Pronunciation Assessment: Traditional vs Modern Modes

Ali Babaeian

Faculty of Arts and Social Science

Center for Educational Measurement and Assessment, The University of Sydney

3 Parramatta Rd, Camperdown NSW 2050, Australia

Tel: 61- 2 9351 6383 E-mail: abab5934@uni.sydney.edu.au

Received: Sep. 10, 2023 Accepted: Nov. 1, 2023 Published: November 1, 2023

doi:10.5296/jse.v13i4.21417 URL: https://doi.org/10.5296/jse.v13i4.21417

Abstract

Pronunciation assessment is a crucial component of language teaching and learning. Nonetheless, there seems to be a lack of consensus on the best methods and criteria for evaluating learners' pronunciation. This review provides an overview of the traditional and modern modes of pronunciation assessment, highlighting their strengths and limitations and their approaches to assessing pronunciation. Traditional modes solely rely on human raters' judgments and linguistic knowledge and may be encroached upon on various grounds. Modern modes often use artificial intelligence, automatic speech recognition, and complicated programs to measure and score pronunciation features. The review concludes with some implications and future directions for modern modes of pronunciation assessment.

Keywords: pronunciation assessment, mode, traditional, modern, human rater, artificial intelligence

1. Introduction

Pronunciation constitutes a remarkable part of linguistic competence and plays a crucial part for a variety of reasons. Primarily, it aids English speakers in communicating effectively, thereby augmenting their communication and confidence (Derwing & Munro, 2015). Failure to produce at least understandable pronunciation of words and utterances may lead to listeners' misunderstanding and frustration (Derwing & Murano, 2005; Levis, 2005), and therefore, teachers, learners, and scholars unanimously regard pronunciation as the pillar of successful verbal communication (Derwing, 2003; Waniek-Klimczak, 2011). The impact of pronunciation on verbal skills is so significant that most language assessments have incorporated pronunciation evaluation into their assessments to measure this essential skill so that they can scale the speaker's pronunciation abilities.

Pronunciation assessment encompasses the process of appraising the quality and accuracy of the speech production of the target language and is traditionally human-rater dependent. Nevertheless, most assessment-designing organizations have been steering away from this tradition in recent years, embracing innovative modes of measuring pronunciation through technology (Issacs, 2017). The use of technology, specifically artificial intelligence (AI)-powered platforms, has made it possible for language assessments to offer swift results within a few hours while maintaining optimum standards of validity.

2. Pronunciation Assessment

Pronunciation assessment entails evaluating the extent to which a speaker is able to produce and perceive the speech features of the target language. Sounded and unsounded speech features function conjointly to produce language sounds, and assessments make judgments about the degree of the speaker's speech productivity. Nevertheless, the emphasis on specific speech features and their subsequent evaluation is contingent upon the purpose of the assessment, and these features may vary across different assessments. The administration of pronunciation assessment fulfills manifold purposes in diverse contexts, including diagnostic (i.e., identifying specific areas of pronunciation that learners may require assistance with), achievement (i.e., deciding whether the learners have acquired certain phonological features), and proficiency (i.e., deciding whether the speaker is overall understandable); each examines the speech features differently (Derwing & Munro, 2015; Harding. 2013; Pennington & Rogerson-Revell, 2018). The first two assessment methods are typically used in educational settings, whereas large-scale language tests often offer proficiency tests for admission reasons. The International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL), Test of English for International Communication (TOEIC), and Pearson Test of English Academic (PTE) are prominent language tests that have been designed to assess language proficiency. These tests seem to adhere to comparable perspectives in establishing criteria for pronunciation assessment.

