FROM SYLLABLE-TIMED TO STRESS-TIMED RHYTHM: GESTURE AS A FACILITATOR IN ENGLISH RHYTHMIC TRAINING

DEL RITMO SILÁBICO AL ACENTUAL: EL GESTO COMO FACILITADOR EN LA ENSEÑANZA DEL RITMO DEL INGLÉS

HELENA CASTRILLO GONZÁLEZ
Universidad de las Palmas de Gran Canaria (ULPGC)
helena.castrillo101@alu.ulpgc.es
https://orcid.org/0009-0003-6109-1954

Fecha de recepción: 09-10-2025
Fecha de aceptación: 21-11-2025

ABSTRACT

This undergraduate classroom study investigates the effectiveness of rhythm-based instruction and the integration of gesture as a facilitator to improve the production of English rhythm among native Spanish speakers. Given the typological differences between Spanish (a syllable-timed language) and English (a stress-timed language), learners often struggle with stress placement, and overall prosodic features. Grounded in embodied learning theory, this pilot study explores whether multimodal strategies – particularly hand gestures and kinesthetic movements mimicking the rhythm of English – enhance learners' oral performance. Thirty-five undergraduate students participated in a short training program based on the materials from Estebas-Vilaplana and Soláns' MOOC (UNED, 2019). Pre and post-tests of participants' oral production were assessed considering their comprehensibility, fluency, and accentedness. Results show significant improvements within each group, especially in the experimental group concerning comprehensibility and fluency. These findings support the value of integrating

rhythmic-based instruction and the use of gesture into L2 pronunciation instruction, particularly for learners from syllable-timed language backgrounds.

KEY WORDS: gesture; embodied learning; English stress-timed rhythm; prosody; Spanish syllable-timed rhythm

RESUMEN

Este estudio en el aula universitaria investiga la eficacia de la instrucción basada en el ritmo y la integración del gesto como facilitador para mejorar la producción del ritmo del inglés en hablantes nativos de español. Dadas las diferencias tipológicas entre el español (una lengua de ritmo silábico) y el inglés (una lengua de ritmo acentual), los estudiantes suelen tener dificultades con la colocación del acento y los elementos prosódicos en general. Basado en la teoría del aprendizaje corporizado, este estudio piloto explora si las estrategias multimodales –en particular los gestos manuales y los movimientos cinestésicos que imitan el ritmo del inglés- mejoran el rendimiento oral del alumnado. Treinta y cinco estudiantes universitarios participaron en una breve formación basada en los materiales del MOOC de Estebas-Vilaplana y Soláns (UNED, 2019). Las pruebas realizadas en el pre y en el post-test de la producción oral de los participantes fueron evaluadas considerando la comprensibilidad, fluidez y acento. Los resultados muestran mejoras significativas en ambos grupos, especialmente en el grupo experimental en cuanto a comprensibilidad y fluidez. Estos hallazgos respaldan el valor de integrar la instrucción rítmica y el uso de gestos en la enseñanza de la pronunciación en L2, particularmente para estudiantes cuya lengua materna posee un ritmo silábico.

PALABRAS CLAVE: aprendizaje corporizado; gesto; prosodia; ritmo acentual del inglés; ritmo silábico del español

1. INTRODUCTION

The acquisition of a second language (L2) involves various layers of phonological adaptation, among which prosody plays an important yet often under-addressed role (Kehoe and Stoel-Gammon, 1997). According to a recent study carried out by Baills, Alazard-Guiu and Prieto (2022), there is evidence that prosody plays a crucial role in pronunciation (Trofimovich and

Baker, 2006; Lomotey, 2013; Wang, 2022, as cited in Baills et al., 2022). One of the key dimensions of prosody is rhythm. Learners always focus first on segmental pronunciation (i.e., individual sounds) and then move to suprasegmental features like rhythm and intonation, (Cruttenden, 2014), although the latter are essential for fluency and intelligibility (Barreiro, Estebas-Vilaplana and Soto, 2007). For Spanish speakers learning English, rhythm constitutes a fundamental obstacle to achieving a high degree of comprehensibility. This stems from the typological differences between these two languages: Spanish is syllable-timed, whereas English is stress-timed (Dauer, 1983; Roach, 2009). This distinction is critical because rhythm also affects intelligibility, fluency and the perceptual impression perception of a foreign accent (Cutler, 2015; Munro and Derwing, 2015).

