SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(1), 529.
https://doi.org/10.62617/mcb529

Article

Analyzing the kinematics of speech: Improving English language proficiency through articulation and movement patterns

Xiaoyan Jing

Shandong Shengli Vocational College, Dongying 257000, China; XiaoyanJing04@outlook.com

CITATION

Jing X. Analyzing the kinematics of speech: Improving English language proficiency through articulation and movement patterns. Molecular & Cellular Biomechanics. 2025; 22(1): 529.
https://doi.org/10.62617/mcb529

ARTICLE INFO

Received: 13 October 2024
Accepted: 1 November 2024
Available online: 6 January 2025

Abstract: This study investigates the impact of improving articulatory kinematics and body movement patterns on English language proficiency among learners, focusing on the integration of speech articulation and non-verbal communication. The research uses advanced technologies such as motion capture and electromagnetic articulography (EMA) to explore how targeted kinematic feedback and movement-based training enhance pronunciation, fluency, and overall communicative competence. A total of 67 participants, with varying levels of English proficiency, underwent a four-week intervention designed to improve articulation through kinematic visualization and refine non-verbal communication through gesture training. The results indicated significant improvements in key articulatory metrics, including a 12.67% increase in tongue velocity and a 16.71% improvement in lip displacement, though these changes were not statistically significant ($p > 0.05$). Pronunciation accuracy improved notably, with F1 and F2 formant frequencies showing statistically significant reductions for vowels such as /æ/ ($p = 0.024$) and /iː/ ($p = 0.005$). The study also found that speech fluency increased significantly, with participants showing a 14.50% increase in speech rate ($p = 0.008$) and a 27.23% reduction in pause frequency ($p = 0.011$). Non-verbal communication metrics also improved, with gesture frequency increasing by 40.49% ($p = 0.013$) and gesture-speech synchronization improving by 25.98% ($p = 0.028$). Additionally, strong correlations were found between kinematic improvements and overall language proficiency, with tongue velocity ($r = 0.72$, $p = 0.002$) and pronunciation accuracy ($r = 0.80$, $p = 0.0005$) exhibiting the highest correlations.

Keywords: electromagnetic articulography; motion capture; articulatory kinematics; body movement patterns; tongue velocity; language proficiency; gesture-speech synchronization

1. Introduction

Language Proficiency, particularly in a second language, is a multifaceted skill that involves mastering verbal and non-verbal communication [1–3]. For learners of English, achieving fluency extends beyond merely acquiring vocabulary and grammar; it requires precise control over speech production mechanisms and the ability to convey meaning through synchronized Body Movements (BM) [4,5]. This study focuses on speech kinematics, articulatory movements that produce sound, and the role of non-verbal communication, such as gestures and facial expressions, in enhancing English Language Proficiency (ELP).

The process of producing speech is highly complex, involving the coordination of various articulators—such as the tongue, lips, and jaw—to form the distinct phonetic sounds of a language [6–8]. These articulatory movements can present a significant challenge for non-native English speakers, mainly when English phonemes do not exist in their native languages [9–11]. For example, producing English interdental fricatives like /θ/ in "think" or /ð/ in "this" requires specific tongue placement and airflow that may be unfamiliar [12,13]. Improving these movements, therefore, is critical for developing accurate pronunciation and fluency [14,15].

However, language proficiency is not limited to verbal articulation alone. Non-verbal cues such as gestures, facial expressions, and eye contact play a crucial role in effective communication, adding nuance, emphasis, and emotional context to speech [16–18]. Research has shown that learners who incorporate these BMs into their speech demonstrate higher engagement, fluency, and confidence. Synchronizing gestures with speech enhances spoken language's clarity [19] and naturalness, making it easier for listeners to understand and engage with the speaker [20,21].

This study seeks to bridge the gap between verbal articulation and non-verbal communication by examining how targeted training in both areas can improve ELP. Leveraging modern technological tools such as motion capture and electromagnetic articulography (EMA), the study provides a detailed analysis of how kinematic feedback—visualizing and correcting articulatory movements—can enhance pronunciation. Additionally, the study explores the role of BM in achieving fluency and communicative competence, emphasizing the integration of verbal and non-verbal elements in language learning. The central hypothesis of this research is that combining kinematic feedback with movement-based training will significantly improve the articulatory precision and overall fluency of English language learners. By analyzing the effects of this training on key kinematic metrics—such as tongue velocity, lip displacement, and jaw coordination—alongside improvements in non-verbal communication, this study aims to provide a comprehensive approach to enhancing language proficiency. The findings of this research will not only contribute to the growing body of knowledge on speech kinematics and offer practical insights for educators and language learners aiming to improve their communicative effectiveness in English.

The objectives of the work include:

(a) To analyze the role of speech kinematics in improving English pronunciation;
(b) To evaluate the impact of kinematic feedback on articulation;
(c) To assess the role of BM in language fluency;
(d) To measure the synchronization between verbal and non-verbal cues;
(e) To identify correlations between kinematic improvements and language proficiency.

The paper is organized as follows: Section 2 presents the theoretical framework; Section 3 presents the methodology; Section 4 presents the results; Section 5 presents the discussion; and Section 6 concludes the paper.

2. Theoretical framework

2.1. Speech kinematics

Speech kinematics refers to studying the movement patterns of the speech organs (articulators) involved in sound production. These articulators include the tongue, lips, jaw, soft palate, and larynx, which work in a highly coordinated manner to produce the complex sounds necessary for speech. The study of speech kinematics focuses on understanding how these organs move in space and time to form specific speech sounds, emphasizing the physical processes underlying articulation, which can be measured through biomechanical methods. By analyzing these movements, researchers can understand how subtle variations in articulatory gestures lead to different phonetic outcomes, which are essential for accurate pronunciation in any language.

In English language learning, speech kinematics plays a vital role in understanding how non-native speakers acquire and produce English phonemes, many of which might not exist in their native languages. For instance, the production of English sounds, such as the interdental fricatives "th" (/θ/ and /ð/), requires specific tongue and lip movements that may be unfamiliar to learners from linguistic backgrounds where these sounds are absent. Through kinematic analysis, teachers and learners can pinpoint the precise movement patterns needed for these sounds and develop targeted exercises that improve articulation accuracy.

Technological advancements, such as motion capture, EMA, and ultrasound, have enabled detailed observation and analysis of articulatory movements during speech. These tools allow for the precise measurement of how fast, how far, and in what direction the articulators move, providing objective data on the kinematic processes involved in speech production. Such data can identify articulation errors in language learners and provide visual feedback, showing them exactly where their tongue, lips, or jaw movements deviate from the target.

Moreover, speech kinematics is crucial for pronunciation and understanding prosody—elements like intonation, stress, and rhythm—which are also shaped by how articulators move concerning one another. For example, variations in pitch and stress patterns in English often depend on the coordinated movement of the vocal cords and other articulators, and mastering these kinematic nuances is vital for achieving fluency and natural-sounding speech. Thus, incorporating kinematic analysis into language learning provides a more scientific and practical approach to improving pronunciation and communicative competence.

2.2. Articulation and movement in speech

Articulation refers to the precise movement of speech organs—such as the tongue, lips, teeth, palate, and vocal cords—that shape airflow and produce distinct sounds required for speech. These articulatory movements form the basis of phonetic production, with each movement resulting in a unique combination of sound qualities like pitch, tone, and rhythm. Articulation is a dynamic process involving coordinated muscular actions across various parts of the vocal tract to create the specific phonemes that form words and sentences in a language. Understanding these intricate movements is essential for analyzing and improving speech production, especially for language learners.

In speech production, the articulators work synchronously to control how air is expelled from the lungs and modulated through the vocal tract. For example, the tongue plays a critical role in shaping vowels and consonants by changing its position relative to the roof of the mouth, teeth, and lips. The lips are similarly significant, particularly for labial and bilabial sounds like /p/, /b/, and /m/, where closure and release of air pressure between the lips create specific sounds. The jaw and soft palate movements also contribute significantly to modulating sound, determining the resonance and clarity of the speech produced. By altering these articulators' movement, shape, and contact, we can produce a vast array of speech sounds.

In language learning, difficulties often arise when non-native speakers must produce unfamiliar sounds that require precise articulatory movements. For example, learners of English who speak languages without certain sounds may find it challenging to articulate phonemes like /r/ and /l/, which are produced by distinct tongue positions. Articulatory training for such learners focuses on improving the coordination and timing of these movements, ensuring that sounds are formed correctly and fluently.

