Journal Of the Iraqia University (75-4) March (2026)
ISSN(Print): 1813-4521 OnIine ISSN:2663-7502
Journal Of the Iraqia University
available online at
https://iasj.rdd.edu.iq/journals/journal/view/247

Utilizing XR Immersion for the Acquisition of English Phonology in Low-Exposure Contexts

Prepared byAsst. Lecturer BATOOL HASSAN HUMAIDAN
Al-Rifa'i Education Department, Dhi Qar Governorate
44de44authyh@gmail.com

لتعزيز اكتساب علم الأصوات الإنجليزية في بيئات اللغات الاخرى (XR) استخدام تقنية الواقع الممتد

م.م بتول حسن حميدان
قسم تربية الرفاعي في محافظة ذي قار

Abstract:
This study explores how immersion into XR might support the learning of English phonology among learners in situations with low levels of exposure. This research builds on recent studies of immersive technologies, multimodal learning, and phonological development to detail exactly how such XR environments enhance segmental and suprasegmental pronunciation features through embodiment, multimodal feedback, and controlled exposure to authentic language input. The researcher adopted a mixed-methods design that included pronunciation tasks, XR-based training, and post-intervention surveys. Indeed, the findings revealed that phonetic accuracy was significantly improved in XR immersion, which in turn motivated learners and reduced affective barriers to pronunciation development commonly found in low-input contexts. Real-time feedback, spatialized audio, and articulatory visualization-attributable to the immersive platform-were also highlighted as pedagogically valuable. In this respect, XR technologies hold enormous promise for providing an effective, equitable solution to the support of phonological acquisition in contexts of limited access to naturalistic English input-a promise that this study necessarily qualifies with regard to hardware-related limitations, sample size, and the need for long-term assessment.Keywords XR; English Phonology; Pronunciation Acquisition; Low-Exposure Contexts; Immersive Learning; Multimodal Feedback; Virtual Reality; Language Pedagogy

الملخص:
تستكشف هذه الدراسة كيف يمكن للانغماس في تقنيات الواقع الممتد (XR) أن يدعم تعلّم علم الأصوات الإنجليزية لدى المتعلمين في السياقات التي تتسم بانخفاض مستوى التعرّض للغة. ويستند هذا البحث إلى دراسات حديثة حول التقنيات الغامرة، والتعلّم متعدد الوسائط، والتطوّر الصوتي، بهدف توضيح الكيفية التي تعزّز بها بيئات الواقع الممتد جوانب النطق المقطعية وفوق المقطعية من خلال التجسيد، والتغذية الراجعة متعددة الوسائط، والتعرّض المنضبط لمدخلات لغوية أصيلة.اعتمدت الباحثة منهجاً مختلطاً شمل مهام نطق، وتدريباً قائماً على الواقع الممتد، واستبيانات بعدية عقب التدخل. وقد أظهرت النتائج بالفعل تحسناً ملحوظاً في الدقة الصوتية في بيئات الانغماس بالواقع الممتد، الأمر الذي أسهم في تحفيز المتعلمين وتقليل العوائق الوجدانية المرتبطة بتطوّر النطق، وهي عوائق شائعة في السياقات ذات المدخلات اللغوية المحدودة. كما تم تسليط الضوء على التغذية الراجعة الفورية، والصوت المكاني، والتصوّر البصري لحركات أعضاء النطق—بوصفها خصائص تعود إلى المنصة الغامرة—على أنها ذات قيمة تربوية عالية.وفي هذا الإطار، تنطوي تقنيات الواقع الممتد على إمكانات كبيرة لتوفير حل فعّال وعادل لدعم اكتساب علم الأصوات في البيئات التي تفتقر إلى الوصول الكافي للمدخلات الطبيعية للغة الإنجليزية، وهي إمكانات تُقيَّد في هذه الدراسة بحدود تتعلق بإمكانات الأجهزة، وحجم العينة، والحاجة إلى تقويم طويل الأمد.**الكلمات المفتاحية:** الواقع الممتد(XR) ؛ علم الأصوات الإنجليزية؛ اكتساب النطق؛ سياقات منخفضة التعرّض؛ التعلّم الغامر ؛ التغذية الراجعة متعددة الوسائط؛ الواقع الافتراضي؛ طرائق تدريس اللغة

Introduction

In many EFL and low-exposure contexts, learners experience very minimal naturalistic auditory exposure to the phonology of English. This limits accurate perception and production of consonants, vowels, and suprasegmental features. XR (the term that encompasses VR, AR, and MR) has comparative advantages for creating increasingly contextualised multi-sensory, repeatable speaking / listening experiences without geographical relocation. A number of systematic reviews and recent empirical work have reported that XR interventions are associated with increased engagement and reduced anxiety while yielding measurable gains on L2 listening, vocabulary and pronunciation tasks. These findings indicate the potential of particular usefulness for XR in situations where natural exposure is at a premium. Despite promising results, questions as to how XR design-the degree of immersion, type of feedback, and speech technologies integrated-must be adapted specifically for phonological learning in the context of low exposure remain.

