NAS Journal of Artificial Intelligence and Computer Science (NASJAICS)

Homepage: https://naspublisher.com/nasjaics/

eISSN: XXXX-XXXX [Online] | ISSN: XXXX-XXXX [Print]

Volume 1 • Issue 1 • Nov-Dec, 2026

A Multimodal Generative AI Tutor for Teaching Phonological Processes and Feature Geometry: Enhancing Metaphonological Awareness, Articulatory Accuracy, and Transfer in L2 Learners

Dr Akamo O. Princewill

Department of English, University of Ibadan

*Corresponding Author: Dr Akamo O. Princewill | ORCID: [0009-0003-2445-2562]

Received	Accepted	Published
10.09.2026	10.10.2026	01.11.2026

Copyright © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract

This study created and assessed a multimodal generative AI tutor that provides explicit instruction on phonological processes and feature geometry to second language learners. This study combined a multilingual model, text to speech synthesis, interactive articulatory visuals and adaptive rule induction. This was incorporated into the classroom for 8 weeks. 90 L2 English learners were asked to take part in the mixed-methods quasi-experimental design. Learners were separated into two groups: AI tutoring and traditional explicit instruction. The tutor administered tests to measure learners' knowledge of processes (assimilation and feature spreading), phonological awareness, accuracy in controlled production and knowledge transfer into free speech. The results revealed significantly larger gains in the AI group for explicit knowledge and metaphonological awareness. Proper pronunciation added value significantly, for process-based errors within the AI condition, which notably notable transfer to spontaneous speech. Learners conveyed that the instructional framework were highly useful, though an insignificant number

noted occasional AI hallucinations in the rule explanations. The findings portrayed that multimodal generative AI can link theory in phonology with real human speech pronunciation. Furthermore, it proposals a tutor system that can be reprocessed and shows that teaching phonological structures function better when delivered in an suitable multimodal way. Likewise, analysis showed that gains were strong for phonological processes where features cut across ordered nodes. Looking at how learners interacted with the system, those who used the feature tree tools often scored higher on transfer tasks. In all, working dynamically with structure helps more, than getting generative clarifications.

Keywords: generative explanations, Artificial Intelligence, phonological processes, explicit instruction, production accuracy.

1.1 Introduction

Pronunciation is a branch of language that always gives second language learners a lot of difficulties. Many learners, despite considerable practice, continue to generate utterances that reflect interference from their first language phonology or demonstrate insufficient mastery of target language phonology. Research has found that form-focused learning can have positive effects on both controlled and spontaneous production when applied correctly. However, current methods and tools when applied to pronunciation focus primarily on simple segmental accuracy or certain suprasegmental features like stress or tone. Teaching materials and methods that provide higher-level descriptions that would allow recognizing typical patterns or processes of error formation are much less widespread and underrepresented in both computer-assisted and traditional classroom instruction.

Feature geometry groups distinctive features into hierarchical nodes that define natural classes and restrict possible applications of phonological processes. Assimilation, deletion, insertion, and spreading processes act upon this geometry that create regular patterns of change for learners to learn. Development of explicit understanding of these processes enables the learner to increase their metaphonological awareness which has been shown to contribute to self-monitoring, error detection, and generalization to unpracticed items. The traditional approach to instruction is primarily didactic: concepts are learned through lectures and diagrams that require extensive working memory resources. Learners receive limited opportunities to explore the conceptual domain, test informal hypotheses, and receive immediate multimodal feedback. Recent advances in the domain of generative artificial intelligence offer novel instructional opportunities.

Large language models can be used to provide explanations based on students answers and hypothesis testing behavior. Speech synthesis can generate target exemplars. Graphical models can illustrate articulatory configurations. Update hierarchical feature trees in real time. Learners who operate such a system can develop mastery of rule induction, receive unified verbal and visual feedback, and practice in an environment which adapts to their skill level. This study was designed to test such a framework. We hypothesized that extended learning in this system would build metaphonological awareness and explicit knowledge than traditional explicit instruction. We also hypothesized that these gains would come with performance on production tasks and, on spontaneous speech.

This research looks at an important area that hasn't been studied enough. Even though computer-assisted pronunciation training has gotten better at teaching speech sounds and rhythm, let us that AI tools have

been used for early reading and basic language tasks, there is not enough solid research on AI that creates content in multiple forms and is made specifically for teaching how sounds work and how they are structured. Importantly very few studies follow the entire path of learning from first understanding to practicing and finally using the skills, in real speech. By filling this gap this work gives both ideas and useful improvements.