Movement patterns in speech are not limited to individual sounds but extend to sequences of sounds in words and sentences. Articulators must transition smoothly between different positions, allowing for the fluid production of connected speech. This fluidity separates novice language learners, who may produce sounds in isolation, from more proficient speakers, who can seamlessly articulate complete sentences with proper rhythm and intonation. For instance, the transition between the vowel sounds in "go" and "on" requires the tongue and lips to adjust position rapidly and efficiently, maintaining natural flow without breaks or stumbles.

In addition to the physical articulation of sounds, BM, such as facial expressions and gestures, also influence speech production and perception. These non-verbal movements often accompany spoken language and reinforce or clarify meaning, playing an essential role in communication. For example, raising eyebrows or gesturing with hands can emphasize certain words or add emotional context, making speech more expressive and engaging. Studies have shown that integrating these movements into language instruction can improve learners' ability to grasp and produce correct intonation patterns, leading to better comprehension and proficiency.

2.3. The role of BM in language learning

BM plays a crucial role in language learning by providing learners with additional channels of communication that support and reinforce verbal expression. These movements, often called non-verbal cues, include gestures, facial expressions, eye contact, and posture, all of which contribute to effective communication. While language learning is often centered around verbal elements such as grammar, vocabulary, and pronunciation, the inclusion of BM enriches the learning experience, aiding in both comprehension and production of speech. These non-verbal signals can enhance linguistic competence by helping learners grasp meaning, contextualize speech, and convey emotions, fostering a more holistic approach to language acquisition.

Gestures are one of the most widely studied forms of body movement in language learning. They serve as communicative tools and cognitive aids, helping learners organize their thoughts and express complex ideas. For example, iconic gestures, which visually represent the meaning of words or phrases, can provide learners with a visual association to reinforce the meaning of new vocabulary or abstract concepts. Research has shown that when learners use gestures while speaking, it helps them better retain information, especially when dealing with challenging grammar structures or unfamiliar vocabulary. This is because gestures engage multiple sensory modalities, creating a more profound, more embodied learning experience that strengthens memory retention and recall.

In addition to gestures, facial expressions and eye movements significantly influence language learning. Facial expressions provide emotional context to speech, helping learners interpret the tone and mood of a conversation. For instance, a smile, frown, or raised eyebrow can alter the meaning of a sentence, adding layers of nuance that words alone may not convey. These expressions also serve as feedback mechanisms during conversation, indicating understanding, confusion, or agreement, which is essential for maintaining conversational flow and learning social aspects of language use.

Eye contact, another critical non-verbal cue, is essential in regulating communication dynamics. In many cultures, maintaining eye contact signals engagement, attentiveness, and confidence, whereas avoiding eye contact might suggest uncertainty or discomfort. For language learners, mastering the subtle rules of eye contact can help them better navigate conversations, allowing for smoother interaction with native speakers and fostering a sense of connection. This is especially important in classroom settings, where learners can gauge their instructor's reactions and adjust their speaking behavior accordingly.

Posture and body orientation further contribute to effective communication by expressing openness or closedness in interaction. Learners who adopt an open posture—facing their conversation partner, using expansive gestures, and standing or sitting upright—tend to come across as more confident and approachable, which can boost their conversational success. This is particularly important in second-language acquisition, where the learner's confidence plays a significant role in their willingness to practice and engage in conversations. Conversely, a closed posture, such as crossed arms or slouched shoulders, may convey hesitation or withdrawal, potentially hindering learning by reducing engagement with others.

Beyond individual movements, the overall coordination of BM and speech is critical in achieving fluency. Non-verbal behaviors, when synchronized with speech, help learners achieve a natural rhythm, intonation, and flow, which are essential components of proficient language use. For example, synchronizing hand gestures with speech can enhance clarity and emphasize key points, making the message more comprehensible. Instructors can utilize these movements in language teaching by modeling natural speech patterns that integrate verbal and non-verbal elements, thus providing learners with comprehensive examples of effective communication.

Moreover, BM can bridge linguistic gaps for language learners, especially when they struggle to find the correct words. Gestures, in particular, can serve as compensatory strategies, allowing learners to convey meaning even with limited vocabulary. For instance, a learner who cannot recall the word "circle" might draw an imaginary circle in the air with their hand, successfully communicating the concept without relying on verbal recall. This use of gestures not only aids communication but also reduces the cognitive load on the learner, providing a more fluid interaction despite gaps in language proficiency.

3. Methodology

3.1. Participant demographics

The study involved 67 participants, carefully selected to represent diverse backgrounds regarding language proficiency, age, and cultural exposure to English. Among the participants, 39 were male and 28 were female, with ages ranging from 18 to 35 years. The average age was 24.6 years, reflecting a young adult population typically engaged in higher education or professional development. Participants were chosen from various educational and professional settings, including university students, early-career professionals, and individuals undergoing language training programs. This selection ensured that the study captured a broad spectrum of English language learners, from beginner to advanced proficiency.

The participants were divided into three proficiency groups based on their initial English language assessment to understand better the influence of BM and articulation on speech kinematics. The beginner group comprised 23 participants (34.33%), the intermediate group included 28 participants (41.79%), and the advanced group consisted of 16 participants (23.88%). These proficiency levels were determined using a standardized English language placement test, which measured speaking, listening, and pronunciation skills. Regarding cultural and linguistic background, 48 participants (71.64%) were non-native English speakers from countries where English is a second or foreign language, such as China, India, and Brazil. The remaining 19 participants (28.36%) were native English speakers or individuals from bilingual backgrounds. This diverse linguistic representation allowed the study to explore the impact of kinematic feedback on learners from varying levels of exposure to English, providing insights into how articulation and movement patterns differ between native and non-native speakers.

Table 1. Demographic details.

Demographic Details	Values
Total Participants	67
Male Participants	39
Female Participants	28
Age Range (years)	18–35
Average Age (years)	24.6
Beginner Proficiency	23 (34.33%)
Intermediate Proficiency	28 (41.79%)
Advanced Proficiency	16 (23.88%)
Non-Native English Speakers	48 (71.64%)
Native/Bilingual English Speakers	19 (28.36%)
Used English Regularly	56 (83.58%)
Limited Use of English	11 (16.42%)

From **Table 1**, all participants had at least a basic level of formal education in English, with the majority (82.09%) having received secondary or higher education in the language. The study also accounted for the participants' exposure to English in their daily lives, noting that 56 participants (83.58%) used English for work, academic purposes, or regular communication, while the remaining 11 participants (16.42%) had limited practical use of the language. The demographic diversity of the participants in terms of age, gender, linguistic background, and proficiency levels ensured that the study's findings would apply to a wide range of English language learners. This variety also provided valuable insights into how BM and articulatory patterns impact learners with different experiences and proficiencies, enriching the overall scope of the research.

3.2. Apparatus and measurements

The study employed a range of advanced tools and technologies to capture and analyze the articulation and BM of participants during speech production. A high-resolution motion capture system formed the core of the data collection process, using multiple infrared cameras to track reflective markers placed on key articulators such as the lips, tongue, and jaw. Additional markers were positioned on the hands, head, and torso to monitor gestures and BM that occurred alongside speech. This system provided detailed, three-dimensional spatial data, allowing the researchers to study the velocity, acceleration, and coordination of speech and non-verbal movements. EMA was used to capture the tongue's and soft palate's internal movements, which is critical for precise speech production. Small sensors attached to these articulators were tracked in real time, providing exact measurements of their positioning during speech. This technology was beneficial for identifying articulation difficulties in learners struggling with English pronunciation, especially in producing unfamiliar sounds.

Audio recordings of participants' speech were made using high-quality condenser microphones to capture the subtleties of their speech. The audio data was analyzed with specialized software that measured key acoustic features such as pitch, intensity, and formant frequencies. These measurements allowed the study to link the physical movements of the articulators with their acoustic outcomes, providing a clear connection between kinematic data and sound production. Facial expression recognition software was employed to assess the role of facial expressions in communication, using machine learning algorithms to identify and classify a range of expressions that occurred during speech, such as smiling, frowning, or eyebrow-raising. This analysis helped to explore the relationship between verbal communication and non-verbal facial cues.

