Xu (2022) [23]. It chooses pronunciation, acceleration, and rhythm as the differentiators to assess the quality of English speech. The findings of our assessment and the outcomes of the personal assessment clearly demonstrate the higher reliability of the deep learning Language speech detection and pronouncing quality model presented in this work. Just thirty-two of the 240 samples that were analyzed had a grade difference; the remaining samples were all similar.

Spring and Tabuchi (2021) evaluated the efficacy of using an ASR tool to assist L2 Japanese students in improving their pronunciation in an online EFL course [24]. The few learners have dealt with a total distance learning environment or attempted to employ the data to decide which courses were the most significant, previous research indicates that ASR tools can be useful in the area. The findings indicate that in general, students felt positive about the ASR-assisted training and that, particularly for those who started off less proficient, they clearly increased their clarity.

In order to derive Mel-frequency cepstral coefficient (MFCC) characteristics from the language signal, the English pronunciation features were examined by Xiong (2023) [25]. After that, the instances of faulty and precise pronunciation were distinguished using the support vector machine (SVM) approach. In order to enhance the recognition impact, deep features were recovered through the use of the deep brief network (DBN) as the SVM’s input. The sparrow search algorithm (SSA) was then employed to enhance the configurations of both the DBN and SVM. The suggested approach for pronouncing English recognition of features was shown to be reliable by the results, and it can be used in spoken language instruction.
III. LISTENING

In this section, we discuss the character of listening, types of listening, procedures, and techniques that help to improve listening skills. Figure 1 represents the general process of listening.

Figure 1: General process of listening

A. Character of listening

The cognitive process of listening is how we give meaning to sounds. It is the active intellectual process of deciphering, comprehending, analyzing, and assessing communication. It’s a particular type of interaction that’s just as important as writing, reading, and speaking. As the globe grows more interconnected and information becomes sophisticated, we must improve our capacity to listen.

When we listen, we prioritize the information that we want to hear and that we can identify. Include another way, we choose what data is essential for comprehension in order to understand what somebody says and respond accordingly. Listening is an artistic ability. When people listen, they do so for many reasons. In order to assist students, organize their ideas and utilize intelligent guessing to make sure they accomplish their listening goal, it is crucial to clarify what listening purpose they have when learning a new language, such as listening for specific information or listening for an overarching meaning or idea.

The primary ability that allows students to use their other abilities is listening. A student won’t have any trouble speaking if they can comprehend what they are hearing. Since listening gives the learner input, it is essential. Moreover, learning cannot start if students do not understand the information that they are given and listening is essential to language learning.

B. Listening procedure

Recognizing the listening process might assist us in reevaluating the ways in which listening instruction is delivered. For this reason, top-down and bottom-up processing are the two essential elements in explaining the listening process.

1) Top-down

A top-down approach uses the data offered by words and sounds to infer the meaning of a listening text based on prior knowledge and experience. The listener uses what she knows about the context the topic, the speakers, the circumstances, and what she can make sense of the auditory input in order to interpret a text. Maintaining searching for specific information, collecting, predicting the future, thinking, and drawing conclusions are examples of top-down listening techniques.

2) Bottom-up

In order to comprehend the text, learners can identify vocabulary and sound aspects with the aid of bottom-up analysis. Bottom-up activities are very helpful to learners who need to increase their vocabulary since they directly target language patterns at both word and phrase levels. They will perceive and process aural material more quickly and accurately as they grow more conscious of the language components of the input. Learners could be asked to discern particular sounds, word boundaries, and syllables that are stressed, determine grammatical structures and
operations, identify abbreviations and related speech, and recognize connecting words in order to cultivate bottom-up learning.

C. Types of Listening

Listening can be divided into seven categories. Every kind supports learners in developing a variety of abilities and approaches. Figure 2 displays the types of listening.

![Figure 2: Types of listening](image)

1) **Interactive**

   It refers to creating suitable answers and concentrates on assisting listeners in becoming aware of variations in cultural feedback styles and methods for offering them. Learners can participate in collaborative discourses more easily and effectively if they are aware of the alternatives and methods available to them as listeners.