In syllable-timed languages, syllables are perceived to be of approximately equal length. By contrast, stress-timed languages like English, exhibit stressed syllables at regular intervals, with intervening unstressed syllables being weakened and reduced. Thus, while in the former type of languages a rhythmic beat tends to be perceived on every single syllable, in the latter, a rhythmic beat is aligned with stressed syllables only. This results in the grouping of a stressed syllable and any following unstressed syllables (excluding the next stressed syllable) into a larger unit called *foot* (Cruttenden, 2014). This structural difference may be a challenge to produce English prosody by native Spanish speakers, as it requires mastering stress placement (Estebas-Vilaplana and Soláns, 2020b) and the shortening and reduction of vowels in unstressed syllables. Failure to achieve these adjustments results in Spanish learners often exhibiting a "foreign sounding" rhythm, which can affect their comprehensibility, fluency and accentenedness.

Recent studies on the benefits of integrating prosody instruction into pronunciation training suggest that gesture may serve as a powerful aid in acquiring prosodic features (Jusslin et al., 2022; Baills et al., 2022; Li et al., 2023, Zhang et al., 2020). Theories of embodied learning (Golding-Meadow, 2003; Jusslin et al., 2022) posit that cognition is grounded in bodily actions and that physical movements can enhance learning outcomes. In language learning specifically, hand gestures mimicking the rhythm of English have been shown to facilitate prosodic acquisition (Esteve-Gilbert and Prieto, 2018).

Furthermore, in higher education studies, one promising approach that emerged in 2008 has been the implementation of MOOCs, *Massive Open Online Courses* (Downes, 2008), and LMOOCs, *Language Massive Open Online Courses* (Bárcena et al., 2014). Estebas-Vilaplana and Soláns (2020b) explored the role of a pronunciation-focused LMOOC and found that it

contributed significantly to learner awareness and engagement in suprasegmental features like rhythm and stress. According to their findings, learners who completed the MOOC performed significantly better in oral production than those who only followed standard materials (2020b). The MOOC or LMOOC, "The Acquisition of English Pronunciation through Songs and Literary Texts" (2019), can be found at the UNED official website.[1] This four-module MOOC offers multimodal learning opportunities that combine rhythmic-based instruction with songs and literary texts, with a large variety of texts extracted from *Antología Oral de la Literatura Inglesa: From Beowulf to Virginia Woolf* (Barreiro, Estebas-Vilaplana, Soto, 2009).

By adopting this framework as a learning tool, the present pilot study investigates whether the incorporation of gesture into a rhythmic-based instruction can develop even further the rhythm of English in Spanish learners. Using materials based on Soláns and Estebas-Vilaplana's MOOC (2019), the study compares two groups: control and experimental. The control group only receives the rhythmic-based instruction found in the MOOC, whereas the experimental group receives the same instruction while, in addition, integrating gesture.

This paper is organized as follows: in the following section, a contextualization is provided, including a review of relevant studies on rhythm in English and in Spanish, as well as the role of gesture in language acquisition. The following section describes the methodology of the study, with details about participants, instructional materials, training procedures, training sessions, oral production tasks and data analysis. After that, the results are presented in section 4, highlighting statistical values, group comparisons and graphical representations of the data. Then, in the discussion section these findings will be interpreted, emphasizing the implications for rhythmic-based instruction and the implementation of gesture. Finally, the paper concludes by summarizing the main outcomes of the study and acknowledging its limitations.

2. CONTEXTUALIZATION

2.1. Rhythm in English and Spanish

Rhythm, "a concept difficult to grasp and generally assumed to relate to speech cadences, is the temporal organization of sounds in terms of grouping" (Jun, 2005 *as cited in* Lahoz, 2012).

