

Research Article

Spoken English Assessment Using Confused Phoneme Assessment Model

Zhaoxia Jiang 

School of Foreign Languages, Yangtze Normal University, Chongqing 408100, China

Correspondence should be addressed to Zhaoxia Jiang; 20040033@yznu.edu.cn

Received 7 April 2021; Revised 23 April 2021; Accepted 18 May 2021; Published 29 May 2021

Academic Editor: Fazlullah Khan

Copyright © 2021 Zhaoxia Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under normal circumstances, in the process of evaluating the phoneme of spoken English which is easy to be confused, it is impossible to accurately assess the oral ability of different groups of people according to their oral ability. The assessment process has problems of poor robustness and low stability. We propose a spoken English assessment method based on an easily confused phoneme assessment model to address these problems. We design an English easily confused phoneme-based evaluation model in the proposed framework by adopting fuzzy logic for the assessment task. We also present HPD set for confused phonemes and introduce the easily confused phonemes in spoken English. Moreover, we derive four fuzzy measure assessment grades of E/G/NI/GR and present the assessment model for them. We continuously recognize and annotate spoken English to find the best-matched statement and complete the recognition and assessment of easily confused phoneme. Then, we also focus on spoken English assessment based on an easily confused phoneme assessment model. Empirical results demonstrate the superior performance of our proposed models over the conventional evaluation methods. Our proposed models improve the spoken English assessment method by 30% and the stability by 45%. Besides, our model is also suitable for the spoken English assessment of different groups of people.

1. Introduction

Conventional assessment of easily confused phoneme of spoken English cannot be accurately evaluated based on the different speaking ability of different groups of people. Its applicability is narrow, and its robustness and stability are low [1]. In this paper, we propose a spoken English assessment method based on an easily confused phoneme assessment model. We design a fuzzy logic μ -based Sugeno integral [2]. Furthermore, we integrate the Sugeno integral framework with a customized HDP set of confused phonemes. Then, our model proposes four kinds of fuzzy measure ratings (E/G/NI/GR) to evaluate the language score [3].

We design an easily confused phoneme evaluation model. For the assessment model with “simple word list grammar” syntax, we collect Chinese-based HDP and classify them into various HDP sets using Fourier transform and Mel cepstrum filtering. Each HDP set comprises phonemes that are not easily recognized by Chinese students [4].

The credibility of the assessment model to discriminate different HDP sets is based on the standard corpus, and the phoneme recognition results are integrated into the Sugeno integration framework [5]. Based on the algorithm of finding the maximum matched statement, we use liaison annotation and liaison recognition of spoken English and process the recognition of HDP in batches [6]. To ensure the effectiveness of the proposed method, the population test environment of spoken English is simulated. Two different assessment methods of spoken English are used for robustness and stability. Experimental results show that the proposed spoken English assessment method is highly effective [6].

2. Building Assessment Model of English Easily Confused Phoneme

2.1. Determining Fuzzy Logic of Assessment. In the real world, everything cannot be described by exact numbers. For example, there is no exact value for how many degrees of heat

and how much space is enough for the sea, for example, hot weather and the great sea. To model such a scenario, fuzzy logic was introduced [7]. Fuzzy logic is a mathematical method to describe an uncertain problem [8]. The fuzzy set is introduced as follows.

The definition of complement, inclusion, union, and the intersection of a fuzzy set is given below:

$$f_{A'} = 1 - f_A, \quad (1)$$

where A' represents a complement of fuzzy set A [9].

B includes A if and only if

$$\begin{aligned} f_A &\leq f_B, \\ A \subset B &\Leftrightarrow f_A \leq f_B, \\ f_C(x) &= \max[f_A(x), f_B(x)], \\ f_C(x) &= \min[f_A(x), f_B(x)]. \end{aligned} \quad (2)$$

Fuzzy integral is a crucial application in the field of fuzzy sets. In the existing fuzzy integral operation, Sugeno integral is the most popular integral operation [10]. We have given below a brief overview of some essential concepts used in the easily confused phoneme evaluation model.

Assume X is a non-empty finite feature set, and 2^X is its power set. Then, $\mu: 2^X \rightarrow [0, 1]$, $\mu(x)$ represents the reliability of the element $x \in 2^X$. $\mu(x)$ is the fuzzy measure over $(X, 2^X)$. Then, if it satisfies [11]:

$$\begin{aligned} \mu(\emptyset) &= 0, \\ \mu(X) &= 1. \end{aligned} \quad (3)$$

Similarly, for $A, B \subseteq X$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$. We can follow the definitions of the intersection, union, not operation described in [12] as follows:

$$\begin{aligned} \wedge: \mu(x) \wedge \mu(y) &= \min\{\mu(x), \mu(y)\}, \\ \vee: \mu(x) \vee \mu(y) &= \max\{\mu(x), \mu(y)\}, \\ \sim: \sim \mu(x) &= 1 - \mu(x). \end{aligned} \quad (4)$$

Let $X = \{x_1, x_2, \dots, x_n\}$, and μ be the fuzzy measures over $(X, 2^X)$, then the Sugeno integral of the function $f: X \rightarrow [0, 1]$ of the fuzzy degree μ is defined as follows:

$$S_\mu(f) = \sum_{i=1}^n \mu(A_i) \wedge f(x_i), \quad (5)$$

where $A_i \subseteq 2^X$, and $A_i = \{x \in X | f(x) \geq f(x_i)\}$.

