Fuzzy Systems and Data Mining X
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA241399

A Fuzzy Measurement-Based Algorithm for Spoken English Evaluation

Jinwei Dong[1], Ziting He , Hao Chen, Jinchun Zhang and Limin Zhi
Guangdong University of Foreign Studies, Guangzhou, China
ORCiD ID: Ziting He https://orcid.org/0009-0009-3795-4741,
Hao Chen https://orcid.org/0009-0004-2883-4178

Abstract. This study explored an improved English language evaluation algorithm for virtual learning environments, particularly in metaverse educational settings. By integrating big data analysis with fuzzy measure theory, the model extracted learners' language usage habits, progress, and difficulties from large datasets, overcoming the limitations of traditional subjective evaluation methods. Fuzzy measures quantified subjective factors such as clarity, fluency and intonation, while a Sugeno integral approach combined these measures into an overall score. Comparisons with traditional methods have shown significant improvements in the assessment of speaking skills across a range of proficiency levels.

Keywords. English Speaking Evaluation, Metaverse Learning, Fuzzy Measure

1. Background Information

The digital transformation of education has given rise to the Metaverse, an emerging virtual reality space that has the potential to revolutionise the field of education, particularly in the context of English language learning. Among the various forms of English learning available in the Metaverse, oral English learning is particularly prominent. The evaluation of spoken English based on a big data analysis differs from the traditional evaluation of spoken English. In the process of determining the weights of the indicators, the machine learning algorithm can employ a purely mathematical approach to calculate the degree of importance of the indicators that are susceptible to human experience. Consequently, the impact of subjective factors on the evaluation results is minimal, thereby effectively addressing the problem of evaluation accuracy.

2. Current Research at Home and Abroad

In recent years, scholars at home and abroad have explored the potential applications of meta-universe in language learning. This has included the proposal of a meta-universe learning system [1], the creation of experiential learning in meta-universe environments [2], and the establishment of campus meta-universe models. These models emphasise

[1] Corresponding Author: Jinwei Dong, School of English Education, Guangdong University of Foreign Studies, China; Email: 200110581@oamail.gdufs.edu.cn.
This study is financially supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2024.

the potential benefits of meta-universe applications in enhancing teaching effectiveness and promoting pedagogical reform. In terms of spoken English evaluation, research has concentrated on utilising intelligent speech recognition technology, natural language processing and machine learning algorithms to automatically assess pronunciation accuracy, fluency, vocabulary utilisation and grammatical structure. For instance, studies have been conducted on intelligent assessment algorithms for spoken English pronunciation quality [3], a spoken language assessment system capable of providing real-time feedback [4] and a system for correcting and improving spoken pronunciation [5]. The implementation of these studies on English-speaking learning platforms has shown preliminary outcomes. However, more precise and individualized speaking assessment models can be developed through the integration of big data analysis with learners' linguistic habits and learning requirements.

3. English Speaking Evaluation Algorithm

Despite the automation and intelligence that have been incorporated into the existing spoken English evaluation system, it still needs improvement in terms of accuracy and personalisation. In order to further align with learners' language preferences and learning objectives, it is necessary to integrate big data analysis technology to enhance and refine the spoken English evaluation algorithm.

3.1. Features of Spoken English and Speech Recognition Technology

Some features of spoken English include liaison and pronunciation confusion. As a phonological phenomenon, Liaison occurs when, according to specific phonological rules, the final sound of a word blends seamlessly with the beginning sound of the word immediately following it. This smooth transition in pronunciation gives the listener the impression that the two words share a syllabic boundary. Pronunciation confusion, on the other hand, arises when the pronunciations of specific words become so similar as to be indistinguishable. The evaluation of speech typically relies on continuous speech recognition technology, which is scored by evaluating criteria such as pronunciation accuracy, speech similarity, speech rate and duration. Moreover, a posteriori probabilities are employed to evaluate mathematical models to ensure the accuracy of the assessment results. The process of automatic speech recognition typically comprises three principal stages, as illustrated in Figure 1.