2) **Selective**

   It refers to the data input provided to tasks that are designed to assist students in extracting specific details from writings, regardless of whether those texts are much beyond the scope of the student's current language and content understanding.

3) **Discriminative**

   It specifies differentiating behavior for the hearing and for recognizing the audible and visual information, and it forms the foundation for all other listening activities.

4) **Intensive**

   It is the formalized feedback to activities that, after the significance of the text for certain material has been created, tries to direct learners' attention to language system elements.

5) **Appreciative**

   Listening is taking in information for the purpose of enjoyment or education, which may involve experiencing pleasure or acquiring sensory insights. The process of taking in and interpreting auditory data in order to find fulfillment or meaning is included.

6) **Comprehension**

   It is crucial to the comprehension of the content to ascribe what the speaker meant rather than imposing one's own interpretation in order to avoid making a critical assessment of the message.

7) **Critical**

   It is defined as assessing what is being said, differentiating and understanding the message to make a decision regarding the communication in order to approve or disapprove of the arguments that are convincing.

D. Effective Techniques in Educating Listening

Listening is essential to learning a language. Through listening, students can engage in spoken language interaction and get the auditory input necessary for language learning. Teachers show students how to modify their listening strategies to meet a range of contexts, input kinds, and listening goals. Cognitive, metacognitive, and socio-affective strategies are the three categories into which listing strategies reside.

1) **Cognitive**
Learners employ cognitive strategies, which are problem-solving procedures, to manage learning activities and expedite the purchase of information or abilities. Cognitive methods entail the direct handling or alteration of instructional material in relation to a particular learning activity. Language learners process, store, and retrieve new knowledge with the aid of cognitive processes. Four cognitive techniques will be examined in the following manner:

1. When speakers attempt to understand the input task without translating, they employ the first cognitive method. As a result, this technique focuses the listener's focus on the organization and significance of the language being studied.

2. To comprehend the new words, the second cognitive method is to concentrate on the essential terms. By applying what they know about words in the intended languages to phrases, the listener gives meaning to the phrase's words. This method is particularly helpful for inexperienced listeners who rely on their limited vocabulary to improve their listening skills.

3. The third cognitive technique involves understanding the entire text by relying on the key idea. This tactic aids listeners in identifying the main idea first and the specifics subsequently. Reading is one of the tactics used in this approach. When a student employs this method, they discover the primary idea quickly and pick up on the aural wording immediately.

4. Scrambling to deduce the meaning based on hints is the fourth cognitive technique. This tactic is employed by listeners who are either unfamiliar with the entire phrase or do not comprehend all of the terms. When they haven't listened intently sufficiently or when the purpose of the phrase is unclear, both native and non-native speakers employ this technique.

2) **Metacognitive**

Learners can take charge of their education by using metacognitive approaches, which are methods of management that involve organizing, tracking, assessing, and changing their learning. In order to use metacognitive methods of planning, listeners can, for instance, make clear the goals of the listening assignment that they were expecting and focus on specific elements of the spoken language or contextual information that helped them understand the spoken word.

When engaging in class, viewing TV, listening to the radio, or conversing with others, this technique is quite helpful. Selecting ahead of time what to focus on is the second mental process tactic. Selective focus is a strategy used by listeners to speed up the process of understanding. For instance, in order to comprehend spoken language with various accents, some listeners decide to concentrate on grammar and accents. However, concentrating too much on accents might hinder understanding as they can be distracting and produce misunderstandings.

3) **Socio impacting**

The final type of method is called socio-affective, and it covers efforts to elicit and encourage favorable emotional responses and perspectives regarding language acquisition. Socio-affective tactics are methods that listeners use to reduce anxiety, confirm comprehension, or work with others. Considering student’s social-psychological factors that is, their feelings about the learning and the learning situation are closely linked, the affective tactics employed to manage learning experiences are crucial. Low anxiety was found to be significantly correlated with high listening efficiency, which implies that effective methods could be used to improve and facilitate listening. The learning environment was instantly impacted by the four management techniques: cognitive, social, emotional, and social strategies for listening comprehension.