2.2. Introduction of Easily Confused Phoneme in Spoken English. For various foreign language learners, a few phoneme sets are always indistinguishable. Each such set is called an HDP (phoneme that is hard to distinguish) set. For example, for most Indian speakers, it is complicated to distinguish between English phonemes /t/ and /d/. For Chinese English learners, it is not easy to distinguish between /w/ and /v/ in spoken English [13]. It is instrumental for language learners to improve their speaking skills and their ability to understand the foreign language if they can successfully master the pronunciation skills of different HDP sets. Providing accurate HDP

assessment results and feedback to language learners is also an essential requirement for speaking assessment. For different native speakers, HDP sets are usually different. In this paper, we discuss the problem of native Chinese English speakers. The established assessment model can also be applied to other non-native English speakers [14].

To reduce the error recognition rate and consider the difference between SR and language learning (LL), it is inappropriate to use the existing SR framework directly to identify which phonemes were pronounced by the language learner [15]. Therefore, it is implemented by other methods in this paper.

Figure 1 describes the HDP sets used in our model. To make the figure clear, we introduce two nodes. One is the “begin” node, which occurs before pronouncing the first word of the sentence. The other is the “end” node, which occurs after pronouncing the last word of the sentence. Since the assessment model has been provided with all the possible pronunciation before recognition, the actual pronunciation of the practitioner can be easily detected by the assessment model [16].

The HDP assessment task can be described as providing an HDP cluster script to the speaking practitioner, and then the script is recorded when the speaker reads these sentences [17]. Then:

- (1) The actual pronunciation of each HDP of the practitioner is annotated
- (2) According to the standard phonetic string found in the dictionary, the proportion of correctly identified HDP and erroneously identified HDP is statistically calculated
- (3) The language score and feedback were provided to the language practitioner

The most challenging problem of assessing pronunciation level based on the pronunciation of language learners is the instability of the speech processing evaluation model [18]. The HDP recognition results of the local recording of 1,032 sentences were statistically analyzed to illustrate this problem. Considering the native speaker’s pronunciation is usually correct, the native pronunciation corpus is considered the standard corpus in this paper. The recognition result of the SR assessment model is obtained based on the recognition result of the standard corpus. Table 1 shows the statistical results. The meanings of the symbols in Table 1 are as follows:

- (i) P represents the phoneme in the corpus set
- (ii) q represents the actual recognition result for the phoneme P assessment model
- (iii) n represents the number of different recognition results of q
- (iv) nt represents the number of occurrences of the phoneme P in the corpus

2.3. Determining a Fuzzy Measure of an Easily Confused Phoneme Assessment Model. In order to evaluate the reliability of the assessment model, two relative measures are

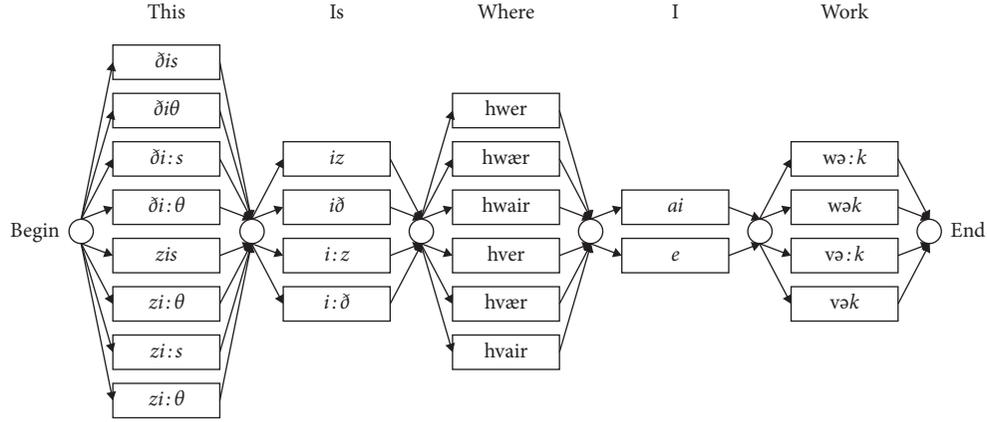


FIGURE 1: HDP sets of the proposed model.

TABLE 1: HDP recognition results based on standard corpus Sphinx-4.

<i>P</i>	<i>Q</i>	<i>n</i>	<i>nt</i>	<i>P</i>	<i>q</i>	<i>n</i>	<i>Nt</i>	<i>P</i>	<i>q</i>	<i>n</i>	<i>nt</i>
<i>E</i>	<i>E</i>	318		<i>i:</i>	<i>i:</i>	455		ɔ	ɔ	234	
	<i>Æ</i>	169	535	<i>i:</i>	<i>ei</i>	80	584	ɔ	<i>au</i>	44	330
	<i>Ai</i>	48			<i>i</i>	49			ɔ:	52	
<i>Au</i>	<i>Au</i>	186		ɔ:	ɔ:	145		æ	æ	398	
	ɔ:	1	197	ɔ:	ɔ	139	322		<i>e</i>	243	682
	ɔ	10			<i>au</i>	38		ə:	ə:	41	
<i>Ei</i>	<i>Ei</i>	272		η	η	146	163	θ	θ	74	94
	<i>i:</i>	1	286		<i>n</i>	17			<i>s</i>	20	
	ʌ	13		<i>ai</i>	<i>ai</i>	647	671	ð	ð	344	385
<i>u:</i>	<i>u:</i>	435		<i>u</i>	<i>e</i>	27	94	<i>z</i>	<i>z</i>	41	422
	<i>U</i>	13	448	<i>u</i>	<i>u</i>	77			<i>z</i>	400	
<i>ə:</i>	ə:	110		<i>v</i>	<i>u:</i>	17	337	<i>n</i>	ð	22	1392
	<i>Æ</i>	284	394		<i>v</i>	324			<i>n</i>	1248	
<i>I</i>	<i>I</i>	1028		η	<i>w</i>	13	163	<i>w</i>	η	144	491
	<i>i:</i>	74	1102		η	146			<i>w</i>	471	
ʌ	ʌ	1516		—	<i>n</i>	17			<i>v</i>	20	
	<i>Ei</i>	80	1596		—				—		