The assessment of pronunciation proficiency in language tests often adopts the holistic approach. This is because pronunciation is an intricate and multifaceted aspect of verbal communication, and a range of factors (e.g., the speaker's linguistic background, task type, and topic) may influence it (Isaacs & Thomson, 2013); therefore, an assessment perspective is required to mitigate these issues. The holistic approach regards pronunciation as a whole entity, focusing on the speaker's overall understandability rather than a compilation of discrete sounds (British Council, 2023). As a result, proficiency tests assess pronunciation based on the speaker's ability to use pronunciation resources to effectively and accurately convey verbal messages in various contexts. The notion of intelligibility is relevant in this context. The term "intelligibility" pertains to the extent to which a listener or a collective audience understands a speaker's verbal message (Field, 2005; Levis, 2018). A holistic pronunciation assessment would consider intelligibility since it aligns with the ultimate goal of communication, that is, mutual understanding.

Conventionally, assessing pronunciation relies on human raters (e.g., teachers and examiners), who judge the spoken output of speakers using a range of criteria. Nevertheless, human raters may be susceptible to limitations and bias (e.g., fatigue, emotion, inconsistency, and lack of expertise), so their judgment may be impinged (Linn & Gronlund, 2000; Mayford & Wolfe, 2003), and test results may not reflect the candidate's abilities. A study by Yates et al. (2011) indicated that human raters of a large-scale language test had a tendency to erroneously assign lower pronunciation scores, and only one-third of the tests aligned with the test-assigned scores. As a consequence, inconsistencies and bias may have significant implications due to the fact that language proficiency tests are frequently used in making decisions on admittance into academic programs or migration to a number of English-speaking countries.

A second issue with the traditional assessment method lies in the paucity of lucid and consistent criteria for rating speech samples. Diverse raters may have disparate expectations, preferences, and judgments about what constitutes acceptable or target-like pronunciation, particularly when evaluating different variations of English or non-native speakers (Pennington & Rogerson-Revell, 2018). Furthermore, raters may be swayed away by a range of factors, including the content, context, purpose of the assessment, background, and identity of the speaker, and their own linguistic and cultural biases (Cambridge, 2023). Hence, these factors have made human rating seem time-consuming and costly, mainly when used for large-scale tests with frequent occurrences (Harding, 2013).

In order to surmount these constraints, language-designing bodies have recently made a shift to newer modes of assessment, devising technologically based alternatives to traditional forms of assessment (Issacs, 2017; Williamson et al., 2012). One such attempt at assessing second language speaking abilities with the centrality of pronunciation is introducing indirect assessment through automated speech scoring. The use of state-of-the-art software programs for assessing speaking skills and pronunciation, in particular, has paved the path for assessing productive language skills. This assessment mode enables the evaluation to be scored outside live assessment conditions in an unbiased situation without human intervention (Issacs, 2017). This assessing approach takes into account specific speech aspects.

3. Speech Features in Pronunciation Assessment

As the holistic perspective being in effect in assessing pronunciation proficiency, intelligibility is the primary criterion in building scoring rubrics. Moreover, other approaches, such as English as a Lingua Franca (ELF) and World Englishes (WE), have consolidated this perspective (Derwing & Munro, 2009; Jerkins, 2000). In other words, phonological features may vary across English variations, influencing understandability. Therefore, a universally accepted phonological pattern is required as a reliable criterion for pronunciation assessment. English as a lingua franca is a language of communication among speakers who have different first languages and may or may not be native speakers of English (Jenkins, 2015). World Englishes is a collection of English variations that have emerged in different world regions due to colonization, migration, education, and media (Kachru, 1985).

Phonological features have a pivotal part in shaping intelligibility. While segmental features (i.e., vowels, consonants, diphthongs, consonant clusters) partially impact intelligible pronunciation, suprasegmental features (i.e., stress, intonation, rhythm, pitch) influence it to a greater extent (Jenkins, 2000; 2002). As a result, suprasegmentals seem to outweigh segmentals in terms of scoring pronunciation proficiency. Celce-Murcia et al. (1996) argue that errors in segmental sounds can result in minor repairable misunderstandings, but errors at suprasegmental levels can affect the entire utterance because they carry more of the overall meaning load of speech. Language proficiency tests, including those administered via AI-based platforms, seem to take this particular scoring criterion into account when evaluating a speaker's phonological control.