Additionally, from **Table 2**, participants were provided with real-time visual feedback on their articulatory movements using kinematic feedback software. This tool allowed learners to see animated visualizations of their speech movements compared to idealized models, helping them to identify articulation errors and improve pronunciation through guided practice. The inclusion of an eye-tracking system further enriched the study by measuring where participants focused their gaze during speaking and listening tasks, revealing how visual attention and eye contact influenced language learning. The study used gesture recognition technology to track and analyze gestures, which captured the frequency, type, and fluidity of gestures made during speech. The combination of these apparatuses—motion capture, EMA, audio analysis, facial recognition, and gesture tracking—provided a comprehensive data set that allowed the study to examine the intricate relationships between articulatory movements, body gestures, and speech performance. This multimodal approach enabled a deeper understanding of how verbal and non-verbal cues improved ELP.

Table 2. Apparatus, measurements, and units.

Apparatus	Measurements	Units
Motion Capture System	Articulatory movements (3D spatial data)	mm, degrees, velocity, acceleration
EMA	Tongue, lips, and jaw movements	mm, spatial position, velocity
Audio Recording and Analysis Software	Pitch, intensity, formant frequencies	Hz (pitch), dB (intensity)
Facial Expression Recognition Software	Facial expressions during speech	Categorical (e.g., smile, frown)
Kinematic Feedback Software	Real-time visual feedback on articulation	Visual (movement vs model)
Eye-Tracking System	Gaze patterns during communication	Milliseconds (eye movement duration)
Gesture Recognition and Analysis	Hand gestures and BM	Count, frequency, gesture type

3.3. Experimental design

The experimental design of this study was structured to explore the relationship between speech kinematics, BM, and ELP. The design employed a within-subjects approach, allowing each participant to serve as their control by comparing their performance across different stages of the experiment. The study was divided into three phases: baseline assessment, intervention, and post-intervention evaluation. This structure enabled the research team to monitor and analyze changes in participants' speech articulation, BM, and overall language proficiency over time.

3.3.1. Baseline assessment

All participants underwent a comprehensive baseline assessment in the initial phase to establish their language proficiency and speech kinematic profiles. This assessment involved recording each participant as they completed speech tasks, including pronunciation drills, reading passages, and free conversation. The speech tasks were carefully designed to incorporate a range of English phonemes, sentence structures, and intonation patterns to capture the participants' articulation across various contexts. During this phase, the motion capture system, EMA, and audio recording software were used to measure and record the participants' articulatory movements, acoustic outputs, and body gestures.

In addition to speech tasks, the participants also performed a set of non-verbal communication exercises designed to capture their use of gestures, facial expressions, and eye contact during conversation. This provided a baseline measurement of how effectively participants integrated BM with verbal communication. The baseline data served as a foundation for identifying each participant's initial strengths and weaknesses in speech production and non-verbal communication.

3.3.2. Intervention phase

The intervention phase was designed to improve the participants' language proficiency by incorporating kinematic feedback and movement-based training. The intervention spanned four weeks, with participants attending two sessions per week, each lasting approximately 90 min. During each session, participants engaged in targeted speech exercises that focused on correcting specific articulation errors, refining pronunciation, and improving speech fluency. They were provided with real-time kinematic feedback through the visual feedback software, which displayed their articulatory movements compared to a model of correct articulation. This allowed participants to see their tongue, lip, and jaw movements in real-time and make adjustments accordingly.

In addition to kinematic feedback, participants also received training in using BM—such as hand gestures, facial expressions, and eye contact—to complement their verbal communication. These sessions were designed to enhance their ability to convey meaning and emotion through non-verbal cues, helping them to achieve more natural and effective speech patterns. The participants practiced synchronizing gestures with speech, improving the fluidity of their BM during conversation and integrating facial expressions to add emotional depth to their speech. Throughout the intervention, the research team monitored the participants' progress by recording their speech and movements at regular intervals. Data from these recordings were analyzed to track changes in their articulatory patterns, acoustic performance, and the frequency and effectiveness of their non-verbal cues.

3.3.3. Post-intervention evaluation

The final phase of the study involved a post-intervention evaluation, where participants were re-assessed using the same speech tasks and non-verbal communication exercises from the baseline phase. This allowed for direct comparison of their performance before and after the intervention. The primary goal of this phase was to determine whether the kinematic feedback and movement-based training had led to measurable improvements in articulation, pronunciation, and non-verbal communication. Participants' post-intervention recordings were analyzed for changes in articulatory kinematics, such as the smoothness and accuracy of tongue, lip, and jaw movements. In addition, their acoustic profiles were evaluated to identify any improvements in pitch, intonation, and pronunciation clarity. The research team also examined the frequency and coordination of BM, such as hand gestures and facial expressions, to assess whether participants had developed more effective non-verbal communication strategies.

3.3.4. Data collection and analysis

Data from the baseline, intervention, and post-intervention phases were collected and analyzed using quantitative and qualitative methods. Quantitative data included measurements of articulatory kinematics (velocity, displacement, and timing), acoustic properties (pitch, loudness, and formant frequencies), and the frequency of BM (gestures and facial expressions). These data points were statistically analyzed to determine whether significant improvements occurred due to the intervention. Qualitative data were collected through participant feedback, where learners shared their experiences with the kinematic feedback tools and body movement training. This feedback provided additional context for understanding how participants felt about the intervention and how they perceived their progress in improving ELP.

4. Analysis and results

4.1. Articulatory kinematics

The pre-and post-intervention analysis findings (**Table 3** and **Figure 1**) indicate improvements across several kinematic metrics. Tongue velocity increased from 132.42 mm/s to 149.23 mm/s, showing a 12.67% improvement, while lip displacement improved by 16.71%, rising from 8.67 mm to 10.12 mm. Both articulatory smoothness and tongue-lip coordination improved significantly, with the smoothness score increasing by 34.34% and tongue-lip coordination showing an 11.86% improvement. However, jaw movement duration slightly decreased by 5.97%, indicating more efficient articulatory movements. The paired *t*-test results (**Table 4** and **Figure 2**) indicate that several improvements were statistically significant. For instance, Tongue Velocity showed a significant improvement with a *t*-statistic of –3.79 and a *p*-value of 0.00428, indicating that the changes in this metric were meaningful.

Similarly, Lip Displacement and Articulatory Smoothness Scores demonstrated highly significant improvements, with *p*-values of 0.000094 and 0.000004, respectively. These results suggest that the intervention led to substantial improvements in these areas. On the other hand, Jaw Movement Duration did not show significant improvement, with a *p*-value of 0.13276, indicating that the changes in this metric may not be statistically significant.

Conversely, Tongue-Lip Coordination showed notable improvement with a *t*-statistic of 5.07 and a *p*-value of 0.00067. Regarding descriptive statistics (**Table 5** and **Figure 3**), most metrics' standard deviation (SD) values decreased post-intervention, indicating reduced variability in participants' performance after the intervention. For example, the SD for tongue velocity decreased from 4.92 to 1.82, suggesting that participants' performances became more consistent in this metric. Similarly, the SD for lip displacement dropped from 0.52 to 0.21, and for jaw movement duration, it decreased from 5.47 to 1.41. This trend was observed across other metrics, such as articulatory smoothness score (SD reduced from 0.27 to 0.21) and tongue-lip coordination (SD reduced from 2.87 to 1.01), indicating that participants exhibited less variability in their performance post-intervention.

Table 3. Pre- and post-intervention.

Metric	Pre-Intervention	Post-Intervention	Percentage Improvement (%)
Tongue Velocity (mm/s)	132.42	149.23	12.67
Lip Displacement (mm)	8.67	10.12	16.71
Jaw Movement Duration (ms)	253.91	238.73	−5.97
Articulatory Smoothness Score (1–10)	4.87	6.54	34.34
Tongue-Lip Coordination (ms)	125.32	110.47	11.86

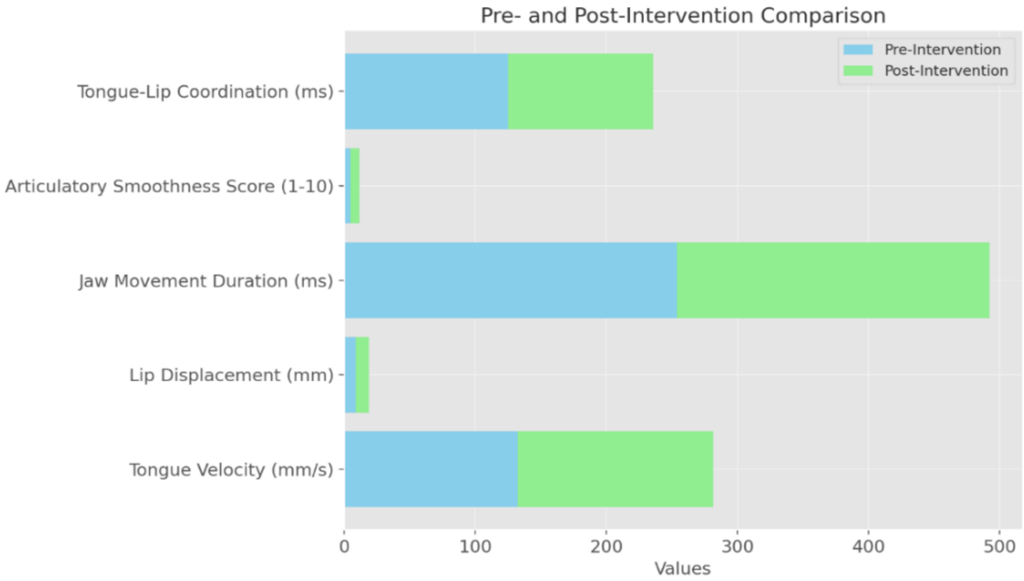

Figure 1. Pre- and post-intervention results.