introduced: the correct recognition rate r and the false recognition rate e_i^j that are defined as follows:

$$\begin{cases} r = \frac{N_{\text{right}}}{(N_{\text{right}} + N_{\text{error}})}, \\ e_i^j = \frac{N_c(i, j)}{n_i(i)}, \end{cases} \quad (6)$$

where N_{right} is the number of phonemes that are correctly recognized, and N_{error} is the number of phonemes that are erroneously recognized. The function $N_c(i, j)$ is the number of phonemes being identified as the j^{th} , but actually being the i^{th} phonemes, $i \neq j$. For example, in Table 1, $N_c(w, v) = 20$, then:

$$e_w^v = \frac{N_c(w, v)}{n_i(w)} = \frac{20}{491} = 0.04. \quad (7)$$

The HDP set is an attribute set $A = \{x_1, x_2, L, x_{10}\}$. x_1 represents the phoneme $/i:/$, and other placeholders can be derived. The fuzzy measure will only depend on the potential of an attribute set. The HDP assessment is still using a fuzzy approach. There are four assessment levels, which are excellent, good, medium, and need to be improved (NI). The actual meaning of fuzzy measure is that a speech instance belongs to or is better than some assessment level. The definitions are given as follows:

- (i) Fuzzy measure of “belong to or better than ‘medium’”:

$$\mu(A) \begin{cases} e^{-((p(A)-c)^2/2\sigma^2)} & 1 \leq |A| \leq L \text{ and } p(|A|)^2 < c, \\ 1.0, & p(|A|) \geq cL < |A|, \end{cases} \quad (8)$$

where L is the length of the HDP cluster, and A is the subset of attribute set composed of the HDP placeholders.

(ii) Fuzzy measure of “belong to or better than ‘good’”:

$$\mu(A) \begin{cases} e^{-((p^{(1A)}-c)^2/\sigma^2)} & 1 \leq |A| \leq L \text{ and } p(|A|) < c, \\ 1.0, & p(|A|) \geq cL < A|. \end{cases} \quad (9)$$

(iii) Fuzzy measure of “belong to ‘excellent’”:

$$\mu(A) \begin{cases} e^{((p^{(A)}-c)^2/2\sigma^2)} & 1 \leq |A| \leq L \text{ and } p(|A|) < c, \\ 1.0, & p(|A|) \geq cL < A|. \end{cases} \quad (10)$$

$$\left\{ \begin{array}{l} \text{for Excellent, } c = 1.0, \sigma = 0.15 \\ \text{for Good, } c = 0.9, \sigma = 0.45 \\ \text{for Medium, } c = 0.7, \sigma = 0.40 \end{array} \right\}.$$

Fuzzy measures of medium, good, and excellent are shown in Figure 2.

2.4. Completing the Building of the Evaluation Model. In order to obtain robust assessment results, we consider the credibility of different speech-processing evaluation models using two measures: r and e_i^j . For HDP assessment, these two cases must be considered independently.

- (1) If the phoneme x is correctly identified, the credibility of the phoneme is defined as:

$$f(x) = r(x). \quad (11)$$

- (2) When the phoneme x is recognized as i^{th} , the phonemex belongs to the same HDP set, rather than the x itself. The credibility of the phoneme is defined as:

$$f(x) = e_x^i. \quad (12)$$

3. Easily Confused Phoneme Assessment Model

Because of the recording module for the test set in this paper, the configuration of the front end of the evaluation model is the recording mode recognition [19]. The overall architecture of the assessment model is shown in Figure 3. The subprocessing units of pre-enhancement, windowing, Fourier transform, Mel cepstrum filtering, discrete cosine transform, batch CMN, and feature extraction were used for speech signal processing.

The assessment model is required to change the system’s syntax at any time, so it is necessary to improve its syntax construction to accept the new syntax dynamically. Thus, we adopt and improve “simple word list grammar” and add the construction method of the search tree with the grammatical sentence [20].

3.1. Annotation and Recognition of Liaison in Spoken English. To enable Sphinx-4 to recognize liaison, it is first necessary to accept the grammar of liaison and then build a search grid according to the input syntax node. The liaison annotation module or the HDP annotation module completes the expansion of the syntax node. As the assessment model does not know the input syntax, the vocabulary size of the assessment model and the acoustic model library should be large enough. The used dictionary type is FullDictionary as it contains more detailed phoneme information of words. WSJ-8gau-13dCep-16k_40me1-130Hz_6800Hz.jar is used as the acoustic model library of the assessment model. The aim is to make the acoustic model enough so that the assessment model can find the desired acoustic model [15].