The initial stage is the construction of the model library [6], which establishes sound databases in accordance with the pronunciation characteristics of natural language. This enables the system to utilise these pronunciation models to perform matching searches. The second stage is feature extraction, which encompasses the collection of acoustic signals, the conversion of human pronunciation into a format that can be understood by the machine, and the elimination of background noise. The final stage is information matching, whereby the extracted speech features are fed into a decoder. The computer will then search for the most appropriate matches in the sound database based on the decoding results. The most similar matches are the result of computer recognition. The assessment of spoken English pronunciation represents a fundamental prerequisite for the successful implementation of automatic speech recognition technology.

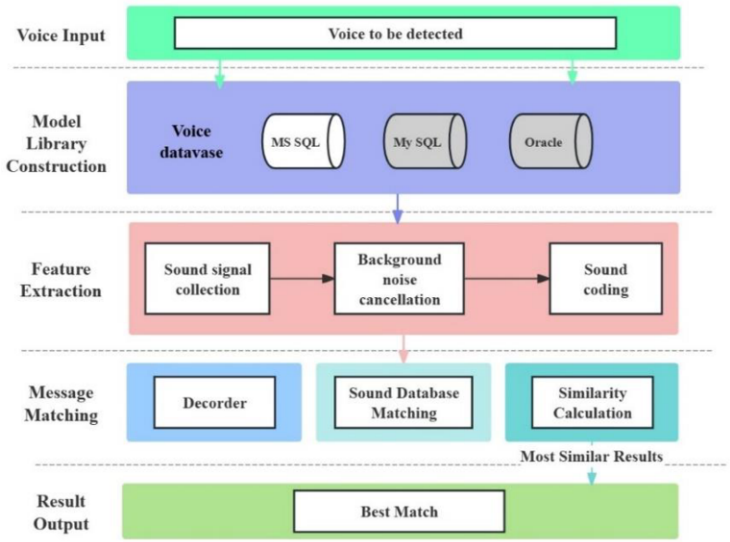

Figure 1. The process of automatic speech recognition

3.2. A Fuzzy Measurement-Based Algorithm for Spoken English Evaluation

In order to accurately quantify and address the problems of subjectivity, fuzziness and uncertainty in the field of language assessment, this paper proposed a flexible and refined evaluation framework that not only integrated multiple evaluation indicators, but also adapted to the specific needs of various types of learners. The objective of this framework is to develop a more precise, individualised and efficient system for evaluating oral English proficiency. In this paper, the features of spoken pronunciation were extracted using the deep learning technique Convolutional Neural Network (CNN). Subsequently, these features were applied to a fuzzy measure-based English-speaking evaluation algorithm to ensure comprehensive and in-depth evaluation results, thereby providing learners with more instructive advice.

3.2.1. Definition of Fuzzy Measures

The definition of fuzzy measure [7], as defined in spoken English evaluation algorithms, is based on fuzzy set theory and serves as a means of quantifying and addressing uncertainty and ambiguity in evaluation. A fuzzy measure μ is a mathematical function that maps a *set A* to a value between 0 and 1, indicating the degree of affiliation or importance of *set A* to an attribute. The closer this value is to 1, the higher the degree of membership of *set A* on that attribute, i.e., the better it represents or satisfies that attribute.

Formally, a fuzzy measure can be defined as follows:

$$\mu: 2^X \rightarrow [0,1] \tag{1}$$

Here, X represents a non-empty finite set that encompasses all potential evaluation features. 2^X denotes the power set of X, which is defined as the set of all subsets of X. The term (A) is used to denote the value of *set A* under the concept of fuzzy measure, reflecting the degree to which A fulfils the evaluation criteria.

In spoken language evaluation, the fuzzy measure can be employed to quantify features such as accuracy and fluency of pronunciation. This enables the transformation of qualitative evaluations of these features into quantitative analysis, thereby establishing a consistent and objective foundation for the evaluation of the algorithm. The membership function enables the fuzzy measure to represent the degree of affiliation of pronunciation features on different evaluation levels such as "excellent", "good", and so on, thereby providing a foundational framework for comprehensive evaluation.

3.2.2. Data Collection and CNN-Based Feature Extraction

In the meta-universe learning environment described in the previous section, a substantial corpus of spoken data from learners is available, including audio and video recordings. It should be noted that the data may contain background noise, unclear pronunciation and non-verbal sounds. The objective of data cleaning is to eliminate these distractions and guarantee that only unambiguous samples of spoken language are employed for analysis.