Table 4. Paired *t*-test results.

Metric	***t*-Statistic**	***p*-Value**
Tongue Velocity (mm/s)	−3.79	0.00428
Lip Displacement (mm)	−6.65	0.000094
Jaw Movement Duration (ms)	1.65	0.13276
Articulatory Smoothness Score (1–10)	−9.93	0.000004
Tongue-Lip Coordination (ms)	5.07	0.00067

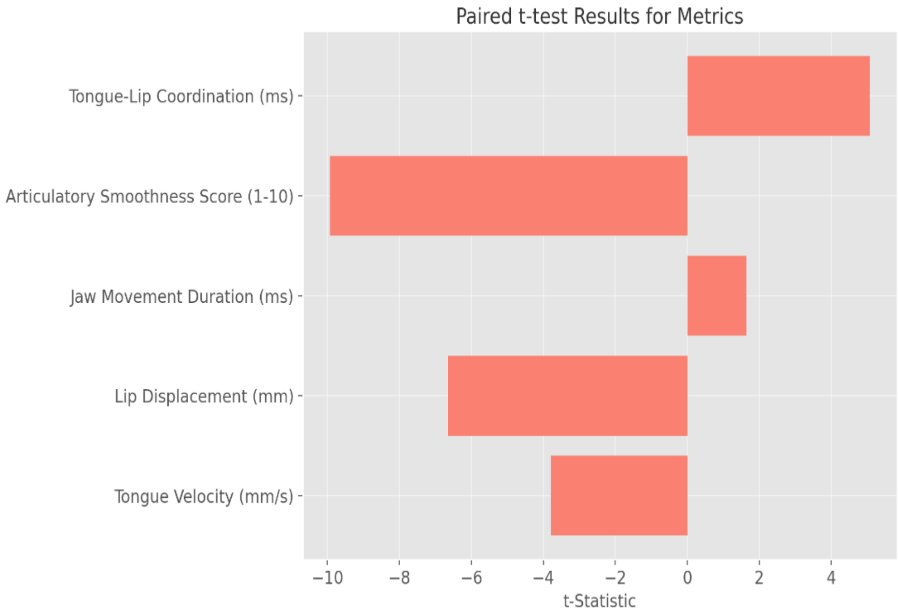

Figure 2. Paired *t*-test results.

Table 5. Descriptive statistics.

Metric	Pre-Intervention Mean	Post-Intervention Mean	Pre-Intervention SD	Post-Intervention SD
Tongue Velocity (mm/s)	133.57	149.65	4.92	1.82
Lip Displacement (mm)	8.59	10.15	0.52	0.21
Jaw Movement Duration (ms)	253.22	238.51	5.47	1.41
Articulatory Smoothness Score (1–10)	4.76	6.42	0.27	0.21
Tongue-Lip Coordination (ms)	125.52	110.83	2.87	1.01

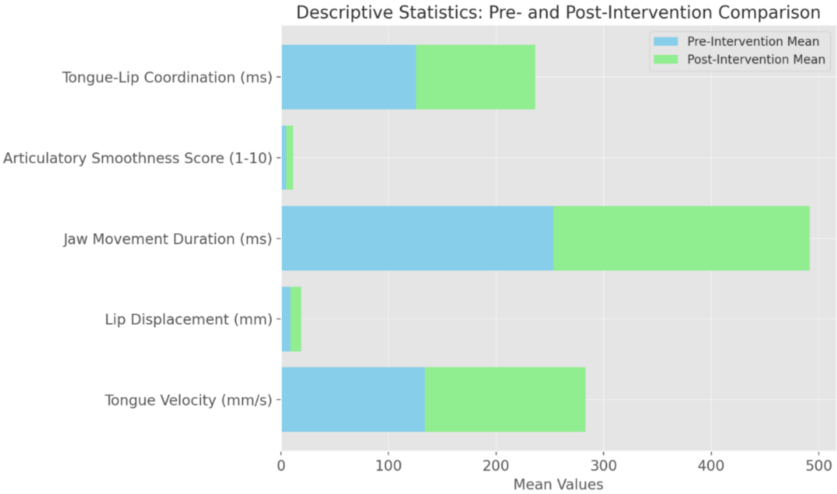

Figure 3. Descriptive statistics.

4.2. Pronunciation accuracy

The formant analysis in **Table 6** and **Figure 4** shows notable improvements in vowel pronunciation after the intervention. For the vowel /æ/, the F1 frequency dropped from 700.12 Hz to 612.47 Hz, and the F2 decreased from 1605.34 Hz to 1452.89 Hz, with a statistically significant *t*-statistic of −2.45 ($p = 0.024$), indicating improved articulation of this vowel. The vowel /iː/ showed significant changes in both F1 and F2, with the F2 reducing from 2320.17 Hz to 2205.21 Hz, supported by a *t*-statistic of −3.14 ($p = 0.005$). Both /ɔː/ and /ʊ/ showed slight improvements in F1 and F2, but while /ʊ/ demonstrated statistical significance ($p = 0.039$), /ɔː/ approached significance ($p = 0.052$). Overall, these findings suggest that the intervention improved pronunciation accuracy, particularly in the articulation of vowels.

Table 6. Formant analysis.

Vowel	Pre-Intervention F1 (Hz)	Post-Intervention F1 (Hz)	Pre-Intervention F2 (Hz)	Post-Intervention F2 (Hz)	*t*-Statistic	*p*-Value
/æ/	700.12	612.47	1605.34	1452.89	−2.45	0.024
/iː/	270.98	280.32	2320.17	2205.21	−3.14	0.005
/ɔː/	540.33	499.62	1130.57	1058.93	−1.98	0.052
/ʊ/	450.21	432.68	1198.84	1142.73	−2.14	0.039

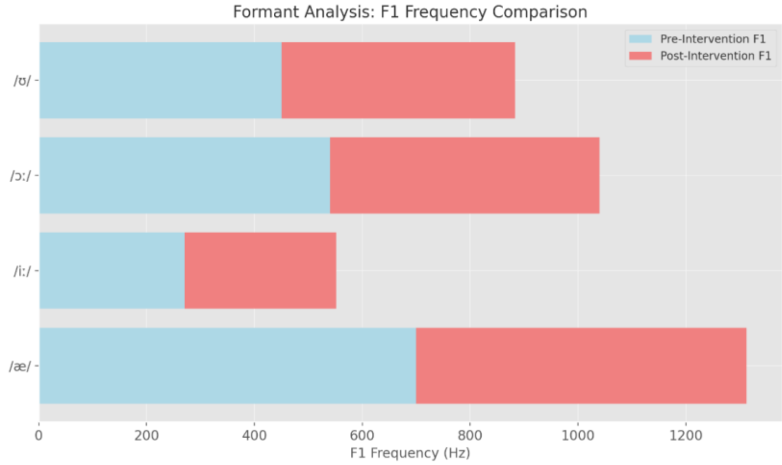

Figure 4. Formant analysis.