The implementation method is to increase the ability to generate new syntax nodes dynamically based on the original “simple word list grammar” class so that the assessment model can accept the syntax nodes processed by the liaison rules. Then Sphinx-4 system is used to recognize the corresponding results for the input speech. Since the assessment of liaison is carried out, the recognition result needs to be compared with the statement processed by the liaison rules to give an assessment result. For liaison, the core module is liaison annotation and recognition result analysis.

The first is the liaison annotation. The model’s primary function is that for a given syntax statement, all the possible liaisons have to be annotated in the grammatical text according to the existing liaison rules. Moreover, all possible liaison extensions have to be generated, and the new synthetic words have to be added to the dictionary of the assessment model. In this evaluation model, this function is implemented by “liaison marker” and “liaison rule” classes. The liaison rule is stored in a hash table in the liaison rule class. The rule of storage is that two liaison phonemes are underlined and used as key values. In the hash table, the key value is the pronunciation of the two phonemes after liaison.

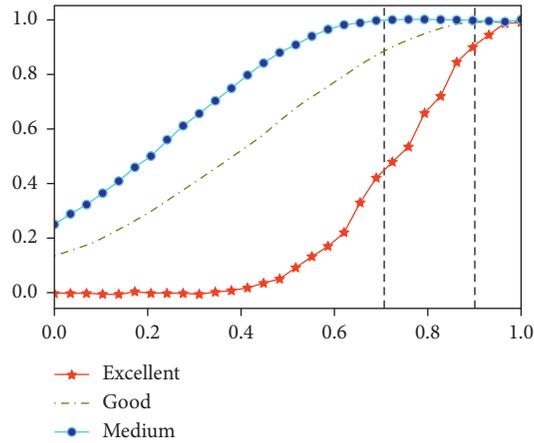


FIGURE 2: Fuzzy measures of medium, good, and excellent.

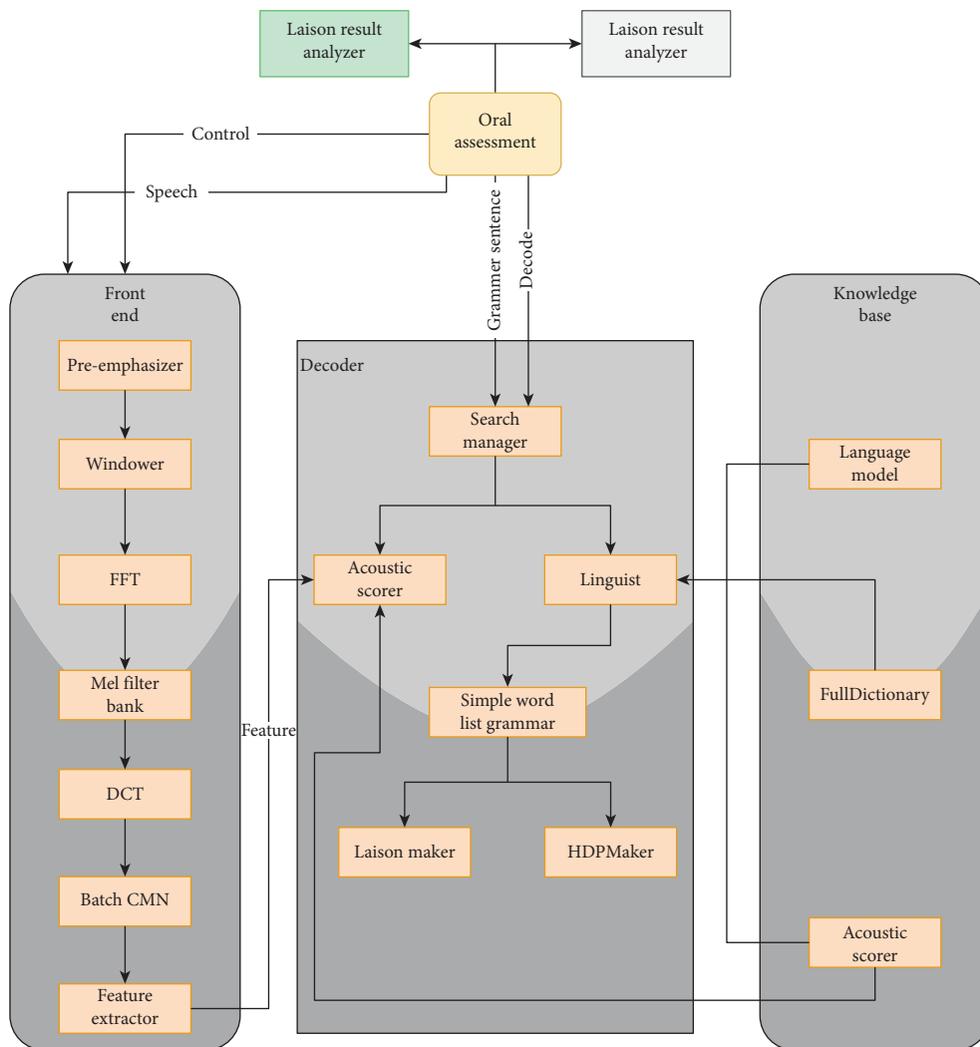


FIGURE 3: Structure of speaking assessment model.