IV. **METHODOLOGY**

In this section, we provide a comprehensive explanation of the dataset, both groups and the proposed technique. Figure 3 represents the flow of methodology.
A. Dataset

We utilized 120 signal datasets from the learners who have studied the English language since elementary school for this experiment and we equally divided the learners into two groups including baseline and study groups.

B. Baseline Group

In this group, there are 60 learners who have been trained through regular instruction only.

C. Study group

We use effective strategies and new techniques using AR-HMM for 60 learners in this group.

D. Proposed technique

1) Artificial Rabbit algorithm (ARA)

As a component of the diversion foraging strategy, the rabbit is compelled to consume the grasses near other people's nests, which can prevent attackers from discovering its nest. Furthermore, a rabbit may choose at random to hide in any of its own shelters by using the randomized hiding strategy, which may reduce the possibility that its enemies could capture.

Every iteration updates the location of each individual bunny by applying the rules of the provided algorithm, which is then assessed by the function of fitness. As the process proceeds, the approaches get more refined. Equation (1) assigns each beginning populated place to a random place within the searching space:

\[ Z_j = ka + [va - ka] \times \text{rand}(1, \text{dim}) j = 1, 2, \ldots, m \]

\[ Q_i(js + 1) = Z_i(js) + Y \times (Z_i(js) - Z_i(js)) + \text{round}(0.5 \times (0.05 + u_i)) \times \text{SND}, i = 1, \ldots, i \neq j \]
\[ Y = d \times K \]  
(3)

\[ d(l) = \begin{cases} 1 & \text{if } l = h(k) \\ 0 & \text{else} \end{cases} \quad l = 1, \ldots, \dim k = 1, \ldots, [u_2, \dim] \]  
(4)

\[ h = \text{randperm}(c), m_1 \sim M(0,1) \]  
(5)

\[ K = \sin(2\pi u_3) \times \left( f - f(\frac{j - 1}{\max})^2 \right) \]  
(6)

b) \textbf{Unpredictable Concealing}

A rabbit will often build multiple tunnels near its nest so that it has somewhere to hide when it needs to run from harm. The following equation is given in this regard.

\[ a_{ji}(js) = Z_j(js) + G \cdot H_Z_j(js), j = 1, \ldots, m \text{ and } i = 1, \ldots, \dim \]  
(7)

\[ G = \frac{S_{\max} + 1 - js}{S_{\max}} \cdot u_4 \]  
(8)

\[ H(l) = \begin{cases} 1 & \text{if } l = i \\ 0 & \text{else} \end{cases} \quad l = 1, \ldots, \dim \]  
(9)

In order to survive, rabbits need to find a safe location to hide. As a result, they get discouraged from choosing any old hole amongst the ones they must hide in to stay hidden. The random concealing technique can be expressed numerically as follows:

\[ q_j(js + 1) = Z_j(js) + Z \times (u_5 \times a_{ji}(s) - Z_j(js)) \quad j = 1, \ldots, m \]  
(11)

The location of the rabbit is updated in the following manner when either randomized concealment or detour feeding is effective:

\[ Z_j(js + 1) = \begin{cases} Z_j(js) & f(Z_j(js)) \leq f(Q_j(js + 1)) \\ Q_j(js + 1) & f(Z_j(js)) > f(Q_j(js + 1)) \end{cases} \]  
(12)

c) \textbf{Energy Diminishing}

An energy element is taken into account while modeling the change from the detour foraging-related discovery phase to the exploitation phase symbolized by randomized concealment. The electricity factor in the above formula is defined by the following terms:

\[ B(js) = 4 \left( 1 - \frac{js}{S_{\max}} \right) \ln \frac{1}{q} \]  
(13)

2) \textbf{HMM}

The probability Hidden Markov model originated from the Markov chain. Its primary purpose is to characterize the signal's statistical properties in a random process using various parameters. These days, investigation into topics connected to HMM-based voice recognition has grown very popular in this sector. We consider \( \{W(s), s \in S\} \) to be a stochastic process. It is said to be a Markov process if it meets the inefficiencies in both time and nation, that is, if the future is independent of the past.