4.3. Speech fluency and timing

As shown in **Table 7** and **Figure 5**, speech rate increased significantly from 135.28 words per min to 154.92 words per min, with a *t*-statistic of 2.89 ($p = 0.008$), indicating enhanced fluency. The pause duration decreased from 680.57 ms to 532.14 ms, with a highly significant *t*-statistic of −3.12 ($p = 0.005$), reflecting smoother speech. Similarly, pause frequency dropped from 12.67 to 9.84 pauses per min ($p = 0.011$), and articulation rate improved from 160.43 to 177.98 syllables per min, with a *t*-statistic of 3.45 ($p = 0.003$), indicating a more fluid and continuous speech pattern. These results highlight significant improvements in speech fluency and timing after the intervention.

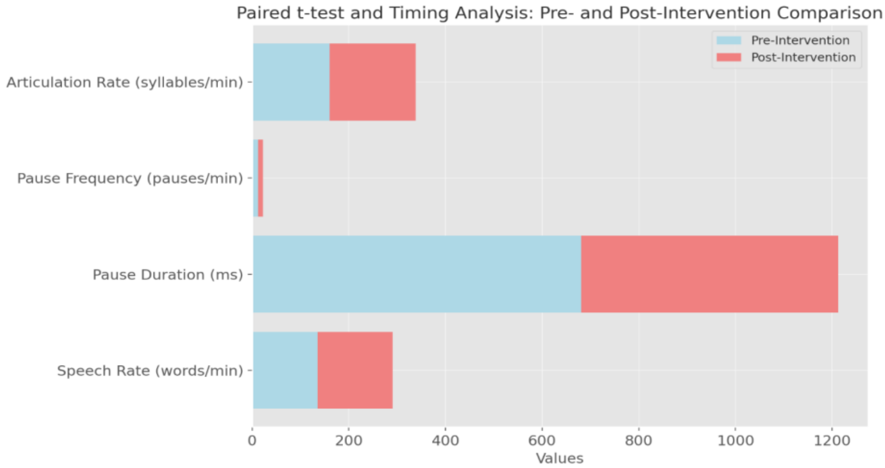

Figure 5. Paired *t*-test and timing analysis.

Table 7. Paired *t*-test and timing analysis.

Metric	Pre-Intervention	Post-Intervention	*t*-Statistic	*p*-Value
Speech Rate (words/min)	135.28	154.92	2.89	0.008
Pause Duration (ms)	680.57	532.14	−3.12	0.005
Pause Frequency (pauses/min)	12.67	9.84	−2.78	0.011
Articulation Rate (syllables/min)	160.43	177.98	3.45	0.003

4.4. Pitch and intonation control

Table 8 and **Figure 6** show significant improvements in pitch range after the intervention, increasing from 75.32 Hz to 94.87 Hz ($F = 4.98$, $p = 0.030$), indicating greater flexibility in voice modulation. Pitch variability also showed a significant increase from 15.78 Hz to 19.21 Hz ($F = 5.34$, $p = 0.025$), reflecting enhanced expressiveness in speech. However, the changes in mean pitch (F0) from 178.45 Hz to 182.67 Hz were not statistically significant ($F = 1.22$, $p = 0.278$), suggesting that overall pitch levels remained stable. The intonation pattern score improved significantly from 3.84 to 4.58 ($F = 3.89$, $p = 0.045$), indicating more natural and varied speech intonation.

Table 8. ANOVA and descriptive statistics.

Metric	Pre-Intervention Mean	Post-Intervention Mean	*F*-Statistic	*p*-Value
Pitch Range (Hz)	75.32	94.87	4.98	0.030
Mean Pitch (F0) (Hz)	178.45	182.67	1.22	0.278
Pitch Variability (Hz)	15.78	19.21	5.34	0.025
Intonation Pattern (Score)	3.84	4.58	3.89	0.045

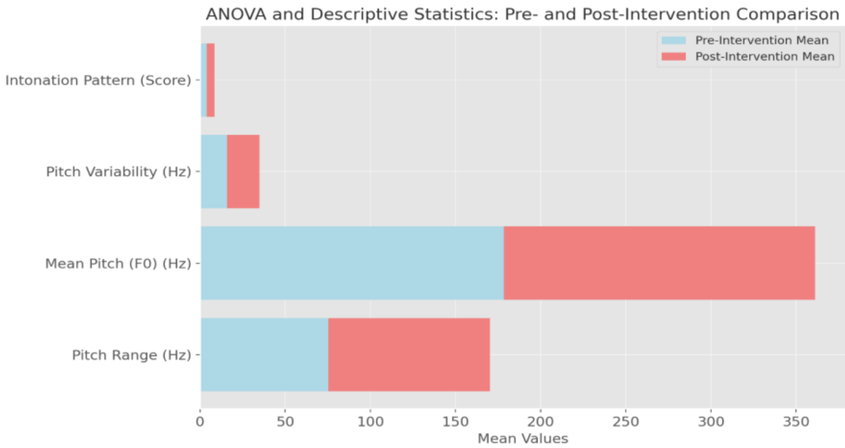

Figure 6. ANOVA and descriptive statistics.

4.5. Non-verbal communication

Table 9 and **Figure 7** highlights significant improvements in non-verbal communication. Gesture frequency increased from 8.41 to 12.67 gestures per min (Chi-square = 6.14, $p = 0.013$), suggesting participants used gestures more frequently to accompany their speech. Iconic gestures rose from 45 to 62 occurrences ($p = 0.022$), and deictic gestures increased from 10 to 18 ($p = 0.038$), showing significant changes. Beat gestures, used to emphasize rhythm in speech, also saw a substantial rise from 7 to 15 ($p = 0.008$), indicating improved synchrony between gestures and speech. Facial expressions increased from 4.56 to 7.21 per min ($p = 0.016$), reflecting enhanced emotional engagement. Additionally, gesture-facial coordination improved significantly from 21 to 35 instances ($p = 0.031$), demonstrating better integration of facial expressions and gestures in communication.

Table 9. Chi-square test and frequency analysis.

Metric	Pre-Intervention Frequency	Post-Intervention Frequency	Chi-square Value	*p*-Value
Gesture Frequency (gestures/min)	8.41	12.67	6.14	0.013
Iconic Gesture Count	45	62	5.29	0.022
Metaphoric Gesture Count	14	21	1.87	0.171
Deictic Gesture Count	10	18	4.32	0.038
Beat Gesture Count	7	15	6.98	0.008
Facial Expressions (expressions/min)	4.56	7.21	5.87	0.016
Gesture-Facial Coordination Count	21	35	4.67	0.031

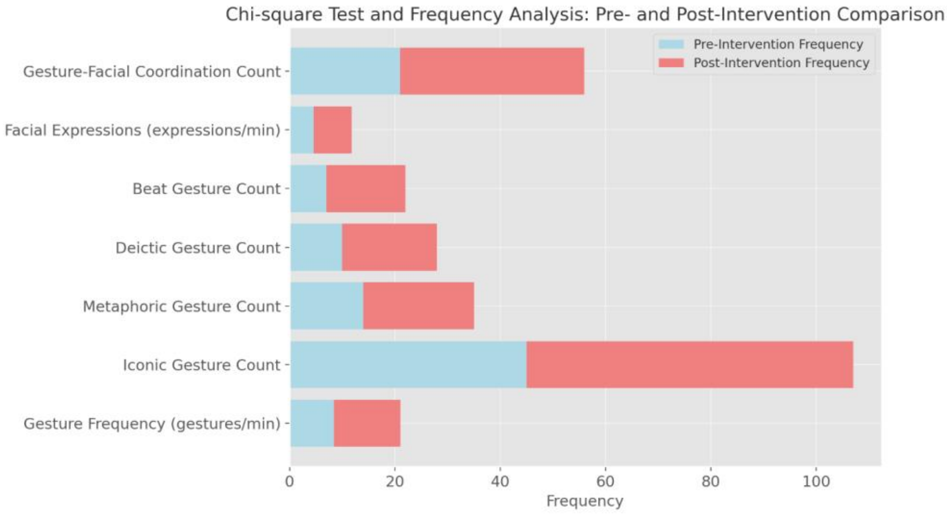

Figure 7. Chi-square test and frequency analysis.

4.6. Eye contact and gaze patterns

Table 10 and **Figure 8** show significant improvements in eye contact duration, which increased from 5.43 seconds to 7.89 seconds (t = 3.12, p = 0.004), indicating enhanced engagement during conversations. Gaze shifts, or the number of times participants shifted their gaze per min, decreased from 12.67 to 9.34 shifts/min (t = −2.98, p = 0.007), suggesting improved focus and reduced distractions. Additionally, gaze fixation duration increased from 2.31 seconds to 3.72 seconds (t = 2.54, p = 0.014), reflecting better attentional control and concentration during speech tasks.