[1] "Aprendizaje de la pronunciación inglesa a través de canciones y textos literarios (Permanente) | UNED Abierta": https://iedra.uned.es/courses/course-v1:UNED+Proning+Permanente/about

Lahoz (2012) also mentioned that it was Pike (1945) who coined the terms syllable-timed and stress-timed, as well as explaining for the first time the differences between these two concepts:

> English sentences are spoken with recurrent bursts of speed, with long or short pauses or with intonation breaks between […]. Many non-English languages (Spanish, for instance) tend to use a rhythm which is more closely related to the syllable than the regular stress-timed of English; in this case, it is the syllables, instead of the stresses, which tend to come at more-or-less evenly recurrent intervals, and syllables and vowels are less likely to be shortened and modified. (1945: 35)

In addition to this, Roach (2009) points out that English follows a stress-timed rhythm, meaning that the intervals between stressed syllables are generally similar, regardless of how many unstressed syllables occur in between. By contrast, other languages are syllable-timed, where both stressed and unstressed syllables tend to appear at regular intervals, so the duration between stressed syllables changes according to the number of unstressed syllables. Spanish contrasts with the stress-timed rhythm of English, often resulting in perceptual and production difficulties for Spanish-speaking learners (Dauer, 1983).

This distinction between Spanish syllable-timed and English stress-timed rhythm has profound implications for L2 learners. Estebas-Vilaplana and Soláns (2020b) note that Spanish learners often struggle with English rhythm, leading to a syllable-timed delivery that sometimes sounds unnatural. Issues like this affect not only comprehensibility but also fluency and accentedness.

To help learners of English, the use of rhythmic-based instruction has gained awareness as a method for reducing rhythmic challenges for L2 learners of English. Barreiro, Estebas-Vilaplana and Soto (2005) advocated for a new method "from prosody to sounds". This innovative approach implied starting from suprasegmental features such as prosody and then, following with segmental features like individual vocalic or consonantal sounds. Similarly, Barreiro, Estebas-Vilaplana and Soto (2007) incorporated poetry and singing into their courses about English pronunciation/phonetics for Spanish learners to offer them a more intuitive path to internalize stress-timed rhythm.

Moreover, Estebas-Vilaplana and Soláns (2019) created their MOOC or LMOOC "The Acquisition of English Pronunciation through Songs and Literary Texts", an online and free-access course with video tutorials that combine practice and description to help learners

improve their English pronunciation. Then, Estebas-Vilaplana and Soláns (2020a) examined the role of their MOOC in higher education studies and found that learners who completed their course performed significantly better in oral production than those who only followed the standard materials. This course follows the previously mentioned order: first, concepts related to suprasegmental aspects are explained, and then the segmental aspects are presented. In this course, the phonetic features of English are presented through recited and sung literary texts. The aim is always to work with rhythm as a foundation, since literary texts like poetry (and, of course, songs) come with a rhythm which is fixed by the author, and hence, there is greater change for the student to identify it and then produce it at their early stage of training.

2.2. The Role of Gesture in Language Learning

Gesture has been recognized in different studies as a pedagogical utility in language instruction. Research by Goldin-Meadow (2005) established that gesture supports cognitive processes and can serve as a tool in language learning. Iverson and Goldin-Meadow (2005) also studied how gestures can help children visualize linguistic structures and they suggest that including gestures in language instruction "may be paving the way for future developments in language" (2005:2)

More recently, Li et al. (2023) and Baills et al. (2022) provide empirical support in their studies showing that rhythm-based approaches with gesture implementation significantly reduce accentedness and improve suprasegmental accuracy in L2 learners. In addition to this, Prieto and Esteve-Gilbert (2018) propose that gestures can improve learners' ability to perceive and produce natural rhythm, while Jusslin et al. (2022) confirm that the gesture-based training enhances both perception and production in L2 learners. This idea builds on findings from Baills, Alazard-Guiu and Prieto (2022), who explained that embodied prosodic training improves suprasegmental accuracy and reduces foreign accent when implemented consistently. In addition to this, several studies have stated that "prosodic features play a crucial role in producing a natural, native-like speech" (Barreiro, Estebas-Vilaplana and Soto, 2005).