A Markov chain is created for a process known as Markov if the time and state that characterize it are discrete. Our definition of a Markov chain is \( \{W_i = W_{i_1}, W_{i_2}, \ldots, W_{i_m}\} \). This chain must meet the Bayesian criteria.

\[ O = (W_1, W_2, \ldots, W_m) = O(W(s_1), j_1, W(s_2), j_2, \ldots, W(s_m), j_m) \]  
(14)

\[ O(W_j) \prod_{j=2}^{m} O(W_j|W_{j-1}) = O(W_j) \prod_{j=2}^{m} O(W_j|W_{j-1}) \]  
(15)

Among these, the time series \( W_i^{-1-j} = W_j, W_{j+1}, \ldots, W_{j-1} \) is represented by \( s_1, s_2, \ldots, s_m \). Equation (15) analyzes the Markov chain's observing sequencing \( X \) with the probability distribution rule from the viewpoint of time series. Transitions between various states can also be used to characterize the Markov chain when the time series \( s_1, s_2, \ldots, s_m \) is not taken into consideration. If the system's state is specified as \( s_t \) at time \( t \), then the likelihood of precisely transitioning from state \( i \) to state \( j \) is given by:

\[ b_{ji} = O(t_t = i|t_{t-1} = j) \quad (1 \leq j, i \leq M) \]  
(16)

Another way to think of it is as a transition to the same state when the system's state doesn't change. As a result, the total probability of a state transition at any given time should equal 1, meeting the following criteria:

\[ \sum_{i=1}^{M} b_{ji} = 1 \quad (1 \leq j, i \leq M) \]  
(17)
The state sequence is the collection of all states, and the underlying framework of the Markov chain is the individual connection between the data series and the state sequence. An HMM model requires the parameters that follow to be defined:

1. Sequencing of findings $O$ is equal to $P = \{p_1, p_2, \ldots, p_S\}$. Each finding sequence in speech signal analysis can be represented as a series of speech parameter values for features at the appropriate time, where $P$ is the amount of speech signal frames.

2. A collection of state sequence $T = \{t_1, t_2, \ldots, t_M\}$ make up $T$. It denotes the collection of states that the system consists of, $M$ is the total amount of states in the system, and every observation sequence is in one of those states.

3. $B$ is the state change probability distribution $B = \{b_{ij}\}$. It includes the collection of odds that every observation sequence's state will change. It is possible to ascertain that the sum of the numbers contained in every row of the information matrix is 1 by using Equation (16) and (17).

4. Set $A$ of output probabilities $A = \{a_i(p_s)\}$. When the equipment is in state $a_i(p_s)$ is a chance that the observing sequences will be output. If it is a discontinuous HMM (DHMM), it can be represented as a matrix; if it is a continuing HMM (CHMM), it can be represented as a density function of probabilities. The output chance of various sequences in various states can be represented, using the CHMM as an example, using the usual normal distribution density function with probability and its linear combinations. Here is the equation:

$$a_i(p_s) = \sum_{k=1}^{N} d_{ik} b_i(o_s, \mu_{ik}, V_{ik}) \quad (1 \leq i \leq M)$$ (18)

5. Probability set for the starting state, $\pi = \{\pi_j\}$. It depicts each state's initial distribution of chances, specifically:

$$\pi_j = O(t_1 = j), \quad (1 \leq i \leq M)$$

$$\lambda = (B, A, P, \pi)$$ (19) (20)

It is not possible to directly witness the state sequencing set $\{T = t_1, t_2, \ldots, t_m\}$ since it is hidden. As a result, we may identify an HMM using the equation that follows.

3) AR-HMM

Enhancing listening skills can be achieved through a hybrid strategy that utilizes the Artificial Rabbit Optimized Hidden Markov Model (AR-HMM). The probabilistic basis for AR-HMM is able to extract subtle patterns from audio input. By providing instantaneous correction and encouragement, AR-HMM greatly improves the learning process. It facilitates more successful language learning by encouraging involvement, more in-depth understanding, and lasting retention. Algorithm 1 represents the AR-HMM algorithm.