1.2 Background of the Study

Evidently, language phonological learning is often being influenced by native language. According to Brown (1998, 2000) whether learners can hear and say sounds that are not in their first language depends on whether those sounds exist in their first language. Feature geometry is both a way to study sounds and a helpful tool for teaching. When students learn this structure they understand better why certain sound changes happen. Teaching students about these ideas leads to results. Wrembel and others found that when learners become more aware of sounds they check themselves more and make mistakes. Also, going through several studies on this subject matter show that teaching directly, especially when it includes quick and specific feedback, has a big positive impact. Even though computers can help with practice and checking work most programmes only focus on surface-level accuracy. Not, on the deeper structure of language.

Language learning is getting tools with the help of generative AI. Large language models make it possible to have conversations about grammar rules. These models can also work with types of information like text, sound and pictures, such as diagrams showing how speech sounds are made and how they are organised. Some early studies show that AI tutors can help with learning how sounds work and improving how clear speech is. These AI tutors do not fully use all the ideas from the study of speech sounds in a system that changes based on what the user needs. This study builds on those steps. It presents a tutor that is made for teaching the ideas behind how speech sounds are structured and how they change.

1.3 Statement of the Problem

There is a wealth of research supporting the effectiveness of conscious pronunciation teaching. A meta-analysis of quasi-experimental research found large to medium effect sizes for overt segmental and suprasegmental teaching (Lee, Jang & Plonsky, 2015). This has been supported by recent research which shows that phonetic explicit instruction can improve learners' controlled and spontaneous speech production when optimal conditions are applied (Saito & Plonsky, 2019). Theoretical knowledge of phonology can be converted into procedural knowledge through skills acquisition theory. Metaphonological knowledge, a form of overt analytic knowledge of L2 phonology (Wrembel, 2015), regularly improves learners' monitoring, accuracy and can even improve their productions of L3 tones.

Feature geometry is based around Clements' (1985) hierarchy of feature trees, natural classes and spreading/delinking. It has been shown that L1 feature geometry affects learners' ability to acquire contrasts in L2 (Brown, 2000). It has had limited use in teaching but provides an interesting explanation of why learners make consistent errors and how their phonology changes. In addition, research has shown that Computer Assisted Pronunciation Training (CAPT) such as Automatic Speech Recognition

(ASR) feedback, articulatory visualisations and multiple modalities improves accuracy, engagement and retention (Mahdi & Al Khateeb, 2019; Ngo et al., 2024; Almusharraf, 2024).

Generative AI has recently been used in phonological teaching. Pilots have been conducted which show that using AI tutors based on GPT and feedback generated from AI stories can improve learners' metaphonological reflection and prosodic productions (Łodzikowski et al., 2025; Hirschi, 2025). However, problems with hallucination mean there needs to be some structure in place. While research has shown that overt teaching, feature geometry, CAPT and GenAI can all help learners, there has been no research conducted to evaluate a multimodal generative tutor specifically using feature geometry to measure learners' progress across the entire spectrum from metaphonological awareness to spontaneous speech. This study fills this gap.

1.4 Aim and Objectives of the Study

The main aim of this research in the area of second language acquisition and applied linguistics is to create, put into practice and test a multimodal AI tutor. This tutor is meant to provide instruction on phonological processes and feature geometry. The study will look at how this tutor affects learners metaphonological awareness, their knowledge of phonology how accurately they produce sounds and whether these skills can be used in real second language contexts.

1.5 Research Objectives

1. To. Improve a multimodal generative AI tutor that uses large language models for conversation, speech synthesis, interactive visualizations of feature geometry and representations of articulatory gestures. The purpose is to support teaching of phonological processes.

2. To compare how effective the AI tutor is compared to methods of explicit teaching when it comes to improving learners' metaphonological awareness and their understanding of phonological processes and feature geometry.

3. To explore how gains in phonological knowledge and metaphonological awareness affect production accuracy. This includes testing, in both controlled experiments and natural spoken language situations.

1.6 Research Problems

1: Does sustained interaction with a multimodal generative AI tutor produce significantly greater improvement in L2 learners' metaphonological awareness and explicit knowledge of phonological processes and feature geometry than traditional classroom based explicit instruction?

2: To what degree do measured gains in explicit metaphonological knowledge transfer to higher production accuracy for process related errors in controlled speaking tasks?

3: How effectively do improvements in explicit knowledge and controlled production carry over into spontaneous speech, and how do learners evaluate the accuracy, usefulness, and pedagogical value of the AI generated visualizations and explanations?

2.1 Literature Review

Research on explicit pronunciation instruction has established that focused attention to form yields reliable gains. A good number of works have shown that explicit pronunciation instruction works. To corroborate this fact, Lee, Jang and Plonsky (2015) opine that systematic segmental and suprasegmental training has medium-to-large effect sizes. Also, Saito and Plonsky (2019) argue that explicit pronunciation instruction helps learners produce spontaneous speech when conditions are good. These findings correlate with skill-acquisition models which is of the fact that knowing structure as facts can become a skill with practice. The key part of this process is awareness, which Wrembel (2015) defines as explicit, analytical knowledge of L2 sound properties. Metaphonological awareness always helps learners check their speech more accurately and even produce L3 tones better.