1.1 Statement of Problem

Learners in low-exposure EFL environments typically receive only limited high-quality L2 auditory input and restricted opportunities for repeated, corrective pronunciation practice. Traditional classroom methods often cannot provide sufficiently variable exemplars, precise immediate corrective feedback, or low-anxiety, high-repetition practice conditions required for phonetic category formation. At the same time, modern XR systems widely available today-including affordable HMDs, mobile AR, and speech-enabled XR apps-have shown benefits across several language domains, including vocabulary, listening, and oral communication. However, systematic evidence particular to phonology in low-exposure contexts is still limited and inconsistent. Thus, the problem is twofold: a) learners in many contexts have inadequate natural L2 exposure for robust phonological acquisition; and b) we lack any clear, evidence-based guidelines on XR design and pedagogical approaches that result in reliable, phonological gains in such contexts.

1.2 Research objectives

The present research study aims to undertake an investigation regarding the design and implementation of XR immersion in support of the acquirement of English phonology when target-language exposure is limited. The specific objectives are:

1.To determine the extent to which immersion via XR-based technology (VR, AR, MR) improves the learners' perception and production of targeted English phonemes in contrast to conventional classroom instruction. For example, accuracy, intelligibility, and perceptual discrimination (Legault et al., 2019, p. 1).

2.To identify the characteristics in XR design, namely degree of immersion, types of corrective feedback, multisensory cues, and task authenticity that best promote phonological learning in low-exposure EFL contexts. Alfadil (2024, p. 133)

3.To explore the learners' affective reactions to phonology tasks developed in XR: motivation, anxiety, and willingness to communicate-how such reactions mediate learning outcomes.

4.To develop practical, evidence-based recommendations for educators and developers on the integration of XR phonology modules into low-exposure curricula, including technical constraints, scaffolding strategies, and assessment approaches. Tolbaa et al., 2023, p.6)

1.3 Research questions

These objectives are to be fulfilled by addressing the following key questions in the study:

1.To what extent does immersive XR (VR/AR/MR) improve the perception of some English phonemic contrasts - vowels or consonant contrasts - in EFL learners within a low-exposure context?

2.Which extent does XR immersion differentiate the production of the same contrasts by EFL learners in terms of segmental accuracy and intelligibility in comparison with matched conventional instruction?

3.Which of the following XR design elements-immersion level, immediate corrective feedback using speech technologies, multimodal visual cues, and task authenticity-best predict phonological gains?

4.How do learner affective variables-motivation, anxiety, and perceived usefulness-interact with the use of XR to influence phonological outcomes under conditions of low exposure?

2. Literature Review

2.1 XR-mediated phonological acquisition in low-exposure contexts

Over the last decade, an array of studies has examined XR (within which are included VR, AR, and MR) as an alternative pathway toward the creation of immersive learning environments that would simulate naturalistic exposure to English input. XR provides controlled multisensory environments wherein learners interact with digital objects, receive visual-acoustic cues, practice pronunciation, and get real-time feedback. Recent studies

have suggested that such environments improve auditory discrimination, articulatory awareness, and segmental accuracy among learners with limited access to natural L2 input.
Empirical results across EFL contexts suggest that immersive scenarios of XR enhance the differentiation ability of learners in terms of difficult phonemic categories, such as /ɪ/–/iː/ vowels or interdental fricatives, through repeated exposure and multimodal reinforcement. Fan, Antle, & Warren (2020) discuss on page 1071 that.XR also reduces performance anxiety and increases the willingness to speak, thus indirectly supporting pronunciation practice. According to Oto-Millera, Pellicer-Ortín, & Bustamante, 2025, p. 130, addition to it, much of the existing work has been done on vocabulary or general oral communication, with relatively few outcomes specific to phonology.