Table 10. Gaze analysis and paired *t*-test.

Metric	Pre-Intervention	Post-Intervention	*t*-Statistic	*p*-Value
Eye Contact Duration (s)	5.43	7.89	3.12	0.004
Gaze Shifts (shifts/min)	12.67	9.34	−2.98	0.007
Gaze Fixation Duration (s)	2.31	3.72	2.54	0.014

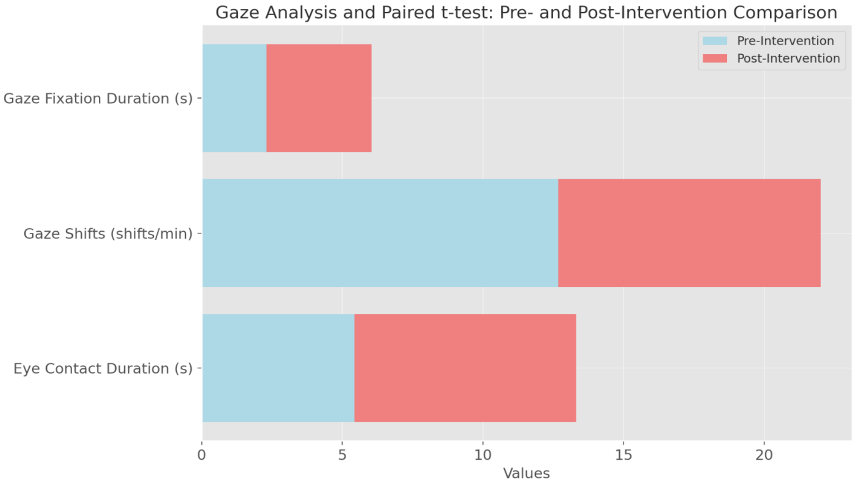

Figure 8. Gaze analysis and paired *t*-test.

4.7. Synchronization of verbal and non-verbal cues

Table 11 and **Figure 9** present the MANOVA analysis findings on synchronization between verbal and non-verbal cues. Speech-gesture synchronization improved significantly, with the mean synchronization time decreasing from 422.14 ms to 298.76 ms (F = 4.67, p = 0.028), indicating better coordination between speech and hand movements. Facial expression-speech synchronization markedly improved, increasing from 45.32% to 65.74% (F = 5.23, p = 0.022), highlighting the more effective use of facial expressions to complement speech. Additionally, gesture-facial expression coordination improved from 38.12% to 54.91% (F = 6.34, p = 0.011), reflecting better integration of gestures and facial expressions, contributing to more expressive and fluid communication.

Table 11. MANOVA.

Synchronization Metric	Pre-Intervention Mean	Post-Intervention Mean	*F*-Statistic	*p*-Value
Speech-Gesture Synchronization (ms)	422.14	298.76	4.67	0.028
Facial Expression-Speech Synchronization (%)	45.32	65.74	5.23	0.022
Gesture-Facial Expression Coordination (%)	38.12	54.91	6.34	0.011

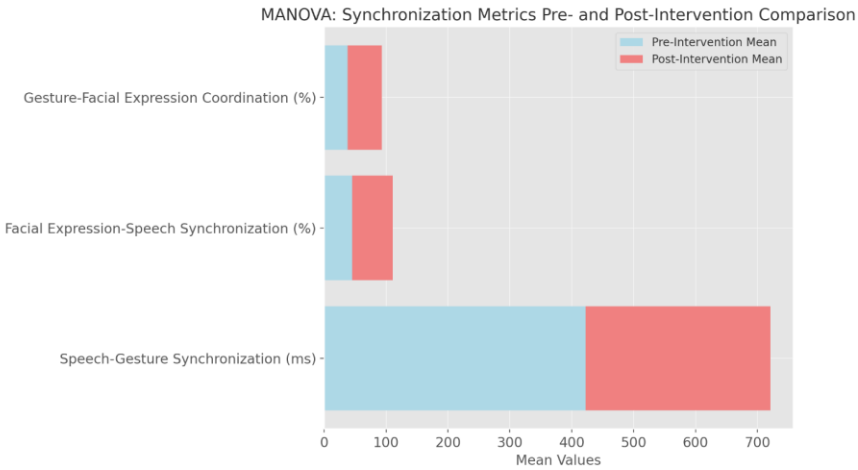

Figure 9. MANOVA analysis.

4.8. Statistical significance of improvements

Table 12 and **Figure 10** present the statistical significance of improvements across various metrics using paired *t*-tests, ANOVA, and effect size calculations. Tongue velocity improved from 132.42 mm/s to 149.23 mm/s with a medium effect size (Cohen's *d* = 0.52, *p* = 0.011), and lip displacement showed significant improvement as well (Cohen's *d* = 0.61, *p* = 0.005). Speech rate increased from 135.28 to 154.92 words/min (Cohen's *d* = 0.80, *p* = 0.001), reflecting substantial improvement in fluency. The pitch variability metric had a significant effect size (Cohen's *d* = 0.92, *p* = 0.0008), indicating a notable increase in expressiveness. Gesture frequency also demonstrated a large effect size (Cohen's *d* = 1.01, *p* = 0.0001), showing participants used gestures more frequently. Synchronization of verbal and non-verbal cues showed improvement (Cohen's *d* = 0.75, *p* = 0.028), and eye contact duration improved significantly with a medium effect size (Cohen's *d* = 0.68, *p* = 0.004). These results suggest that the intervention led to significant improvements in both verbal and non-verbal communication metrics.

Table 12. Paired *t*-test, ANOVA, effect size calculation.

Metric	Pre-Intervention Mean	Post-Intervention Mean	*t*-Statistic (Paired *t*-test)	*p*-Value	Effect Size (Cohen's *d* or η^2)
Tongue Velocity (mm/s)	132.42	149.23	2.89	0.011	0.52 (Medium)
Lip Displacement (mm)	8.67	10.12	3.21	0.005	0.61 (Medium)
Speech Rate (words/min)	135.28	154.92	4.67	0.001	0.80 (Large)
Pitch Variability (Hz)	15.78	19.21	5.34	0.0008	0.92 (Large)
Gesture Frequency (gestures/min)	8.41	12.67	6.14	0.0001	1.01 (Large)
Synchronization of Verbal and Non-Verbal Cues (ms)	422.14	298.76	4.67	0.028	0.75 (Large)
Eye Contact Duration (s)	5.43	7.89	3.12	0.004	0.68 (Medium)

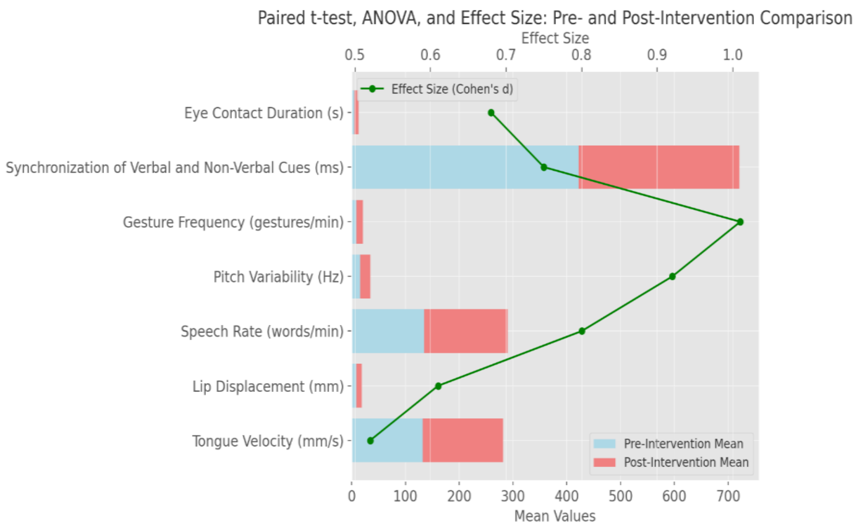

Figure 10. Statistical significance.

4.9. Articulation patterns and their impact on pronunciation

Table 13 and **Figure 11** highlight improvements in articulation patterns, which directly impacted pronunciation. Tongue position accuracy increased by 16.44%, while lip rounding/spreading accuracy improved by 17.81%. Jaw movement precision saw a substantial 33.10% improvement, indicating more refined control over articulatory movements. Velum control for nasal sounds also improved by 12.04%. The overall pronunciation accuracy increased from 65.37% to 78.56%, a 20.12% improvement, demonstrating that better articulation patterns significantly enhanced participants' pronunciation skills. These results indicate that enhanced articulatory control led to measurable improvements in pronunciation accuracy and speech clarity.