Feature geometry gives the theoretical structure for explicit pronunciation instruction. Based on Clements (1985) hierarchy of feature trees, natural classes and spreading/delinking operations, Brown (2000) showed that L1 feature geometry limits how learners acquire L2 contrasts. Even though feature geometry is not used much in teaching design, it explains why learners keep making the mistakes and why their sound inventories change. At the time research on Computer-Assisted Pronunciation Training (CAPT) shows that automatic speech recognition (ASR) feedback, visual articulatory models and multimodal inputs can improve accuracy, engagement and retention (Mahdi & Al Khateeb 2019; Ngo et al., 2024; Almusharraf, 2024).

Recently, generative AI tools have moved into training. Pilots by Łodzikowski et al. (2025) And Hirschi (2025) show that GPT-driven tutors and AI-generated narrative feedback improve reflection and prosodic performance.. Problems, like model confabulation show that we still need strong safeguards. While many studies confirm the value of instruction, feature-based frameworks CAPT and GenAI tools no one has tested a multimodal generative tutor that is built around feature geometry and that tracks learning from metaphonological awareness all the way to spontaneous speech. The study gives special attention to this gap.

2.2 Theoretical Framework

Four theories were meticulously utilized in conducting this research. According to DeKeyser 2007 Skill Acquisition Theory says that declarative knowledge gained by teaching turns into procedural skill and then into automatic skill when practice and feedback are organized. In this study the AI tutor gives the base through interactive rule induction and adaptive production tasks help turn that knowledge into practice. Schmidt 1990 in the Noticing Hypothesis says that paying attention to target features is key for learning. The system utilizes a multimodal design that portrays the unification of text, audio and pictures, like feature trees, to concentrate on how the features function in order. Mayer 2009 with the Cognitive Theory of Multimedia Learning adds ideas about how to use two channels and active learning. The tutor will be able to regulate the pace of the information provided and connect the spoken words to images meaning that there will be no excessive loading of the minds of learners. This enables learners to think more profoundly about concepts. Finally, Feature Geometry Theory regulates the content of the domain. The theory structures the teaching around feature trees, naturally grouped items and formal phonological processes. This combines the learning mechanisms, focus of attention, multimedia design and linguistic content to create a teaching model.

3.1 Research Methodology

There are avalanche studies backing up the resourcefulness of focused pronunciation teaching strategies. An analysis of quasi experimental studies showed that systematic instruction on suprasegmental and segmental pronunciation can produce moderate to large effect sizes (Lee, Jang & Plonsky, 2015). In addition, optimum conditions produced gains in learners' ability to produce both controlled and spontaneous speech after explicit phonetics instruction (Saito & Plonsky, 2019). This is in line with skill acquisition theories which advocate that declarative knowledge can be transformed into procedural knowledge. Meta-phonological awareness (explicit analytical knowledge of L2 sounds) consistently improves self-monitoring, accuracy and production of L3 tones (Wrembel, 2015).

The theory behind feature geometry is based on Clements' (1985) hierarchy system including feature trees, natural classes and spreading/delinking. Brown (2000) showed that L1 feature geometry influences the acquisition of contrasts in L2. Although underused in designing teaching strategies, feature geometry can provide an insightful explanation for recurring errors in learners and the reorganisation of learners' sound inventory. There is evidence in the CAPT literature that speech recognition, articulatory displays and multi-modal stimuli improve accuracy, engagement and retention (Mahdi & Al Khateeb, 2019; Ngo et al., 2024; Almusharraf, 2024).

Generative artificial intelligence has more recently been used in phonological training strategies. Initial trials show that using GPT based tutoring and AI generated storytelling feedback can improve meta-phonological awareness and prosody (Łodzikowski et al., 2025; Hirschi, 2025). Nonetheless, issues like hallucinations mean there is need to be a structure on ground. While there is evidence to support all aspects of a multi-modal generative tutoring strategy using feature geometry, none of the research has looked into a tutoring system specifically designed around feature geometry to measure learning outcomes across the entire spectrum from meta-phonological awareness to spontaneous speech.