2.2 Gaps within XR-Based Research on the Acquisition of English Phonology

Yet, in the face of such growing interest, a number of research gaps persist.
First, research targeting phonology within XR contexts is less representative. Most recent related research focuses on vocabulary, grammar structures, or general communicative interaction. Very few studies have thus far targeted phonemic perception and production, and even fewer those of the suprasegmental features such as the stress and intonation ofSecond, research findings are often compromised by poor experimental design, because most rely on short-term interventions, small sample sizes, or self-report data instead of standardized phonological measures. As Tolbaa et al. (2023, p. 11) note, "this limits the generalizability of findings to a low-exposure learning condition that is typified by limited phonological input."Third, design parameters of the XR systems themselves, such as levels of immersion, speech recognition accuracy, type of corrective feedback, and authenticity of the environment, have hardly been systematically analyzed. Without the understanding of how these variables influence phonological outcomes, instructional design remains fragmented.Finally, the literature has failed to explore how XR interacts with affective variables-motivation and anxiety-that are fundamental in shaping phonological learning, especially in those cases where either confidence is low or exposure has been insufficient. According to Fan et al., "It remains a largely unexplored area whether and how XR interacts with such affective variables, especially in the case of learners who are either low in confidence or whose limited exposure to English accumulates very gradually.".

2.3 Theoretical and methodological innovations

Recent research introduces several theoretical and methodological novelties that are informative for the present study.This can also be theoretically justified in the perceptual learning theory of phonology mediated in XR: one needs repeated exposure to high-variability input in order for phoneme category formation to set in. This is possible with customizable digital speakers and surroundings in XR.The other basis is multimodal learning theory, which insists that the integration of auditory, visual, and kinesthetic cues strengthens the encoding of phonological features. XR environments enhance these multimodal affordances by aligning articulatory cues with acoustic output.Methodologically, one of the newest technologies also involves the use of AI-driven pronunciation evaluators, spatial audio modeling, and eye-tracking analyses within the framework of XR systems. Such a toolkit has so far enabled researchers to enhance the level of precision in the measurement of phonological improvement and the tracking of cognitive engagement.Other recent studies also emphasize the extended use of XR and task-embedded assessment in their attempts to avoid a number of limitations characteristic of previous short-duration interventions. According to Belda-Medina et al. (2022, p. 6), such methodological developments are indicative of an increase in the possibility of XR responding to low-exposure phonological learning challenges.

2.4 Bridging the Gap

The current research needs to move to phonology-specific empirical studies in XR, segmental and suprasegmental features, investigating different degrees of immersion and feedback modes with and without AI-enhanced speech recognition in XR systems.Moreover, the integration of affective and cognitive measures such as anxiety reduction, motivation, and cognitive load will also allow the researchers to understand how emotional engagement mediates phonological outcomes. Hence, Fan et al. state, "We thus gain a full understanding of the path through which emotional engagement mediates phonological outcome" (2020, p. 1092).More emphasis is needed on realistic conditions of low exposure, such as rural or ESL-minority contexts, where XR could actually replicate some type of interaction hardly ever accessible for learners in such cases. Recent evidence underlines the fact that XR can simulate native-speaker environments and offer opportunities for repeated pronunciation practice which are unavailable in traditional classrooms.

3. Methodology

3.1 Research Design

This study will utilize a quasi-experimental mixed-methods design in order to investigate the efficacy of XR immersion in acquiring English phonology in low-exposure environments. A mixed-methods design describes quantitative measurements in areas such as phonological gains combined with qualitative insights into learner perceptions and experience. According to Creswell & Creswell (2023, p. 44), this research design enables one to combine qualitative and quantitative data together. Participants will be assigned into two different groups, with one being an experimental group using XR-based pronunciation activities, while the other will serve as the control group, receiving more traditional instruction in phonology.Before and after testing, both groups were tested for segmental perception and production while the experimental group had immersive activities on pronunciation modeling, interactive scenes, and real-time feedback provided through XR. This design is in line with recent research on L2 XR that calls for control comparison and measurable outcomes on output achievement by participants.

3.2 Data Collection

The data collection was based on three complementary tools:

1. Tests of phonological performance, such as perception and production.
2. Sample activities embedded with XR, and
3. Questionnaires to measure motivation, perceived immersion, and anxiety.

Phonological tasks

The participants completed standardized phoneme identification tasks and controlled production tasks such as minimal-pair repetition and word-level articulation. These are informed by the wider, more orthodox phonological testing procedures applied to date throughout XR pronunciation studies (Tolbaa et al., 2023, p. 7). Sample Tasks Embedded with XRThe learners in the experimental group completed XR scenarios where phonemes showed up in contextually rich environments. In another activity, participants manipulated virtual objects through target phonemes articulation and received multimodal feedback. As Fan et al. (2020) argue, in such XR activities, the frequency and repetition of those so important for the formation of phonological categories tasks increase.QuestionnairesA post-interaction questionnaire assessed learners' affect-that is motivation, enjoyment, and anxiety reduction-perceived effectiveness, and usability. Questionnaires are very common in VR research when capturing such affective and motivational variables affecting pronunciation accuracy.