For example, there is an entry “KAH K AH” in the file of liaison rules. After the hash table is loaded, the corresponding entry is the key value of “K_AH”. In this manner, the liaison rules can be used. For an input statement, liaison

marker first uses the dictionary in the assessment model to detect each word’s phonetic symbols and then process every two adjacent words in sequence according to the liaison rules. If it confirms the liaison rules, the first word is

annotated as the liaison available for all the liaison possibility extensions and new word extensions. Then the annotated sentence is extended, extending the statement to a set of statements that can contain all the possibilities of liaisons.

The second is the dynamic addition of words. To make the assessment model able to detect and identify all the possibilities of liaison, we need to add these liaisons to the knowledge base of the assessment model. If two words, such as “link” and “up”, are adjacent in the input statement. By examining the dictionary of the assessment model, the last phoneme of the first word is /K/, and the first phoneme of the last word is /AH/. By looking at the hash table of the liaison rules, the key value is “K AH”, and the corresponding value is “K AH”, that is, two phonemes can be linked to read, and the pronunciation after liaison is “K_AH”. In this way, a new word is formed and added to the dictionary. According to the “link” and “up” entries “LINK LIH1 NG K” and “UP AH1 P” in the original dictionary, they are merged into a new word, “link_up”, and its phonetic symbol is “L IH1 NG K AH1 P”. The acoustic statistical model corresponding to each phoneme is then associated with the phonetic symbol. In this way, new words are added dynamically to the assessment model. If the assessment model identifies the input speech feature frames, the most matched feature frames of the newly synthesized words are calculated, which means that the assessment model can recognize the liaison.

The last is the analysis of the recognition results. Although the recognition rate of the Sphinx-4 system for the speech signal with a small vocabulary is very high, the error is unavoidable. For example, to recognize the recording of the sentence “that should be good enough for us”, the result is probably “that that should be good enough_for us”, which cannot be evaluated. We need to normalize it and find out the sentence in all liaison extension sentences that matched the recognition result as the final recognition result. The algorithm to find the maximum matched statement is as follows:

- (1) Obtain the recognition results and all liaison extension sentences list (allLiaisonSentence (string list)).
- (2) Set the maximum length, Max Len = 1, and matching result string (result Sentence (string)) to null.
- (3) Take the unprocessed one-sentence extension in allLiaisonSentence and store it in standard Sentence (string).
- (4) Compare recognized result and standard Sentence. Find the location of each word or the liaison cluster in the standard Sentence from the recognized result and store it into the integer array match.
- (5) Find the most extended monotonically increasing sequence in the match and denote the length as len. The elements that do not appear in this sequence are set to -1.
- (6) If len is greater than Max Len, then Max Len is set to len, and result Sentence is set to standard Sentence.
- (7) Add the mark of having been processed to the statement in allLiaisonSentence.
- (8) If all the statements in allLiaisonSentence are processed, turn to step (9). Otherwise, turn to step (3).
- (9) Output result Sentence. The algorithm ends.

At the same time, the assessment model also increases the rejection function. When the number of words in the recognition statement is less than 60% of the sentence, the assessment model refuses to recognize it.

After finding a maximum matched statement, it is compared with a statement that connects all the segments. These segments can be read together to see whether each liaison and its type are recognized. Then the final assessment result is obtained based on the Sugeno integral speech assessment algorithm.

An example of an assessment is given, and local recording is used. The recording text is “that should be good enough for us”. The recording format is PCM_SIGNED, 16000.0 Hz, 16bit, mono, little-endian. The input recording text works as a grammar statement and then generates the syntax node according to the liaison rules. That is, {that, should, be, good, enough, for, us, that_should_be, that_should_be, good_enough, enough_for, for us, good_enough_for, enough_for_us, good_enough_for_us}. The original 7 grammar words are extended to 16 grammar nodes according to the liaison rules. The new compound word is then added into the dictionary of the assessment model to produce all possible methods of liaison pronunciation. Each liaison pronunciation is connected from a grammatical node on a possible path from the “begin” node to the “end” node. Space separates each node. There are 32 ways of pronunciation. Then the assessment model builds up a search grid based on syntactic nodes and identifies the input recording. The output recognition result is “that should be good_enough_for us”. The recognition result is compared with 32 known ways of pronunciation, and one of the best matching results is found. To make the match more accurate, the number and position are used as the matching object. For “that should be good_enough_for us”, the shortest path of this syntax consists of two liaison groups of “that should be” and “good_enough_for us”, with the length of 2 and 3. The first group $x = \{x_1, x_2\}$ is analyzer, where x_1 is the CC liaison type, and x_2 is the CF liaison type. The actual result of recognition is that two places are not linked to being read. According to the reliability building method, $f(x_1, x_2) = 0.7$. After the Sugeno integral of this liaison group, the assessment result of the liaison group is “good”. The assessment result of the second liaison group is “excellent”.

3.2. Recognition and Assessment of Easily Confused Phoneme. Sphinx-4 system is unable to identify easily confused phoneme, and it also needs to improve its grammatical structure to recognize it. The implementation of the assessment model is similar to that of liaison recognition, but the details are different. The main classes are HDPMarker and HDPRule.

First, new words and the form of new words are added to the dictionary of the assessment model. To carry out the assessment, the assessment model must be made to

determine which phoneme the spoken language practitioner has made and then compare the identified phoneme with the standard phoneme.