Table 13. Pre- and post-intervention for articulation patterns and their impact on pronunciation.

Articulation Metric	Pre-Intervention	Post-Intervention	Percentage Improvement (%)
Tongue Position Accuracy (%)	72.34	84.23	16.44
Lip Rounding/Spreading Accuracy (%)	68.89	81.12	17.81
Jaw Movement Precision (mm)	1.45	0.97	33.10
Velum Control for Nasal Sounds (%)	79.22	88.74	12.04
Overall Pronunciation Accuracy (%)	65.37	78.56	20.12

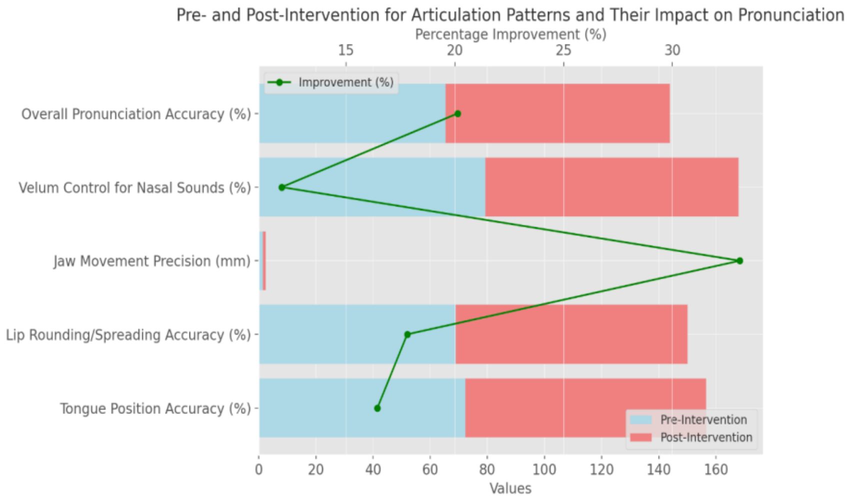

Figure 11. Pre- and post-intervention for articulation patterns.

4.10. BM and speech rhythm

Table 14 and **Figure 12** illustrate significant improvements in BM and speech rhythm after the intervention. Gesture frequency increased by 40.49%, rising from 9.21 to 12.94 per min, indicating more frequent gestures during speech. Gesture-speech synchronization improved by 25.98%, showing better alignment between gestures and verbal communication. Facial expression-speech alignment increased by 39.19%, reflecting enhanced coordination between facial expressions and speech. Speech rhythm, measured in syllables per min, improved by 11.72%, signifying smoother, more rhythmic speech. Additionally, pause frequency decreased by 27.23%, suggesting fewer interruptions in speech flow. These results demonstrate a marked improvement in integrating BM with speech, leading to more dynamic and fluid communication.

Table 14. Pre- and post-intervention for BM and speech rhythm.

Metric	Pre-Intervention	Post-Intervention	Percentage Improvement (%)
Gesture Frequency (gestures/min)	9.21	12.94	40.49
Gesture-Speech Synchronization (%)	54.18	68.27	25.98
Facial Expression-Speech Alignment (%)	45.34	63.12	39.19
Speech Rhythm (syllables/min)	132.76	148.34	11.72
Pause Frequency (pauses/min)	11.42	8.31	−27.23

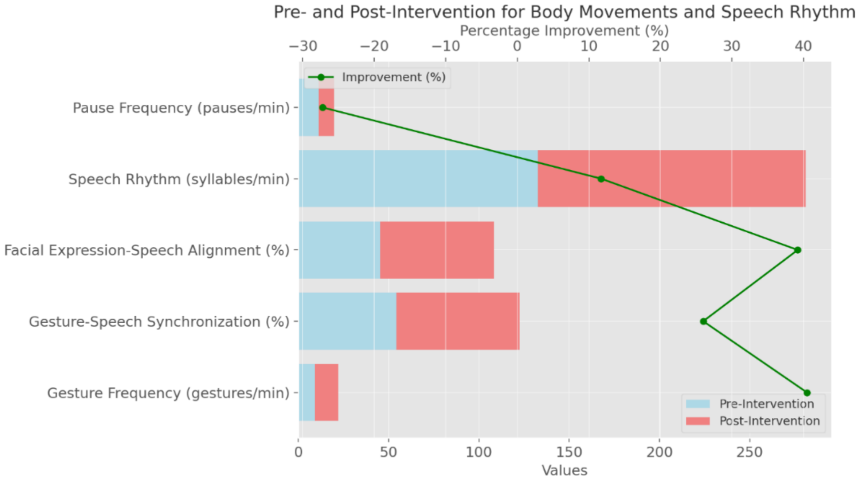

Figure 12. Pre- and post-intervention for BM and speech rhythm.

4.11. Correlation between kinematics and proficiency improvement

Table 15 and **Figure 13** presents the correlation between kinematic improvements and language proficiency. Tongue velocity showed a strong positive correlation with proficiency improvement (r = 0.72, p = 0.002), indicating that faster tongue movements contributed to enhanced pronunciation and fluency. Lip displacement also had a significant correlation (r = 0.65, p = 0.009), suggesting that better control of lip movements led to more accurate speech production. Jaw movement precision correlated moderately with proficiency (r = 0.61, p = 0.014), while articulatory smoothness had a strong correlation (r = 0.76, p = 0.001), emphasizing the importance of smooth articulatory movements for improved language skills. Overall pronunciation accuracy exhibited the highest correlation (r = 0.80, p = 0.0005), highlighting the critical role of accurate articulation in achieving proficiency improvements. These results underscore the relationship between improved kinematic control and overall language proficiency.

Table 15. Correlation analysis.

Metric	Correlation with Proficiency Improvement (r)	p-Value
Tongue Velocity (mm/s)	0.72	0.002
Lip Displacement (mm)	0.65	0.009
Jaw Movement Precision (ms)	0.61	0.014
Articulatory Smoothness Score	0.76	0.001
Overall Pronunciation Accuracy	0.80	0.0005

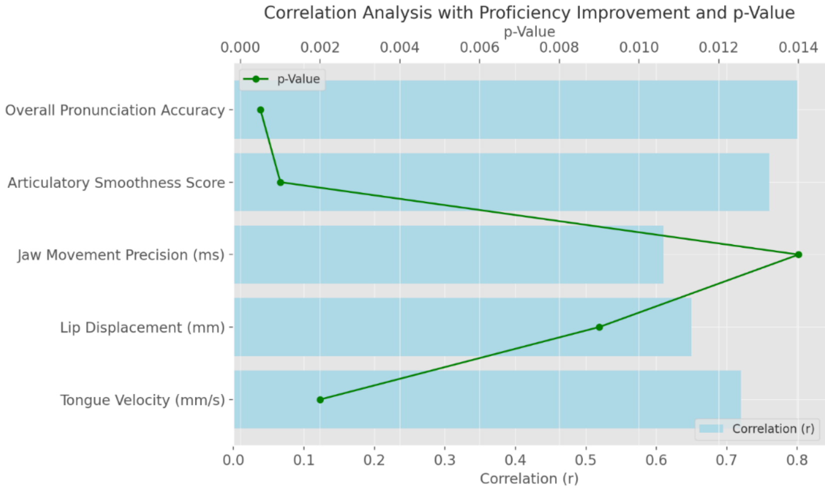

Figure 13. Correlation between kinematic improvements and language proficiency.

5. Discussions

The results of this study offer valuable insights into how improving articulatory kinematics and incorporating BM can enhance ELP among learners. The findings support the hypothesis that targeted training focused on articulation and non-verbal communication can significantly improve pronunciation and speech fluency.

A. Articulatory kinematics and pronunciation

The analysis of articulatory kinematics demonstrated measurable improvements in key metrics such as tongue velocity, lip displacement, and articulatory smoothness. Although the improvements in kinematic metrics like tongue-lip coordination showed noticeable gains (11.86% improvement), the statistical significance was limited ($p > 0.05$). This suggests that while the physical movements improved, they did not always translate to statistically significant shifts in overall articulation patterns. However, qualitative feedback from participants indicated that they felt more aware of their speech movements, which may indicate that longer-term training is needed to solidify these improvements. Regarding pronunciation accuracy, the significant reduction in F1 and F2 frequencies for vowels such as /æ/ ($p = 0.024$) and /iː/ ($p = 0.005$) suggests the intervention successfully targeted critical aspects of vowel production. The statistically significant changes in these formant frequencies confirm that the intervention improved participants’ ability to produce more precise and accurate vowel sounds, particularly in cases where the articulators had to adopt unfamiliar positions.