4.1 Findings and Discussion

The numbers showed benefits for the AI tutor. For knowledge items the AI group did much better than the traditional group in learning gains and those gains stayed strong at the delayed posttest. Gains in awareness followed the same pattern. The two groups developed precision on process-related errors. The AI group had stronger gains, especially in assimilation and feature-spreading contexts. Broadly speaking, there was a moderate amount of transfer to speech however, the AI group showed increased amounts. The interaction data shows that members of the AI group utilised the feature trees very heavily and requested justifications for their actions. The stimulated recall data also suggested that the novelty of new visuals and having to verbalise their thought processes helped highlight the underlying relationships that the traditional static display did not. The questionnaire responses show that the participants had positive feelings towards the visualisations and scaffolding provided by the AI tutor. Some responses also stated that they saw the AI tutor provide very slightly incorrect explanations for the rules however these were quickly rectified within further conversation or by the teacher. This assists in facilitating Acquisition Theory because the Artificial Intelligence tutor gave declarative contributions and lots of procedural practice tasks. These tasks helped learners transform their understanding into adeptness. The amount of noticing went up because of the double coding of the AI interconnection. This supports Schmidt's theory.

The use of multimedia may have helped reduce cognitive load due to all the channels working together and the learners being able to control what was coming in to encourage active engagement with the abstract hierarchical structure. The importance of teaching learners about the geometry of features was apparent. Learners who had a good understanding of classes and limitations of processes made fewer process based errors and demonstrated more intelligent transfer. Building off previous theories by Brown and others this study provides evidence that generative AI can provide useful teaching tools. This study has been limited due to using a design aimed at intermediate L2 learners and only English was used. There was still a risk of confabulation. Future studies should look at how AI tutors affect process driven errors for learners at and other target languages.

5.1 Conclusion

The study showed that a designed multimodal generative AI tutor can give clear step-by-step teaching on phonological processes and, on the shape of sounds. Learners who interacted with the system developed stronger metaphonological awareness and unambiguous understanding than those who acquire traditional teaching. These gains transferred to enhance pronunciation precision and, to a reckonable degree, to spontaneous speech. The teacher therefore connects a lasting problem between what is taught in phonology and how it is actually taught in classrooms. It also explains how artificial intelligence, when used properly and with types of input can be a strong helper in teaching, not just a tool for simple answers. This idea can be used again and again. Works well in different situations. It offers benefits, for schools that have lots of resources and also for those that do not. The tutor thus links a long standing gap between theoretical phonology and practical pronunciation pedagogy.

5.2 Recommendations

1. Add pictorial tools to re-work the structure into pronunciation drills. AI tutors using a template based approach to reduce guess work but also allow free flowing conversation.

2. Providing a single resource for speech synthesis, articulation and written notes. This enables the learners to achieve clearer understanding of the principles via double methods. Questions are added to prompt thinking after the speech practice.

3. Understanding what the student is capable of and how much they can remember. Eye tracking or additional activities can be used to enhance the teaching methods. Activities which encourage students to work out rules themselves, by writing and self assessing their ideas, and to create active knowledge.

4. An insight for teachers into what the student is doing and where they may have gone wrong. Teachers will also receive training on AI to help explain what the tutor is saying and use it within lessons.

5. Assessing students after a period of time to ensure they can retain their knowledge and apply it in different ways. The tutor is adaptable to work with diverse languages and learners' abilities. Languages with very different phonology can be catered for.

References

Almusharraf, N. (2024). The effect of settings, educational level and tools on computer assisted pronunciation training: A meta analysis. Journal of Computer Assisted Learning. https://doi.org/10.1111/jcal.12974

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), Second language acquisition and linguistic theory (pp. 4–63). Blackwell.

Clements, G. N. (1985). The geometry of phonological features. Phonology Yearbook, 2, 225–252. https://doi.org/10.1017/S0952675700000440

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition (pp. 97–113). Lawrence Erlbaum.

Hirschi, K. (2025). Artificial intelligence generated feedback for second language intelligibility: An exploratory intervention study on effects and perceptions. Language Learning. https://doi.org/10.1111/lang.12719

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta analysis. Applied Linguistics, 36(3), 345–366. https://doi.org/10.1093/applin/amu040

Łodzikowski, K., et al. (2025). Exploring large language models for L2 metaphonological awareness tutoring. Journal of Second Language Pronunciation, 11(1). https://doi.org/10.1075/jslp.24030.lod

Mahdi, H. S., & Al Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. Review of Education, 7(3), 445–464. https://doi.org/10.1002/rev3.3165

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Ngo, T. T. N., Chen, H. J., & Lai, K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta analysis. ReCALL, 36(1), 4–21. https://doi.org/10.1017/S0958344023000113

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed framework and meta analysis. Language Learning, 69(3), 652–708. https://doi.org/10.1111/lang.12345

Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Wrembel, M. (2015). Metaphonological awareness in multilinguals: A case of L3 Polish. Language Awareness, 24(1), 60–83. https://doi.org/10.1080/09658416.2014.890340