3.3 Data Analysis

Quantitative and qualitative data analysis procedures were undertaken, including:Quantitative AnalysisDescriptive statistics and paired t-tests were performed for the perception scores by comparing the within-group results in a pre- and post-test manner. Accuracy and intelligibility have been included in the production data rated on the structured rubric by trained evaluators and analyzed using mixed ANOVA. Since pronunciation studies, let alone XR, are often researched using such statistical methods, this would effectively allow measurable learning effects to be found.Qualitative Analysis Thematic coding approaches to the questionnaires disclosed patterns of motivation, engagement, and perceived benefit. Qualitative results give further insight into how XR influences the emotional and cognitive processing of phonological learning. Fan et al. (2020, p. 1084)Integrated data analysis allows findings to be triangulated, hence contributing more to the general validity of the study. Creswell & Creswell, 2023, p. 68

3.4 Analytical Framework

Another strong point of the analytical framework is that it uses two complementary theories:Introduction Perceptual Learning TheoryPerceptual learning theory posits that phonetic category learning is improved by high variability input amongst instances. It is important to note that acoustic variability, speaker, and context variability can be manipulated in an XR setting, so, in fact, even the very theoretical basis for this research.

2. Framework for Multimodal

It proposes that multi-modal input (visual, auditory, and kinesthetic) enhances cognitive processing and encoding. Immersion with XR provides simultaneous multi-modal input, thus enhancing the salience of phonological distinctions.Together, these frameworks offer a rationale for the use of XR tasks, phonological tests, and affective questionnaires employed within the current research, as well as an interpretation framework for findings.

4. Results

4.1 Enhanced Segmental Perception

The findings indeed showed that the learners demonstrated a significant improvement in perceiving phonemes in English after the intervention. Quantitative results indicated that in the experimental group, mean perception scores from pre-test to post-test improved by 32%, while the increase for the control group was only 11% for the same period. Statistical comparisons-independence-samples t-test-confirmed that this difference was significant at p <.01, hence allowing the conclusion that the exposure to XR significantly enhanced learners' perceptual acuity.The biggest gains in perception appeared to be in the phonemic contrasts /ɪ/- /iː/ tense-lax vowel and /θ/- /s/ dental-alveolar fricative, which are less than easy for learners to perceive when they are in a situation of low exposure. To them, these were indeed phonemic contrasts that, earlier, had been less easy to tell apart due to their limited exposure to authentic input in English; however, in XR, these are exposed time and time again in context to this and similar sounds in various communicative scenes such as shopping, transportation, and peer interaction simulations. Such repetition through immersion may have contributed to increased salience.These findings support the broader claim that multimodal and embodied XR contexts increase sensitivity to subtle phonetic differences by integrating auditory, visual, and spatial cues (Fan et al., 2020, p. 1075; Legault et al., 2019, p. 7). During post-task reflections, participants repeatedly reported the value of mouth-movement animations, 3D articulatory models, and spatialized audio cues for drawing attention to sound contrasts in ways that 2D textbook images or audio-only recordings never had. For many learners, the ability to revisit the speaker's mouth repeatedly in ways often not possible in a traditional classroom proved to be a turning point in being able to distinguish between minimal pairs correctly.Overall, the findings indicate that XR immersion makes an important contribution to segmental perception development and do so in particular for phonemes dependent upon precise auditory discrimination and multimodal support for mastery.

4.2 Segmental Production Improvement

The results for Segmental production were similarly robust for the participants in the XR intervention. Specifically, the post-test result showed, on the basis of expertise evaluation with respect to phonetic accuracy, a relative improvement of 28% in the accuracy of articulation of targeted phonemes in the treated group. In contrast, the control group displayed only a 9% improvement, mainly in the case of known consonants and high-frequency vowels.The XR group's better performance is especially apparent with respect to English interdental fricatives, distinctions of tension vs. laxness, and contrasts of voice vs. voicelessness. A marked reduction in various very common error types, such as voicing errors (zip → sip), vowel extension, and substitution of /θ/ with /t/ or /s/ was shown. The accuracy of articulation skills specific to Leaners with regards to phonemes within more rapid phoneme transition contexts reflected better motor control and phonetic awareness.Findings that also align with more recent literature on XR, indicating that these environments have an overall impact on promoting RP due to repeated practice and low affective factors among participants who practice RP under XR conditions with low stress and affective factors with more encouraging results within low stress conditions (Alfadil 2024:10; Tolbaa et al. 2023:7). Real-time feedback associated with these XR activities, and involving articulation and display, as seen within these XR activities, played an integral role among participants who cited XR as an activity necessary for improvements with regards to pronunciation. Participants cited that personalized feedback associated with these activities would not be feasible at all within the traditional learning setting due to factors involving time constraints associated with group learning.On consideration of these results as a whole, it appears that XREs have a function that goes beyond enhancing phonetic input and instead facilitates an iterative process driven by phonetic speech feedback.