4. Technology in Pronunciation Assessment

The use of technology for educational purposes has been prevalent for the past few decades. Computer-Assisted Pronunciation Training (CAPT) concentrates on the enhancement of pronunciation skills through interactive software, offering a plethora of activities, including read-aloud, repetition, and games. However, newer methods of technology application have assisted education where Automatic Speech Recognition (ASR) and AI work in concert to assess pronunciation. In recent years, assessment tools such as Language Confidence and Speechace have launched platforms to evaluate pronunciation using similar assessing technology. These platforms provide feedback at different levels of granularity, ranging from individual phonemes to the entire utterance, as well as the estimated scores on standardized scales provided by IELTS, PTE, and the Common European Framework of Reference for Languages (CEFR).

An essential constituent of automated speech assessment is that of ASR. Almost all AI-powered pronunciation assessing tools heavily rely on ASR technology. ASR, or voice recognition component, analyzes the speaker's speech, converting voice commands into text for further analysis and scoring. In conjunction with AI, speech-to-text software processes the speech, converting it into text, and the text is compared and matched to the closest lexical entry in the database, forming the entire utterance or sentence. Once the system decides the accuracy of speech, it proceeds to score pronunciation through complicated methods. The entire assessment process, including speech recognition, analysis, and scoring procedures, occurs in an unbiased environment with no human intervention and optimum validity standards (Cámara-Arenas et al., 2023).

The integration of technology has also permeated the domain of large-scale, high-stakes language tests. High-stakes tests are those with significant consequences for the test takers (Marchant, 2004). AI-based systems have found their way into evaluating proficiency in pronunciation through accurate programs and algorithms. TOEFL, for example, uses SpeechRater, an AI-driven platform for recognizing, analyzing, and scoring speakers' pronunciation. Similarly, PTE utilizes a sophisticated system in order to assess the speaker's speech, assigning a discrete score for pronunciation. These commercial language tests emphasize that their tests are of high validity and reliability.

AI-based pronunciation assessments are often touted for their high validity and reliability; nonetheless, it is essential to acknowledge that they also possess some benefits and downsides. On the one hand, AI-powered assessments have enhanced transparency, consistency, and accessibility compared to human raters (Xiao & Chen, 2020). On the other hand, they may not adequately capture the complexity and variability of real-life spoken language use and may not reflect the communicative goals and functions of the test takers (Xiao & Chen, 2020). Furthermore, using AI in high-stakes tests raises ethical and social concerns. These concerns encompass the fairness, validity, reliability, and accountability of the test results and the impact of test administration on the behaviors, attitudes, beliefs, values, and opportunities of both test takers and other stakeholders (Shohamy & McNamara, 2009).

Each of these technologies uses specific operational methods to conduct pronunciation assessment.

5. Operational Mechanism by AI-Powered Systems

AI-based testing platforms use various technologies to assess speaker's pronunciation skills reliably and efficiently. Despite the existence of a variety of technologies for pronunciation assessment, ASR is the leading technology used in conjunction with speech analysis programs (Chen et al., 2020). Speech analysis programs examine several features, including stress, intonation, and rhythm, which are influential variables in intelligibility (Witt & Young, 2000). Language assessments often use these technologies based on an acoustic model to be able to capture sounds (Zhang et al., 2020).

From speech input to score generation, the entire system comprises three essential models: an acoustic model for speech recognition that generates word hypotheses based on a given speech sample, a language model that computes some features measuring different aspects of speech, and finally, a scoring model that maps features to a score using a defined paradigm (Chen et al., 2018; Van Moere & Downey, 2016). All these components collaborate harmoniously to generate speaking scores, including pronunciation. The chart below illustrates the scoring process by automated scoring systems (see Figure 1).