Subsequently, it is necessary to extract key linguistic features from the cleaned data. Convolutional neural networks (CNN) have demonstrated the potent capability for feature extraction in audio data. By inputting the preprocessed time-frequency representations into CNN, critical linguistic features can be effectively extracted from audio signals. In this paper, the CNN model was designated as the primary algorithmic model for audio feature extraction.

The pre-processed audio representations from the preceding stage were employed as the input to CNN. The convolutional layer of the CNN would learn to extract useful local features from the input data. In the convolutional layer, a convolutional kernel (also referred to as a filter) performs a convolution operation with local regions of the input data, thereby generating a feature map. The specific convolution operation can be expressed as follows: Here, x represents the time-frequency representation of the input, w denotes the weight of the convolution kernel, b signifies the bias term, f is the activation function, and y stands for the output of the convolutional layer. Following a series of trials, five 3X3 convolution kernels were ultimately identified, with a step size of 1 and the activation function set to ReLU.

$$y_{ll} = f\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} w_{mn}\, x_{(l+m)(j+n)} + b\right) \tag{2}$$

Here, x represents the time-frequency representation of the input, w denotes the weight of the convolution kernel, b signifies the bias term, f is the activation function, and y stands for the output of the convolutional layer. Following a series of trials, five 3X3 convolution kernels were ultimately identified, with a step size of 1 and the activation function set to ReLU.

After the convolutional layer, the pooling layer operation is introduced. The pooling layer aims to reduce the spatial size of the feature map, thereby increasing the sensory field of the network and reducing the amount of computation. The most common operations of pooling are maximum and average pooling. In this paper, the average pooling operation was selected, with a pooling kernel size of 2X2 and a step size of 2. This configuration enables the calculation of the average value within a local region in the audio as the output, which in turn reduces the dimensionality of the features while retaining the most important information.

After processing through multiple convolutional and pooling layers, the network generates several feature maps. These feature maps contain a set of hierarchical features extracted from the raw audio data. They can be regarded as abstract representations of

the audio signal, capturing key information and being used for subsequent speech evaluation tasks.

3.2.3. Construction of Fuzzy Comprehensive Evaluation Algorithm

The construction of the fuzzy comprehensive evaluation algorithm is a staged process, commencing with the selection of key evaluation indicator factors and the construction of a fuzzy judgement matrix. The algorithm employed the principle of fuzzy change to calculate the judgement matrix of the evaluation indicators, which typically comprises three principal stages: the construction of the fuzzy matrix, the selection of the fuzzy operator and the synthesis of the result vector. In the evaluation process, weights are assigned in accordance with the degree of influence exerted by different factors on the evaluation object. These weights are subsequently synthesised by means of fuzzy operators in order to obtain comprehensive evaluation results. In this paper, the accuracy, fluency, rhythm and intonation of pronunciation were selected as the evaluation indicators.

In constructing the fuzzy matrix, it is necessary to quantify each influencing factor. The degree of membership of each factor in the factor set is determined by defining the membership function. Commonly used membership functions include trigonometric, Gaussian and trapezoidal functions. In this paper, the trigonometric function was selected for calculation. It is defined mathematically as follows.

$$\mu(A) = \begin{cases} 0 & \text{if } |A| = 0 \\ \frac{1+\frac{|A|-\alpha}{\beta}}{2} & \text{if } 1 \le |A| \le L \\ 1 & \text{if } |A| > L \end{cases} \tag{3}$$

Here $|A|$ represents that the basis α and β of the feature *set A* are parameters that adjust the shape of the function and L is a preset threshold.

Furthermore, the Sugeno integral [8] was adopted to address the problem of fusion of different fuzzy measurements in developing the oral continuous reading evaluation model. The application of the Sugeno integral enables the algorithm to effectively assess the continuous reading group in accordance with the specific continuous reading context. To illustrate, if the continuous reading period exceeds two hours, the system will primarily assess the actual number of consecutive readings completed by the learner in order to ascertain the score for connected reading. This approach offers a more precise representation of the learner's performance in oral continuous reading.