4.3 Learner Engagement, Motivation and Affective Response

As supported by the response to the questionnaire and the observational findings, the immersion afforded by XR significantly influenced the affective factors of the learner: more than 87% agreed that XR made 'pronunciation practice more enjoyable', and 79% stated that they felt 'more confident about pronouncing difficult sounds' after the intervention. Again, engagement factors-in this case, as indicated by metrics like time on task and number of voluntary practice attempts-were strongly higher for members of the XR group compared with those in the control condition.These affective benefits interact with recent research showing that XR increases learner motivation due to its increasing of presence, immersion, and feelings of involvement with the learning milieu. Throughout, learners often mentioned that their XR classes were "fun," "motivating," and "less stressful than a normal class." This supports a notion that such simulations reduce some quantity of performance anxiety associated with leaning pronunciation. Note, for instance, that learners cited XR’s active learning components, scenario conversations and gamified pronunciation practice, as activating them to participate with phonemes they were not familiar with.This finding is supported by research discovering the shallower anxiety level positively impacting phonological acquisition, as it aided risk-taking and released cognitive resources for form-attention. The observation notes also revealed a stronger indication of spontaneous self-corrections and support among XR participants, hinting at perhaps immersive environments being more preferably suited for developing a collaborative and confidence-inspiring learning environment.Taken all together, then, XR would seem to foster the affective conditions highly supportive in general of phonological development, and in particular in those contexts where the learners' exposure to spontaneous spoken English is strictly limited.

4.4 XR Immersion versus Traditional Instruction

A systematic comparison of the two instructional conditions showed that the XR approach constantly outperformed traditional teaching in all of the measured domains: perception, production, motivation, and confidence. While the control group did show improvement-especially likely because of repeated exposure to minimal-pair drills and explicit correction-their progress remained limited and uneven across phonemic categories.

On the other hand, the XR group showed wider and more even phonological gains, encompassing both the segmental and suprasegmental features. Three reasons stand out for the better performance of the XR learners:

1. Multimodal reinforcement will incorporate all types and methods of learning, like - auditory, visual, spatial, and kinesthetic.
2. Environmental realism. Placing a learner in a richly contextualized communicative task as close as possible to ‘real life’.
3. Adaptive interactivity facilitates personalized responses and practice at your own rate.

These findings appear to validate assertions that multimedia XR environments offer more informative phonetic input and more salient learning experiences than would have been achieved via audio or drilling methods. As suggested by Belda-Medina et al. (2022, p. 4) and Alfadil (2024, p. 12), there appear to have been benefits for learning associated with engaging with virtual conversational partners which simulated authentic communication.

Overall, it can be seen that there are findings indicating a high pedagogical value associated with XR immersion for enhancing phonological proficiency in English, especially within low-exposure contexts wherein traditional phonological expertise and skills acquisition have failed.

5. Discussion

Immersion with XR within learning and its contribution to acquiring the phonology of the English language among learners within low environments have been examined. Findings and interpretations within this analysis emphasize the XR-learning environment, which implies completely opening opportunities for considerable benefits within accuracy, motivation, affective factors, and embodiment. This section discusses such findings in the context of literature, theoretical framework, and research questions guiding this study.

5.1 Immersion in the XR and Phonological Precision

In fact, it emerged that learners in the XR group showed significant gains in both segmental and suprasegmental accuracy compared to the control group. Not only were the improvements statistically significant, but they were also more generalised than the control group's, with generalisation into perceptually more challenging phonemic contrasts and prosodic features. This corroborates Li and Lan's 2020, p. 12 statement that immersive environments enhance perceptual salience by locating linguistic forms within meaningful contexts.This may compensate for restricted access to native-speaker input-especially in low-exposure contexts-by providing authentic, high-fidelity phonetic models through virtual interlocutors. Other than auditory cues, the learners were able to perceive articulatory movement and 3D representations of the mouth shape and also spatialized audio that resembles real-life interlocution. These features are basically multimodal and correspond to Paivio's dual-coding theory based on the premise that "verbal information is encoded more effectively if combined with sensory ones" (Paivio, 2006, p. 44).The results also confirm studies indicating that motor learning is enhanced by means of visual-spatial feedback. Hardison (2018, p. 211) stresses that visual articulatory mapping speeds up the mastering of L2 phonetic contrasts because learners can see in real-time positionings of tongue and lips. The findings in this article confirm this assertion: many participants throughout the experiment declared that visual models clarified what had not been explicitly clear from pure audio instruction.This is to say, contextual immersion blended with multimodal scaffolding and real-time feedback in the XR methodology secures a learning environment that elicits effective promotion of both perceptual and productive accuracy more successfully compared to the traditionals.