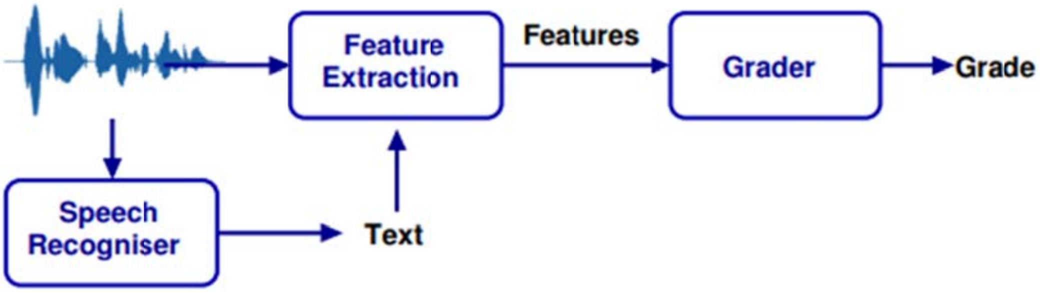

Figure 1. Automated Speech Scoring System

The main component of speech scoring systems is the acoustic model, enabling the system to distinguish phonemes and sounds through the hidden Markov model (HMM) (Young, 2001). The hidden Markov model is a performance-based program used in some automated systems for assessing speaking abilities. Acoustic models must be optimized or "trained" on a set of speech data by pairing the speech with the speech transcription so that the model associates sounds with orthographic representations (Hinton et al., 2012). The language model consists of words likely to be spoken in the responses and uses n-gram frequencies for test takers' responses. This model anticipates test takers' utterances, increasing speech recognition accuracy remarkably (Balogh et al., 2012). The scoring model selects features from the speech recognition process and applies them to predict human ratings (Van Moere & Downey, 2016).

In summary, despite the downsides associated with technology and the use of AI, the advantages of these advancements outweigh the disadvantages in the context of pronunciation assessment. Language assessments, thus, are undergoing a significant transformation by adopting AI-powered platforms, which provide prompt results while maintaining rigorous standards of validity. Moreover, the technologies used in these assessments are continuously progressing, leading to minimizing technical errors that may impact overall assessment results.

6. Future Directions

The surge in popularity of technologically-based assessments will lead to a notable transition from traditional to technology-driven evaluation, where AI-based assessment modes are anticipated to assume a crucial role in assessing pronouncing skills. However, the future of AI-powered platforms for pronunciation assessment is promising and challenging. On the one hand, AI-powered platforms can offer more benefits over human ratings, such as scalability, consistency, transparency, and accessibility. AI-powered platforms can also provide learners with more personalized and adaptive feedback based on their learning goals, preferences, and progress. AI-powered platforms can also integrate with other learning resources and platforms to create a more comprehensive and engaging learning experience. On the other hand, AI-powered platforms also face some limitations and challenges that need to be addressed, such as the validity and reliability of the scoring models (Chapelle et al., 2008), the diversity and representativeness of the speech data (Zhang et al., 2020), the ethical and social implications of using AI for high-stakes testing (Shohamy & McNamara, 2009), and the integration of human and machine feedback for optimal learning outcomes (Neri et al., 2008). Therefore, more research and development are needed to ensure that AI-powered platforms can meet the needs and expectations of learners and teachers in large-scale language tests.

References

Balogh, J., Bernstein, J., Cheng, J., Van Moere, A., Townshend, B., & Suzuki, M. (2012). Validation of automated scoring of oral reading. *Educational and Psychological Measurement*, *72*(3), 435-452. https://doi.org/10.1177/0013164411412590

British Council. (2023). Holistic approach. https://www.teachingenglish.org.uk/article/holistic-approach

Cámara-Arenas, E., Tejedor-García, C., Tomas-Vázquez, C. J., & Escudero-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Language Learning & Technology, 27*(1), 1-19. https://hdl.handle.net/10125/73512

Cambridge. (2023). Pronunciation Assessment. Language Teaching. Cambridge Core. https://www.cambridge.org/core/journals/language-teaching/article/pronunciation-assessment/9F0B4A8F5C9E1A5B0E4D6D9C5F1E8F7A]

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.). (2008). Building a validity argument for the Test of English as a Foreign Language. Routledge.