**Algorithm 1: Process of AR-HMM**

**Step 1:** def AROHMM\_optimization():
  initialize\_population()
  while not convergence:
    evaluate\_fitness()
    select\_parents()
    crossover()
    mutate()
  update\_population()

**Step 2:** def train\_AROHMM():
  audio\_data = load\_audio\_data()
  features = extract\_features(audio\_data)
  AROHMM\_optimization(features)

**Step 3:** def test\_AROHMM():
  test\_audio\_data = load\_test\_audio\_data()
  test\_features = extract\_features(test\_audio\_data)
  evaluate\_performance(test\_features)

**Step 4:** def apply\_AROHMM():
  new\_audio\_data = collect\_new\_audio\_data() new\_features = extract\_features(new\_audio\_data)
  improved\_skills = apply\_trained\_model(new\_features)

return improved\_skills
V. RESULT AND DISCUSSION

In this section, we distribute nine questions to the learners and evaluate our proposed technique with other techniques in post and pre testing based on accuracy, involvement, Effectiveness, and Transferability.

A. Data analysis

The nine questions were distributed to the learners to ask about their experiences. Questionnaires are shown in Table 1. All students answered the questions expressed in Figure 4 and Table 2.

Table 1: Questionnaires distributed to learners

<table>
<thead>
<tr>
<th>Questionnaire no</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I like listening</td>
</tr>
<tr>
<td>2</td>
<td>After the training, I feel confident in my current English listening skill</td>
</tr>
<tr>
<td>3</td>
<td>Improving my listening skills without this technique was challenging</td>
</tr>
<tr>
<td>4</td>
<td>I have noticed a significant improvement in my listening using the technique</td>
</tr>
<tr>
<td>5</td>
<td>This kind of technique is most important in English training class</td>
</tr>
<tr>
<td>6</td>
<td>The process of the method was helpful in enhancing my listening skill</td>
</tr>
<tr>
<td>7</td>
<td>This class improved my nature of listening skills and ability</td>
</tr>
<tr>
<td>8</td>
<td>The new technique generates curiosity about listening</td>
</tr>
<tr>
<td>9</td>
<td>I am satisfied with this class to improve my skill</td>
</tr>
</tbody>
</table>

Table 2: Outcomes of questionnaires

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Strongly Agree (%)</th>
<th>Agree (%)</th>
<th>Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>97</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>95</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>98</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>92</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>95</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>96</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>98</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4: Outcomes of questionnaires
B. Existing techniques

We evaluated the performance of our proposed technique and compared it with the existing techniques such as LSTM (Geng (2023)) and STT (Hirai and Kovalyova (2023)) [19, 20].

LSTM: The long short-term memory (LSTM) is a network-based speech recognition tool for English language instruction.

STT: The speech-to-text (STT) is a technology to has the functional ability for a machine to convert spoken language from audio inputs into text. Both are used to improve the listing training effect in foreign English language.

C. Parameters

1) Accuracy

The ability of accurately and exactly understand spoken words entails fully comprehending the circumstances, details, and meaning of what is being stated without confusion or miscommunication to improve listening training.

2) Involvement

It refers to the level of a person's engagement, focus, and active participation in a listening exercise. It describes specifying the factors that contribute to effective and engaged listening training.

3) Efficiency

It corresponds to the efficacy and profitability of the techniques or tactics used to improve an individual's ability to interpret and understand listening training.

4) Time

The amount of time it takes for someone to become skilled with technology and comprehend it can differ greatly based on a number of variables, including the technology's complexity, the person's learning style, their past knowledge and experience, and the materials available for learning.

D. Performance evaluation

We evaluate the performance of existing and proposed techniques in both pre and post testing.

1) Accuracy

Pre testing: The existing technique LSTM and STT attained 72% and 85%. When compared to the existing technique, our proposed technique AR-HMM achieved 91% accuracy.