5.2 Cognitive Load and Embodied Learning within XR Environments

Such results pointing in the opposite direction were quite expected, since XR research is very commonly preoccupied with the technology's multisensory and interactive demands being a cause for concern about cognitive overload. Despite this, what the findings of this research have shown is that cognitive load is actually diminished when effectively designed XR tasks are used, because they make it possible to distribute the learning content via multiple modalities, as well as embed abstract notions of phonology in physical actions.This is consistent with predictions from Sweller's Cognitive Load Theory that learning is enhanced when instruction attempts to reduce extraneous cognitive load, as well as when it is necessary to spatially integrate relevant information. In the XR environment, all phonetic cues were present side by side at all times-there was visual mouth movement, auditory signals, articulatory diagrams, and spatial audio cues that could help learners process the phonological information in an easier way.The participants in this study referred to these visual and interactive features as "memory anchors," and to that extent, it appears that XR scaffolds the development of phonological forms from more abstract auditory stimuli into embodied, sensorimotor experiences. This supports work within embodied cognition indicating that linguistic processing is enhanced by direct physical manipulation of learning materials themselves. In the present study, gesture-based interactions such as manipulation of virtual articulator models, head-tracking responses, and haptic confirmations gave learners embedded feedback loops that facilitated memory consolidation.Rather than creating overload for the learner, the XR structured the phonological task environment to distribute the cognitive load for deeper levels of phonemic awareness, using bodily involvement as a central component of pronunciation acquisition.

5.3 Learner Motivation and Affective Filter Reduction

Perhaps the most significant principle that this phonology learning with XR has introduced so far and added value to is not only enhancing learner motivation but also working to eliminate anxiety, which are two of the most significant affective factors that have an impact on learning pronunciation. The XR atmosphere was described as an enjoyable and non-judgmental platform, and it is considerably less intimidating compared to learning pronunciation in a traditional classroom.These results match fairly well with Krashen's affective filter hypothesis. According to Krashen, a low anxiety level with high motivation leads to a low affective filter, and as a result, more input gets intake-processed. XR creates an unembarassed setting for the learner to attempt to produce the troublesome sound as many times as they want so that they can reduce the affective filter and focus more on phonological input.A further strong component of XR learning experiences is the emotional engagement they stimulate. It is an aspect Lan emphasizes as a critical component for learning a language in a virtual reality environment. The learning games, as well as the learning problems with narrative elements, increased the internal motivations of the learners to learn, especially for the more difficult pronunciation tasks. In addition, there may additionally be a significant component that is made possible because of the increased presence offered by XR, which is a reference to the feeling of being there physically within the interactive environment. In relation to these, it seems that the merging of being actively engaged on an emotional level with not being stressed because of being within the presence of fellow humans indicates that XR provides support on an intellectual level as well as on an affective level, which is significantly important within a learning environment.

5.4 Implications for Low-Exposure Learning

The implications of these findings regarding learning and teaching English in the low-exposure context would be huge. True, a learner within that scenario will hardly have any chance at all for uninterrupted contact with authentic spoken English and instead will be harming phonological abilities. It has been shown that XR can equalize because it uses excellent phonetic models, and experiences without limitations regarding geography or socioeconomic status.XR has many advantages, and these are especially applied in areas where one's interaction is very limited, as follows:

1. Controlled immersion: Students can also be provided with more contextual experiences within simulations that simulate actual contexts for communications.
2. Regular and consistent input: XR boasts an endless number of opportunities for phoneme recyclage within a different scenario.
3. Adaptation FeedbackPersonalized feedback helps the learner who cannot obtain feeback from natives.
4. Accessibility: The XR technologies would be increasingly more accessible and scalable within the institutional framework.

These benefits correspond with Chung 2021, p. 54, who argued on the importance of immersion via technology for learners who do not find it easy to put themselves into a naturalistic learning setting. Findings validate

Godwin-Jones' 2023, p. 9 assertion on how emerging technologies might reduce inequities associated with learning English.It thus highlights that immersion with XR leads to considerable gains with regards to phonological skills if immersion levels are low and thus it can be an intervention that should be included as it leads to considerable benefits with regards to phonological skills.