Chen, L., Zechner, K., Yoon, S. Y., Evanini, K., Wang, X., Loukina, A., ... & Gyawali, B. (2018). Automated scoring of non-native speech using the speechrater sm v. 5.0 engine. *ETS Research Report Series*, *2018*(1), 1-31. https://doi.org/10.1002/ets2.12198

Chen, Y., Li, J., Wang, N., Zhang, Z., & Wei, X. (2020). *Speech Recognition: Technologies and Applications*. Springer.

Derwing, T. (2003). What do ESL students say about their accents?. *Canadian Modern Language Review*, *59*(4), 547-567. https://doi.org/10.3138/cmlr.59.4.547

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, *39*(3), 379-397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language teaching*, *42*(4), 476-490.

https://doi.org/10.1017/S026144480800551X

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*, 42. John Benjamins Publishing Company. https://doi.org/10.1093/applin/amw041

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, *39*(3), 399-423. https://doi.org/10.2307/3588487

Harding, L. (2013). Pronunciation assessment. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics*, 4521-4529. Blackwell. https://doi.org./10.1002/9781405198431.wbeal0966

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, *29*(6), 82-97. https://doi.org/10.1109/MPS.2012.2205597

Isaacs, T. (2017). Fully automated speaking assessments: changes to proficiency testing and the role of pronunciation. Routledge. https://doi.org/10.4324/9781315145006-36

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford university press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied linguistics*, *23*(1), 83-103. https://doi.org/10.1093/applin/23.1.83

Jenkins, J. (2015). *Global Englishes: A resource book for students* (3rd ed.). Routledge.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge University Press.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL quarterly*, *39*(3), 369-377. https://doi.org/10.2307/3588485

Levis, J.M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Linn, R. L., & Gronlund, N. E. (2000). Measurement and assessment in teaching (81hEdition). *New Jersey, USA: Prentiee Hall.*

Marchant, G. J. (2004). What is at stake with high-stakes testing? A discussion of issues and research. *Ohio Journal of Science*, *104*, 2-7.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of applied measurement*, *4*(4), 386-422.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer-assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning, 21*(5), 393-408. https://doi.org/10.1080/09588220802447651

Pennington, M. C., & Rogerson-Revell, P. (2018). English pronunciation teaching and research: Contemporary perspectives. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-47677-7

Shohamy, E., & McNamara, T. (2009). Language tests and human rights. *International Journal of Applied Linguistics, 19*(1), 5-13. https://doi.org/10.1111/j.1473-4192.2008.00191.x

Van Moere, A., & Downey, R. (2016). 21. Technology and artificial intelligence in language assessment. In *Handbook of second language assessment*, 341-358. De Gruyter Mouton. https://doi.org/10.1515/9781614513827-023

Waniek-Klimczak, E. (2011). *"I Am Good at Speaking, But I Failed My Phonetics Class" Pronunciation and Speaking in Advanced Learners of English*, 117-130. Multilingual Matters. https://doi.org/10.21832/9781847694126-010

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational measurement: issues and practice*, *31*(1), 2-13. https://doi.org/10.1111/j.1745-3992.2011.00223.x

Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech communication*, *30*(2-3), 95-108. https://doi.org/10.1016/S0167-6393(99)00044-8

Xiao, B., & Chen, C. (2020). A Review of Automatic Scoring Systems for Spoken English Proficiency Tests. *IEEE Access, 8,* 14900-14915.

Yates, L., Zielinski, B., & Pryor, E. (2011). The assessment of pronunciation and the new IELTS Pronunciation Scale. In *IELTS Research Reports, 12,* 1-46. Melbourne: IDP: IELTS Australia and British Council.

Young, S. J. (2001). Statistical Modeling in Continuous Speech Recognition (CSR). In *UAI* (Vol. 1, pp. 562-571). https://doi.org/10.48550/arXiv.1301.2318

Zhang, K., Wei, Y., Wang, K., Zhao, S., Liao, Q., Beatman, A., & Adeogba, D. (2020). Hierarchical Transformer for Pronunciation Assessment. In *Proceedings of Interspeech 2020* (pp. 1-5).