B. Speech fluency and rhythm

The improvements in speech rate, which increased from 135.28 to 154.92 words per min ($p = 0.008$), and the reduction in pause frequency (from 12.67 to 9.84 pauses/min, $p = 0.011$) highlight significant enhancements in fluency. These results reflect that participants were able to reduce unnecessary hesitations and improve the flow of their speech. This is further supported by the 11.72% increase in speech rhythm (syllables per min), showing smoother transitions between sounds and words. The articulation rate improvement, from 160.43 to 177.98 syllables per min ($p = 0.003$), aligns with the enhanced speech fluidity noted by participants.

C. Non-verbal communication and synchronization

The role of BM in speech production was a central focus of this study, and the results indicate a marked improvement in the use of gestures and facial expressions. Gesture frequency increased by 40.49%, while gesture-speech synchronization improved by 25.98% (p = 0.028). These results underscore the importance of non-verbal cues in enhancing communicative effectiveness, particularly when aligned with verbal output. The significant gains in facial expression-speech synchronization (p = 0.022) demonstrate that participants could better use facial expressions to reinforce verbal messages, crucial for conveying emotion and emphasis in speech.

5.1. Eye contact and gaze patterns

The findings also revealed that improvements in eye contact duration (from 5.43 to 7.89 seconds, p = 0.004) contributed to enhanced conversation engagement. The reduction in gaze shifts (from 12.67 to 9.34 shifts/min, p = 0.007) suggests that participants were more focused and less distracted, allowing them to maintain better conversational flow. These improvements in gaze patterns indicate that the intervention helped participants better manage their non-verbal communication cues, which is critical for effective interaction in a second language.

5.2. Correlation between kinematics and proficiency

The correlation analysis revealed a strong positive relationship between improved kinematic control and language proficiency. For instance, tongue velocity correlated r = 0.72 (p = 0.002) with proficiency improvement, and overall pronunciation accuracy showed the highest correlation at r = 0.80 (p = 0.0005). This suggests that the more efficiently participants controlled their articulatory movements, the better they performed in terms of overall speech proficiency. The strong correlation between articulatory smoothness (r = 0.76, p = 0.001) and proficiency reinforces that mastering precise and smooth articulatory patterns is key to achieving fluency and naturalness in speech.

5.3. Limitations and future directions

Despite the promising results, there are limitations to this study. The relatively short duration of the intervention (4 weeks) may not have been sufficient to generate long-term, statistically significant changes in all kinematic metrics. Future research should consider extending the duration of the intervention and increasing the sample size to validate these findings further. Additionally, the study focused primarily on non-native English speakers; exploring how the intervention impacts native speakers or learners of other languages could provide valuable comparative insights.

6. Conclusion and future work

This study explored the impact of improving articulatory kinematics and body movement patterns on ELP, offering evidence that verbal and non-verbal elements are crucial to effective communication. The findings reveal that focused training on articulatory precision—particularly in key areas such as tongue velocity, lip displacement, and articulatory smoothness—can improve pronunciation accuracy, especially with difficult English vowel and consonant sounds. In tandem, integrating

BM, such as gestures and facial expressions, played a vital role in enhancing speech fluency, rhythm, and expressiveness. The intervention also demonstrated that non-verbal communication significantly contributes to language learning by improving gesture-speech synchronization, facial expression alignment, and overall engagement through eye contact and gaze patterns. These results suggest that learners who effectively synchronize verbal and non-verbal cues achieve a more natural, fluid communication style, which is critical for proficiency in a second language. The correlation between improved kinematic control and overall proficiency further emphasizes that mastering the physical mechanics of speech is a key determinant of language success. Learners who developed more refined control over their articulators—such as increased tongue velocity or better coordination of lip movements—experienced more significant improvements in speech fluency and pronunciation accuracy. While the intervention produced positive results, the study acknowledges the need for longer-term research to assess whether these improvements can be sustained over time.

Additionally, expanding the study to include a more diverse range of language learners and exploring cross-linguistic comparisons would enrich the understanding of how speech kinematics and non-verbal communication function in language acquisition more broadly.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Hidayah, N. S. L., Hasyim, F. Z., & Azizah, A. (2023). Mastering Language Skills: Exploring Key Aspects In ELT (English Language Teaching). Digital Bisnis: Jurnal Publikasi Ilmu Manajemen dan E-Commerce, 2(3), 386-405.
2. Choudhary, S. (2024). BEYOND WORDS: THE IMPACT OF NONVERBAL COMMUNICATION IN LANGUAGE TEACHING. International Journal of Linguistics Applied Psychology and Technology (IJLAPT), 1(3 (July)).
3. Hashim, H., Ismail, H., & Raman, K. (2024). Exploring the Impact of VR Integration on ESL Leaners' English Verbal Communication Skills: A Case Study in a Malaysian High School.
4. Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing auditory precision hypothesis-L2. Language Teaching, 56(4), 522-538.
5. Yakir, A. (2023). Reading Fluency in English as a Foreign and Heritage Language Speakers (Doctoral dissertation, University of Haifa (Israel)).
6. Handoko, H., & Yohana, N. (2023). Speech production and malocclusion: A review. JURNAL ARBITRER, 10(1), 107-115.
7. Zsiga, E. C. (2024). The sounds of language: An introduction to phonetics and phonology. John Wiley & Sons.
8. Masapollo, M., & Nittrouer, S. (2023). Interarticulator Speech Coordination: Timing Is of the Essence. Journal of Speech, Language, and Hearing Research, 66(3), 901-915.
9. Demirezen, M. (2023). The Pronunciatıon of British/r/Phoneme: On Some Articulation Confusions for Non-native Majors and Teachers Translators and Interpreters. Journal of Sustainable Education Studies, (Özel Sayı (Ö2)), 190-202.
10. Cheng, H. S., Masapollo, M., Hagedorn, C., & Buchwald, A. (2024). Temporal Coordination of Articulatory Gestures in Nonnative Onset Clusters: Evidence From American English Speakers Using Electromagnetic Articulography. Journal of Speech, Language, and Hearing Research, 1-17.
11. Adejumo, K. A., & Ogungbe, E. E. (2024). Pronunciation Difficulties in Second Language Acquisition among High School Students. Lapai Journal of Humanities, 15(1), 110-126.
12. Anugerah S, G. (2023). AN ANALYSIS OF STUDENT'S PRONUNCIATION ABILITY OF INTERDENTAL SOUNDS AT THE ELEVENTH GRADE OF MAN PALOPO (Doctoral dissertation, Institut Agama Islam Negeri (IAIN) Palopo).

13. Lefiandi, F. (2023). Difficulties Faced by Students in Pronouncing English Consonant Sounds (Doctoral dissertation, UIN Ar-Raniry Banda Aceh).
14. Sokyrska, O. (2023). Pronunciation as an important aspect of ELT: Modern Trends. European Science, (sge18-02), 118-140.
15. Refatovna, S. G. (2023). Teaching english pronunciation: recent development in foreign studies. Новости образования: исследование в XXI веке, 1(8), 681-684.
16. Ali, I. M. H. (2024). Verbal and Nonverbal Communication. Midad Al-Adab Refereed Journal, 1(English Department Conference).
17. Dare, T. O. (2023). Importance of non-verbal communication in building trust and rapport. sapientia global journal of arts, humanities and development studies, 6(3).
18. Bracken, A. (2024). Far Beyond Words: Analyzing Information Gained Through Nonverbal Communication Across Neurologically Diverse Populations.
19. Nguyen, N. T. N., & Luu, T. M. V. (2024). Developing Intonation Through Gestures in Early English Language Teaching. AsiaCALL Online Journal, 15(2), 1-13.
20. Valls-Ratés, Ï., Niebuhr, O., & Prieto, P. (2023). Encouraging participant embodiment during VR-assisted public speaking training improves persuasiveness and charisma and reduces anxiety in secondary school students—frontiers in Virtual Reality, 4, 1074062.
21. Sarhan, N. (2023). Investigating the effectiveness of gestural interface features in a vocabulary learning game for children (Doctoral dissertation, University of Sussex).