Unlike extensions of liaison, the unit of expansion here is limited to the phonemes of each word. For each phoneme, by looking up all the easily confused phonemes and then replacing them one by one, all the possible pronunciations can be obtained to make up a new word. For example, in “this is where I work”, for “this”, its phonemes in the Sphinx-4 dictionary are “DH IH S”, while the HDP set has the rules “DH Z”, “IHIY”, and “S TH”, so that the pronunciation of the word can be expanded to 8. Then, the phonetic symbols of each newly extended word are linked to the acoustic model and added to the dictionary of the assessment model. In this manner, if the recognition result of the input speech is an extended new word, the assessment model can detect whether the speaker is wrong and then give it an evaluation result.

The last is the analysis of the recognition results. The recognition results of the assessment model also need to be normalized. Its standard algorithm is the same as the maximum matching algorithm of liaison. If the number of words matching is less than 60%, the assessment model refuses to recognize. The difference is that the recognition result is only compared with the standard phonetic string. For judging whether the two words are the same or not, it is not to determine whether the spelling is the same and consider the easily confused phoneme set. The phonemes in the same HDP can be regarded as the same phoneme. Another difference is that the assessment of confusing phoneme is based on the HDP cluster. After all the sentences in the cluster have completed the maximum matching, then the assessment of the liaison cluster is carried out, and an assessment result based on linguistic variables is given.

- (i) That should be good enough for us.
- (ii) That’s the pleasantest part of it.
- (iii) She’s his sister.
- (iv) He has five thousand pounds a year.
- (v) I just had to come in and tell you the news.

An HDP cluster given in the statement corresponds to the HDP set of $\{/i:/i/\}$. Since the cluster assessment involves five statements, the assessment model needs to be processed one by one. Take “she’s his sister” as an example to illustrate how the assessment model processes every sentence in the HDP cluster. First, the phonetic symbols of all the words in the sentence are found out in the dictionary, and then the phonemes are connected with an underline to form a new spelling form. The original phonetic symbols are added to the new synthetic words to form a new dictionary entry and add to the dictionary. Then the word of the grammar statement is expanded based on the HDP set. Taking “his” as an example, it can be expanded to phoneme mode: $\{f\text{ HH_IH, HH_IY, HH_IH DH, HH_IY_DH}\}$. Other words are expanded following this rule.

Finally, using all the extended syntax nodes, the syntax search tree is established; the input recordings are identified;

and the recognition results are obtained. For the recognition result, it still needs to find its maximum matching. “HHIY DH” and “HHIH Z” will be considered as a word by the assessment model because they originate from the extension of the same word “his”. After finding the maximum matched statement, the actual pronunciation of the phonemes in the HDP set is compared with the standard statement and the recognition statement. After the assessment model has finished all the statements in the HDP cluster, a set of credibility with a prescribed length is obtained, and then Sugeno integral of the set is carried out to obtain the assessment result.

4. Experimental Analysis

In order to verify the effectiveness of the proposed method, simulation analysis is carried out. A group of people with different spoken English abilities is selected to conduct robustness and stability simulation experiments. A simulation study is conducted on the different gender and academic qualifications of the spoken English ability group. The results of simulation experiments are compared with the conventional assessment method. According to the given assessment model of liaison and easily confused phoneme in spoken English, experiments on the assessment of easily confused phoneme are performed to verify the effectiveness of the assessment model. The HDP model training experiment and the validation of the assessment model are described, and the results are analyzed.

4.1. Robust Liaison Experiment and Result Analysis. The recording format of the corpus used in this experiment is PCM_SIGNED, 16000.0 Hz, 16bit, mono, little-endian. The raining corpus T1 consists of 1,032 native spoken language recordings and their scripts. The test corpus with 100 native spoken natural language recordings is denoted as T2.

The liaison experiments include model training and model validation. The former is used to train the fuzzy measure of the assessment model and evaluate the credibility. The latter is used to verify the validity of the proposed model.

The training corpus T1 is used to obtain the credibility of the assessment model. The performance of the Sugeno integral is affected by determining the fuzzy measure and evaluating the credibility of the model. T1 is also used for the closed test. T2, T3, and TN are used for development tests. First, model training is introduced, and the process of training is as follows:

- (1) Linguistics experts annotate the liaison in the corpus T1, which is taken as standard liaison annotation
- (2) All the speeches in T1 are batch-processed to obtain the results after the normalization
- (3) The results of recognition and artificial annotation are compared to obtain the number of different liaison combinations

- (4) The results of recognition and artificial annotation are compared to obtain the number of correct, false, or missing identifications of different liaison types

Tests comprise the open test and the closed test. For the open test, the language material other than the training corpus is selected to test the assessment model. For the closed test, the training corpus is used to test the assessment model.

The results of the liaison assessment model based on Sugeno integrals are shown in Table 2. The model test results are shown in Figure 4.

In Table 2, GR is the ratio of “good”, defined as $(G + E)/$ liaison groups. From Table 2, it can be seen that for the closed test T1, the “good” and “excellent” of the model output reach 78% of the total. For the open test T2, the ratio is 76%. The results show that the liaison assessment model has good robustness. The open test T3 is the same as the T2 test. The conventional evaluation method has low comprehensive performance, and the “good” and “excellent” outputs of the assessment model are only 45%. The robust of the proposed method is improved by 30%.