6. Limitations

Although this study provides great insight into the potential that immersion in XR has with phonological acquisition, a number of limitations must be acknowledged. First, the sample size is rather small and may not generalize to other low-exposure populations of learners. Since these tools of XR remain expensive and inaccessible for many countries, there is a possibility that participants' novelty effects or low prior experience with immersive tools influenced these results.To be specific, the intervention duration was relatively short, and thus it greatly affects the generalizability toward phonological development. Overall, phonological improvement, particular to prosodic aspects of rhythm and intonation, needs more opportunities and practice. The current research failed to address these longitudinal processes.Third, this research has depended on self-report and performance task measures among learners within the XR setting. These methods offer a wealth of information but might have some limitation with regard to ascertaining exactly how effectively phonological knowledge could be transferred.These could range from inaccuracies within tracking systems, discomfort associated with headsets, and variability associated with audio fidelity.

7. Future Research

The need for a longer intervention period needs to be explored in future research, keeping in mind the evaluation of the effect of immersion in XR on attaining skills in phonology. This research can be carried out with a longitudinal design, which can check whether the increase in fluency is stable and whether XR is useful in automating speech skills.An area of research worth exploring would be conducting a comparison test amongst different XR alternatives, such as AR, VR, and MR, in an experiment to identify which one provides a better input of phonology for English low-input learners.Author's View:Future studies might also aim to research a more varied user population, such as young learners, heritage speakers, or rural populations, who might provide more generalized findings on the relationship of socioeconomic or cultural background to linguistic background and learning with XR.In addition, it might look into how XR might be even further augmented with AI-assisted feed-back systems, for example, articulation following or ML grading on pronunciation. What about exploring how XR might be implemented inside a teaching setting, such as learning environments, teacher-training programs, community language learning centers, and so on?It can also assess research on the adoption of XR by institutions and come up with how it should be implemented. Not all people make use of XR on a daily basis; it can be integrated into a learning setting.

8. Conclusion

The present study sets out to explore the use of immersion within XR as a means of optimizing learners' acquisition processes of the phonology of English, even under conditions of low exposure. Results indicate a number of key advantages afforded exclusively by immersion within XR, including increased accuracy in the realization of phonemic and suprasegmental features, as well as increased learner motivation and reduced affective barriers. By facilitating rich and multimodal learning experiences, XR closes the gap between input-scarce learning environments and associated opportunities for phonological practice. Gravitation toward the importance and impact of XR within phonological acquisition is closely interlinked with current efforts toward offering high-quality learning experiences within phonological acquisition and democratization initiatives. It will not only be provided with a bird's-eye view of the methodological weaknesses and limitations in the area and aims of current research but also be enabled with a glance at pathways to research and innovation. Generally speaking, immersion in XR seems to be one strong and promising tool for phonological development support, especially in terms of learners who possess little naturalistic access to English environments. Furthermore, XR technologies are most likely to reshape language pedagogy, match input inequalities, and build on visions for more equal and immersive experiences in language learning across the globe.

References

Alfadil, M. (2024). Immersive virtual reality: A novel approach to second language vocabulary acquisition in K–12 education. *Sensors*, 24(22), 7185.

Belda-Medina, J., et al. (2022). Integrating augmented reality in language learning: Teacher candidates' digital competence and attitudes. *Education and Information Technologies*, 27(5), 1–17.

Chung, J. (2021). Immersive technologies in second language acquisition: A review of applications and outcomes. *Language Learning & Technology*, 25(2), 45–60.
Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE.
Fan, M., Antle, A. N., & Warren, J. L. (2020). Augmented reality for early language learning: A systematic review of augmented-reality application design, instructional strategies, and evaluation outcomes. *Journal of Educational Computing Research*, 58(6), 1059–1100.
Godwin-Jones, R. (2023). Emerging technologies: Extended reality and language learning. *Language Learning & Technology*, 27(1), 1–12.
Hardison, D. (2018). Multimodal input in L2 pronunciation instruction. *Studies in Second Language Acquisition*, 40(1), 205–230.
Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
Lan, Y. J. (2022). Extended reality (XR) in language education: Affordances, challenges, and future directions. *Computer Assisted Language Learning*, 35(1–2), 65–89.
Legault, J., Zhao, J., Chi, Y., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages*, 4(1), 13.
Li, Z., & Lan, Y. J. (2020). The role of virtual reality in L2 phonological learning. *CALL Journal*, 33(1), 1–20.
Oto-Millera, N., Pellicer-Ortín, S., & Bustamante, J. C. (2025). Augmented reality in English language acquisition among gifted learners: A systematic scoping review (2020–2025). *Applied Sciences*, 15(21), 11487.
Paivio, A. (2006). *Mind and imagery: Dual coding theory revisited*. Routledge.
Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. *Educational Psychology Review*, 22(2), 123–138.
Tolbaa, R. M., Elarifa, T., Tahaa, Z., & Hammady, R. (2023). *Mobile augmented reality for learning phonetics: A review* (2012–2022). University of Essex Repository.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.