3.2.4. Fuzzy Comprehensive Evaluation and Evaluation Result Outputs

For each evaluation indicator μ_i, the composite degree of membership (A) is calculated based on its membership function (A) and weight w_i:

$$\mu(A) = \sum_{i=1}^{n} w_i \cdot \mu_i(A) \tag{4}$$

This formula weights and sums the degree of membership of all indicators to obtain the composite membership degree. For the determination of the weights w_i associated with each evaluation indicator, here employed a machine learning approach using the Random Forest (RF) algorithm. This method leverages the ensemble learning technique

of multiple decision trees to estimate the importance of each feature, which in context corresponds to the evaluation indicators. The detailed information of the weights w_i is crucial for understanding the influence of each evaluation indicator on the overall assessment. For instance, consider a model for evaluating English oral proficiency with three assessment indicators, which is an important degree score that comes from the training of a Random Forest model: Pronunciation Clarity (u_1), Fluency (u_2), and Intonation (u_3). Employ the Random Forest (RF) algorithm to ascertain the weights for these indicators.

Dataset: A dataset of 1000 samples, each annotated with scores for the three assessment indicators and an overall evaluation grade.

Features and Target Variable: The features are Pronunciation Clarity (u_1), Fluency (u_2), and Intonation (u_3), with the target variable being the overall evaluation grade.

Training the Random Forest Model: A random forest model is trained on this dataset, which automatically computes the importance of each feature.

Feature Importance: Post training, the RF model provides importance scores for each feature. For instance:

- Pronunciation Clarity (u_1): 0.45
- Fluency (u_2): 0.30
- Intonation (u_3): 0.25

Normalization of Weights: These importance scores are normalized to ensure they sum up to one, maintaining the proportional influence of each indicator on the composite membership degree. The normalized weights w_i are:

$$\begin{aligned} w_1 &= \frac{0.45}{0.45+0.30+0.25} \approx 0.47 \\ w_2 &= \frac{0.30}{0.45+0.30+0.25} \approx 0.31 \\ w_3 &= \frac{0.25}{0.45+0.30+0.25} \approx 0.22 \end{aligned} \tag{5-7}$$

Application of Weights: These weights are applied to calculate the composite membership degree $\mu(A)$ for each sample.

Based on the calculated composite membership degree (A), the evaluation object is classified into the corresponding evaluation level. For example, the evaluation levels can be defined as “excellent”, “good”, “fair” and “poor”, with a threshold range of membership degree set for each level.

Finally, the results of the comprehensive evaluation are conveyed in a transparent and intelligible way to provide suitable feedback to the learners.

3.3. Algorithm Optimisation Strategy

In order to integrate fuzzy measures and provide different levels of assessment results, the algorithm employed the Sugeno integral [8], which is a special kind of fuzzy integral that is able to integrate multiple fuzzy measures to give an overall assessment. This method is capable of not only assessing individual articulations, but also continuous articulations, thereby providing more comprehensive and accurate results.

For each evaluation dimension, the degree of membership is calculated based on its fuzzy measure function, and then these degrees of membership are synthesised into a single value by using the Sugeno integral formula. The Sugeno integral is typically defined as the supremum of the minimum value or product of all membership degrees.

The processing of the difficult pronunciation set (HDP) refers to a specialised technique for the assessment of spoken English, specifically for the identification of confusable phonemes. Firstly, by analysing all possible pronunciations within English sentences, a database encompassing these pronunciations is constructed. Subsequently, when a learner pronounces a word, the system captures the pronunciation features and searches for the most matching pronunciation path in the database. This is then used as a basis for assessing the accuracy of the pronunciation and providing targeted feedback and suggestions for improvement.

The fundamental principle underlying the construction and utilisation of HDP sets is that it facilitates an in-depth analysis of learners' pronunciation patterns, particularly those error-prone phonemes. This approach not only assists learners in identifying and enhancing their challenging pronunciation but also enables a more sophisticated and tailored assessment of their performance, which can effectively enhance their speaking abilities and comprehension.

4. Experimental Results

4.1. Experiment Overview

In order to verify the validity of the English proficiency assessment model combining big data analysis with fuzzy measurement, a series of experiments were conducted with a total sample size of 100, ensuring a diverse representation of different proficiency levels, age groups, and language backgrounds. The aim is to test the ability of the model to accurately extract key information from learner data and effectively quantify the fuzzy factor of English ability through fuzzy measurement. After the experiment, we selected 5 samples from different populations through stratified sampling to compare the results. The study used data from learners of different levels, including individuals of different levels, ages, and backgrounds. In order to ensure the accuracy and reliability of the data, the data were cleaned and preprocessed, and about 5% of the outliers and noise data were eliminated. In addition, the assessment process is carefully reviewed by professional English teachers to ensure the accuracy and credibility of the results.