Post testing: The existing technique LSTM and STT attained 80% and 92%. When compared to the existing technique, our proposed technique achieved AR-HMM 95% accuracy. Figure 5 and Table 3 display the performance of accuracy.

2) Involvement
Pre testing: The existing technique LSTM and STT attained low, medium, and high (22%, 51%, and 72%) and (20%, 49%, and 75%). When compared to the existing technique, our proposed technique AR-HMM achieved (18%, 52%, and 80%) involvement.

Post testing: The existing technique LSTM and STT attained low, medium, and high (20%, 55%, 80%) and (18%, 50%, and 89%). When compared to the existing technique, our proposed technique AR-HMM achieved (15%, 59%, and 93%) involvement. Figure 6 and Table 4 represent the evaluation of involvement.

Table 4: Percentage of Involvement

<table>
<thead>
<tr>
<th>Methods</th>
<th>Involvement (%)</th>
<th>Pre test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L M H L M H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>22 51 72 20 55 80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STT</td>
<td>20 49 75 18 50 89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR-HMM</td>
<td>18 52 80 15 59 93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: L- low, M- medium, and H- high

Figure 6: Evaluation of involvement

3) Efficiency

Pre testing: The existing technique LSTM and STT attained 72% and 69%. When compared to the existing technique, our proposed technique AR-HMM achieved 80% efficiency.

Post testing: The existing technique LSTM and STT attained 81% and 79%. When compared to the existing technique, our proposed technique AR-HMM achieved 90% efficiency. Figure 7 and Table 5 display the performance of efficiency.

Table 5: Values of efficiency

<table>
<thead>
<tr>
<th>Methods</th>
<th>Efficiency (%)</th>
<th>Pre test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>72 81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STT</td>
<td>69 79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR-HMM</td>
<td>80 90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7: Performance of efficiency
4) **Time**

**Pre testing:** The existing technique LSTM and STT attained 20hr and 22hr. When compared to the existing technique, our proposed technique AR-HMM achieved 19hr duration.

**Post testing:** The existing techniques LSTM and STT attained 18hr and 19hr. When compared to the existing technique, our proposed technique AR-HMM achieved 14hr duration. Figure 8 and Table 6 represent the evaluation of time.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Time (hrs)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>STT</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>AR-HMM</td>
<td>19</td>
<td>14</td>
</tr>
</tbody>
</table>

**Table 6: Outcome of Time**

![Figure 8: Evaluation of time](image)

Regarding these four parameters, our proposed technique is outperformance compared to the existing technique to improve listening skills in the foreign English language.

**VI. CONCLUSION**

The efficiency of listening instruction in foreign languages can be greatly increased by incorporating recognition of speech algorithms. This paper introduces the artificial rabbit optimized hidden markov model (AR-HMM) and a comprehensive explanation of listening. The speech signal dataset was collected from 120 learners and separated into two groups. After training, we distributed nine questions to the learners for analysis of the experience of the proposed technique. As a result, we evaluate the performance of the proposed technique with the existing technique in pre and post testing. The parameters included involvement (low (15%), medium (59%) and high (93%)), time (14hr), accuracy (95%) and efficiency (90%). It shows that our proposed technique outperformed others to improve listening training effect in foreign english language. Data collection issues, such as low response rates, inaccurate data, or insufficient datasets, may have an impact on how reliable the findings of the research appear. The speech recognition algorithms are prone to mis-interpretations of accents or dialects, leading to mistakes in voice transcription and interpretation. Methodological flaws that potentially affect the accuracy and dependability of the results include the use of self-reported data or the lack of control over unrelated factors. Future work may examine integration with other linguistic abilities, compare other speech recognition systems, and analyze long-term effects. It might also be investigated to adapt to another language.

**ACKNOWLEDGMENT**

The basic scientific research project of the Liaoning Provincial Department of Education: Research on the Meaning Generation Mechanism of X ЕСТЬ X Structure under the Framework of Cognitive Semantics, Project Number: LJKMR20220632.

**REFERENCES**