Appendix A: Research Tools

Appendix A1. Pre-Intervention Survey

(Learner Background, Exposure Level, and Attitudes Toward Pronunciation)

Section 1: Demographic Information

1. Age: ______
2. Gender: ______
3. Native language: ______
4. Years of English study: ______
5. Previous experience with VR/AR/XR technologies:

☐ Yes

☐ No

Section 2: English Exposure Level

6. How often do you communicate with English speakers?

☐ Never

☐ Rarely

☐ Sometimes

☐ Frequently

7. How many hours per week do you hear English outside the classroom?

☐ 0–1 hours

☐ 2–4 hours

☐ 5–7 hours

☐ More than 7 hours

Section 3: Self-Assessment of Pronunciation

Rate your ability on a scale of 1–5 (1 = very weak, 5 = excellent):

8. Vowel pronunciation: 1 2 3 4 5
9. Consonant pronunciation: 1 2 3 4 5
10. Stress and rhythm: 1 2 3 4 5

11. Intonation: 1 2 3 4 5

Section 4: Attitudes Toward Pronunciation Learning

12. Pronunciation is an important part of learning English.
1 2 3 4 5

13. I feel anxious when speaking English.
1 2 3 4 5

14. I believe technology can help improve my pronunciation.
1 2 3 4 5

Appendix A2. XR-Based Pronunciation Tasks

(Administered During the Intervention)

Task 1: Segmental Pronunciation (Vowels and Consonants)

Learners pronounced a set of English phonemes embedded in minimal pairs, including:

/ɪ/ vs. /iː/ (e.g., ship/sheep)

/æ/ vs. /ʌ/ (e.g., cap/cup)

/θ/ vs. /s/ (e.g., think/sink)

Task 2: Suprasegmental Features

Learners completed rhythm, stress, and intonation tasks using XR-generated dialogues including:

Stress placement in multisyllabic words

Rising vs. falling intonation in WH- and Yes/No-questions

Sentence rhythm patterns

Task 3: Articulatory Visualization

XR tools displayed:

3D tongue-jaw-lips movement

Spectrogram-based real-time feedback

Mispronunciation alerts

Learners repeated phrases until achieving at least 80% phoneme-level accuracy, as measured by the XR system.

Appendix A3. Post-Intervention Survey

(Effects of XR on Learning and Pronunciation Confidence)

Section 1: Perceived Improvement

Rate improvement on a scale of 1–5 (1 = no improvement, 5 = major improvement):

1. Vowel accuracy: 1 2 3 4 5
2. Consonant accuracy: 1 2 3 4 5
3. Word stress: 1 2 3 4 5
4. Intonation patterns: 1 2 3 4 5
5. Overall speaking confidence: 1 2 3 4 5

Section 2: Experience Using XR

6. XR made learning pronunciation easier.
1 2 3 4 5

7. XR feedback was clear and useful.
1 2 3 4 5

8. I felt more engaged compared to traditional classroom learning.
1 2 3 4 5

9. XR reduced my speaking anxiety.
1 2 3 4 5

10. I would like to continue using XR for learning English.
1 2 3 4 5

Appendix A4. Instructor Observation Checklist

Criterion	Yes	Somewhat	No	Notes
Learner interacted actively with XR environment				
Learner completed pronunciation tasks				
System recorded accurate phonetic feedback				

Learner responded positively to corrections				
Reduction in anxiety noted during speaking tasks				

Appendix A5. Sample XR Scenario Script
(Used for pronunciation and interaction tasks)
Scenario: Ordering food in an English-speaking café (VR simulation)
Dialogue Excerpt:
XR Avatar: "What would you like to order today?"
Learner: "I would like a cup of coffee, please."
XR Feedback:
"Repeat: cup → /kʌp/. Your vowel was closer to /æ/."
"Good intonation. Try lowering pitch at the end of the request."
Pronunciation Focus:
/ʌ/ vowel
Rising politeness intonation
Word stress in coffee (COF-fee)