4.2. Stability HDP Experiment and Result Analysis. The training corpus T1 consists of 1,032 sentences of native natural speaking recording and script. The test corpus is composed of 122 HDP clusters of native natural speaking recording, denoted as T2. Two students were chosen: one is graduated in English major, and the other is graduated in computer science. Their speaking ability is different from each other. The recordings of the same corpus of the students are denoted as T3 and T4, respectively.

HDP experiments also include model training tests and model validation tests. The corpus T1 is used for the model training test. T2, T3, T4, and TN are used for model validation tests.

First, the model training experiment is described as follows:

- (1) For the given recording script and HDP set, the HDP assessment model is used to identify the corresponding recordings and obtain the recognition results
- (2) By comparing the corresponding phonemes in recognition strings and annotation strings, the statistical data of the recognition results of different phonemes are obtained

Then the model verification experiment is given. T2, T3, and T4 are used for development tests. Considering the significant difference in speaking proficiency of different people, two spoken English assessment methods are applied. The assessment results of the corpus set T2–T4 are shown in Table 3. The precision validation under different data sets and the different parameter values are shown in Figure 5.

Table 3 shows the speech evaluation results of T2, T3, T4, TA (the proposed method), and TN (conventional assessment method). From Table 3, it can be seen that the assessment results of TA agree with the expected results. The

TABLE 2: Results of the liaison assessment model based on Sugeno integral.

Corpus	Liaison groups	Results (E)	Results (G)	Results (NI)	Results (GR)
T1	1,483	34	1,125	324	0.78
T2	151	6	109	36	0.76
T3	151	6	108	36	0.75
TN	41	3	85	14	0.45

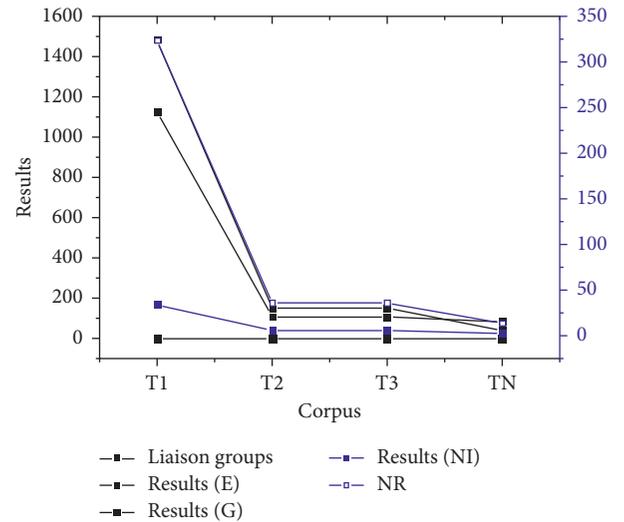


FIGURE 4: Model test results.

TABLE 3: Assessment results of the corpus sets T2–T4.

	T2 (%)	T3 (%)	T4 (%)	TA (%)	TN (%)
Excellent	54	19	5	18/5	15/4
Good	37	53	50	53/50	45/35
Medium	7	14	25	14/25	7/30
NI	2	14	20	13/19	12/8

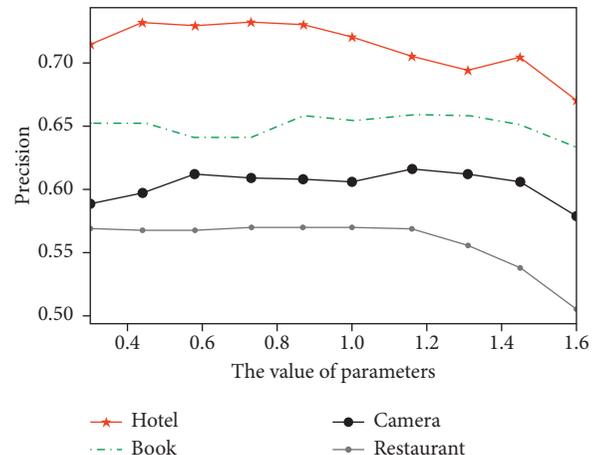


FIGURE 5: The precision validation under different data set and different parameter value.

evaluation results of TN deviate from the predetermined results. By weighted arithmetic analysis, the reliability of the proposed method is increased by 45%.

In this section, the experiment of testing liaison and HDP in spoken English is given. The former used development testing and closed testing. The experimental results show that the proposed spoken English assessment method has high robustness. The latter mainly tests the stability of the algorithm. Experimental results show that the proposed spoken English assessment method is highly reliable.

5. Conclusions

This paper proposes a spoken English assessment method based on an easily confused phoneme assessment model. We design and implement the English easily confused phoneme assessment model, presenting the assessment model's configuration and the related recognition results. Experimental results show that the proposed method is very effective. The research in this paper can provide a theoretical basis for the assessment of easily confused phonemes of spoken English.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This research was financially supported by Chongqing Educational Science Planning Project (No. 2020-GX-325), the Humanities and Social Sciences Research Foundation of Chongqing Municipal Education Commission (No. 18SKH142), and the Fundamental Research Funds for Yangtze Normal University (No. 2016XJQNO3).