4.2. Analysis of Discrepancies

After comparison experiments we analyzed the differences between model and evaluator scores to identify potential areas for model improvement or to understand the variability of human assessments.Our enhanced fuzzy comprehensive evaluation model has been fortified with an expanded and diverse dataset. This enlargement has bolstered the model's robustness and its ability to provide nuanced assessments of learners' spoken English skills.

The membership degree remains a pivotal metric, reflecting the proficiency level of each learner across key dimensions: pronunciation clarity, fluency, and intonation. With these degrees, we can discern the proximity of a learner's performance to the ideal benchmarks, where values nearing 1.0 indicate exceptional proficiency. The incorporation of human evaluator scores has introduced a critical validation step, ensuring that our model's assessments are not only data-driven but also aligned with human judgment. The calculated average accuracies, ranging from 95.23% to 98.70%, reveal a remarkably high degree of congruence between the model's evaluations and

those of human evaluators. This affirms the reliability and trustworthiness of our model's outputs.The results highlight the model's efficacy in evaluating a wide array of learners, offering a nuanced understanding of their spoken English skills. The high average accuracy not only validates the model's dependability but also underscores its capacity to provide constructive feedback aimed at enhancing oral communication skills.

In summation, the fusion of big data analytics with fuzzy measurement theory has marked a significant leap forward in the domain of educational technology. Our model's improved accuracy and its capacity to deliver tailored feedback are instrumental in enriching the learning experience. This advancement lays a solid groundwork for the evolution of groundwork for the evolution of intelligent educational systems.

5. Conclusion

The implementation of English-speaking evaluation algorithms within the context of big data analysis on the Metaverse English learning platform presents a promising opportunity for enhancing the efficacy of English learning and the precision of English teaching. As a consequence of the ongoing advancement of information technology and the enhancement of algorithms through the analysis of big data, metaverse English learning will become increasingly intelligent and individualised. This will provide global English learners with an interactive, expansive and immersive English learning platform, thereby establishing a robust foundation for the development of a more intelligent, efficient and individualised metaverse English learning ecosystem.

References

[1] Suzuki SN, Kanematsu H, Barry DM, Ogawa N, Yajima K, Nakahira KT, Shirai T, Kawaguchi M, Kobayashi T, Yoshitake M. Virtual Experiments in Metaverse and their Applications to Collaborative Projects: The framework and its significance. Procedia Computer Science. 2020 Sep;176:2125-2132, doi: 10.1016/j.procs.2020.09.249.

[2] Sinha E. 'Co-creating' experiential learning in the metaverse- extending the Kolb's learning cycle and identifying potential challenges. The International Journal of Management Education. 2023;21(3):100875, doi: 10.1016/j.ijme.2023.100875.

[3] Xue, N. Analysis Model of Spoken English Evaluation Algorithm Based on Intelligent Algorithm of Internet of Things. Computational Intelligence and Neuroscience. 2022;2022(1):8469945, doi: 10.1155/2022/8469945.

[4] Wang J. Speech recognition sensors and artificial intelligence automatic evaluation application in English oral correction system. Measurement: Sensors. 2024;32:101070. doi:10.1016/j.measen.2024.101070.

[5] Sheng Y, Yang, K. Automatic Correction System Design for English Pronunciation Errors Assisted by High-Sensitivity Acoustic Wave Sensors. Journal of Sensors. 2021;2021(1):2853056, doi: 10.1155/2021/2853056.

[6] Lampert TA, Stumpf A, Gançarski P. An Empirical Study Into Annotator Agreement, Ground Truth Estimation, and Algorithm Evaluation. IEEE Transactions on Image Processing. 2016 Jun;25(6):2557-2572. doi:10.1109/TIP.2016.2544703.

[7] Zhao L, Liu Y, Chen L, Zhang J, Koomey JG. English oral evaluation algorithm based on fuzzy measure and speech recognition. Journal of Intelligent and Fuzzy Systems. 2019 Jul;37(1):241-248. doi: 10.3233/JIFS-179081.

[8] Enemegio R, Jurado F, Villanueva-Tavira J. Experimental evaluation of a Takagi-Sugeno fuzzy controller for an EV3 ballbot system. Applied Science. 2024;14(10):4103. doi:10.3390/app14104103