References

- [1] H. Dong, "Modeling and simulation of English speech rationality optimization recognition based on improved particle filter algorithm," *Complexity*, vol. 2020, Article ID 6053129, 10 pages, 2020.
- [2] Z. Li, K.K. Sun, and S. Young, "The risk assessment method and management strategy of distribution network cable in Macau," in *Proceedings of the 2016 China International Conference on Electricity Distribution (CICED)*, pp. 1–5, IEEE, Xi'an, China, 2016.
- [3] Y. Liu, L. Qi, C. Lv, M. Zheng, and X. Ji, "A study on objective evaluation of vehicle steering comfort based on driver's electromyogram and movement trajectory," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 1, pp. 41–49, 2017.
- [4] C. Fenlon, L. O'Grady, S. Butler, M.L. Doherty, and J. Dunnion, "The creation and evaluation of a model to simulate the probability of conception in seasonal-calving pasture-based dairy heifers," *Irish Veterinary Journal*, vol. 70, no. 1, pp. 1–11, 2017.
- [5] Y. Lv and Z. Sun, "An evaluation of the influence of measurement geometry on the uncertainties of photometric model results based on the laboratory measurements of particulate surfaces," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 187, pp. 338–357, 2017.
- [6] A. Bonfante, S. M. Alfieri, R. Albrizio et al., "Evaluation of the effects of future climate change on grape quality through a physically based model application: a case study for the Aglianico grapevine in Campania region, Italy," *Agricultural Systems*, vol. 152, pp. 100–109, 2017.
- [7] J. E. Till, H. L. Beck, H. A. Grogan, and E. A. Caffrey, "A review of dosimetry used in epidemiological studies considered to evaluate the linear no-threshold (LNT) dose-response model for radiation protection," *International Journal of Radiation Biology*, vol. 93, no. 10, pp. 1128–1144, 2017.
- [8] Y. Ozluk, Y. Caliskan, M. Sevinc et al., "Re-evaluation of glomerulitis using occlusion criteria based on the Banff 2013 revision: a retrospective study," *Transplant International*, vol. 30, no. 6, pp. 579–588, 2017.
- [9] M. Perumal, G. Tayfur, C. M. Rao, and G. Gurarslan, "Evaluation of a physically based quasi-linear and a conceptually based nonlinear Muskingum methods," *Journal of Hydrology*, vol. 550, pp. 740–742, 2017.
- [10] M. M. Rahman, D. W. Lamb, J. N. Stanley, and M. G. Trotter, "Use of proximal sensors to evaluate at the sub-paddock scale a pasture growth-rate model based on light-use efficiency," *Crop and Pasture Science*, vol. 65, no. 4, pp. 400–409, 2014.
- [11] S.-K. Kim, M.-I. Roh, and K.-S. Kim, "Evaluation of feasibility index in the arrangement design of an offshore topside based on the automatic transformation of experts' knowledge and the fuzzy logic," *Ocean Engineering*, vol. 130, pp. 284–299, 2017.
- [12] S. Korol, M. White, E. O'Meara et al., "An evaluation of the effect of spironolactone on the risk of new-onset diabetes in a population-based study of patients with heart failure," *Canadian Journal of Cardiology*, vol. 32, no. 10, p. S90, 2016.
- [13] L. Corich, M. Buseti, V. Petix, S. Passamonti, and A. Bandiera, "Evaluation of a biomimetic 3D substrate based on the Human Elastin-like Polypeptides (HELPS) model system for elastolytic activity detection," *Journal of Biotechnology*, vol. 255, pp. 57–65, 2017.
- [14] L. Pazzagli, A. Möllsten, and I. Waernbaum, "Marginal structural model to evaluate the joint effect of socioeconomic exposures on the risk of developing end-stage renal disease in patients with type 1 diabetes: a longitudinal study based on data from the Swedish Childhood Diabetes Study Group," *Annals of Epidemiology*, vol. 27, no. 8, pp. 479–484, 2017.
- [15] P. Tappenden, S. Ren, R. Archer et al., "A model-based economic evaluation of biologic and non-biologic options for the treatment of adults with moderately-to-severely active ulcerative colitis after the failure of conventional therapy," *Pharmacoeconomics*, vol. 34, no. 10, pp. 1023–1038, 2016.
- [16] N.B. Rached, A. Kammoun, M.-S. Alouini, and R. Tempone, "A unified moment-based approach for the evaluation of the outage probability with noise and interference," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 1012–1023, 2016.
- [17] C. Sun, W. Wei, H. Su, X. Zou, and X. Wang, "Evaluation of sn-2 fatty acid composition in commercial infant formulas on the Chinese market: a comparative study based on fat source and stage," *Food Chemistry*, vol. 242, pp. 29–36, 2018.
- [18] L. Molander, A. Hanberg, C. Rudén, M. Ågerstrand, and A. Beronius, "Combining web-based tools for transparent evaluation of data for risk assessment: developmental effects

- of bisphenol A on the mammary gland as a case study,” *Journal of Applied Toxicology*, vol. 37, no. 3, pp. 319–330, 2017.
- [19] M. Wu, G. Shu, R. Chen, H. Tian, X. Wang, and Y. Wang, “A new model based on adiabatic flame temperature for evaluation of the upper flammable limit of alkane-air-CO₂ mixtures,” *Journal of Hazardous Materials*, vol. 344, pp. 450–457, 2018.
- [20] S. Piao, T. Okura, and M. Irie, “On-site evaluation of Wagyu beef carcasses based on the monounsaturated, oleic, and saturated fatty acid composition using a handheld fiber-optic near-infrared spectrometer,” *Meat Science*, vol. 137, pp. 258–264, 2018.