In this context, this study tests whether the use of rhythmic-based instruction and the implementation of gestures can facilitate the oral production of English prosody by Spanish L1 learners. Specifically, the study examines if participants who incorporated gestures into their training outperformed those who did not.

3. METHODOLOGY

3.1. Participants

A total of thirty-five undergraduate students between 21 and 27 years old (M=24.0, SD=2.16), and in their final year of the Degree in Modern Languages at the University of Las Palmas de Gran Canaria participated in this study. All participants had some knowledge about English phonetics and phonology since they had attended a compulsory, one semester course in their second year. In addition to this, all participants reported Spanish as their first language, and none reported hearing impairments or diagnosed speech difficulties.

To ensure comparability, participants were assigned to one of two groups: a control group and an experimental group. Participants with morning schedule formed the control group (n=20) and those with afternoon schedule formed the experimental group (n=15). Both groups followed the same rhythm-based instruction, but the experimental group received additional training implementing the use of gestures and body movements. In addition to this, only the experimental group was encouraged to perform gestures and reinforce sentence-stress production.

3.2. Instructional Materials

The instructional design materials were adapted from Estebas-Vilaplana and Soláns' MOOC or LMOOC "The Acquisition of English Pronunciation through Songs and Literary Texts" (UNED, 2019), which is specifically targeted at teaching the rhythm of English to Spanish speakers. Its main aim is the improvement of English pronunciation by Spanish speakers. This course also aims to showcase the contrast between the syllable-timed and the stress-timed rhythmic organizations and to propose structured exercises to promote the acquisition of stress-timed rhythm. The MOOC is divided into three different modules, and each module consists of five classes: the first four present the main features of the phonetic trait being studied, and the last class includes a final review and recommendations for further work. It is divided as follows:

- Module 1: The rhythm in English and in Spanish (1. Rhythm in English and in Spanish, 2. Rhythm and accentuation, 3. Double stress and stress shifts, 4. Focus, stress and unstressing, 5. Final review and recommendations)

For this study, only module 1 was included, and more specifically, classes 1 and 2: with class 1 being "rhythm in English and in Spanish", and class 2 "rhythm and accentuation".

The training consisted of two sessions over two days: each lasting 80 minutes. Both groups received the same training, where they first observed video content from the MOOC and then, they followed adapted exercises in two different handouts (one per session) prompting students to identify, discriminate, imitate and produce different types of texts presented during the viewing of the videos, including varying stress patterns and rhythmic alternations. Both handouts included perceptual and production tasks using the same texts from the videos. The experimental group was encouraged to use hand gestures mimicking the rhythm of English to accompany their oral production (the same gestures observed in the videos), while the control group performed the same tasks without gesture instruction.

3.3. Speech Tasks

To measure the impact of the instruction and to evaluate participants' oral performance, three reading tasks were designed and administered at both pre-test and post-test situations. The tasks included two poetry texts and one narrative text, resulting in a total of six speech samples per participant. The same three texts were used for both the pre-tests and post-tests:

1. Trained poetry text: a poem excerpt included in the MOOC, which participants had encountered during training sessions. Text extracted from "Where the bee sucks, there suck I" by William Shakespeare.
2. Trained narrative text: a short prose passage with natural stress-timed rhythm also included in the MOOC. Text extracted from *Pride and Prejudice* by Jane Austen.
3. Untrained poetry text: a poem of similar length and rhythmic complexity but not used during the training. Text extracted from "Sorrow" by Edna St. Vincent Millay.

3.4. Data Collection

Each participant read the three texts aloud in a quiet classroom environment, with recordings made individually using a high-quality microphone connected to a laptop which runs the recording software anonymized *Audacity*. The recording setup was consistent across pre-tests and post-tests: the microphone was placed at a fixed distance with participants seated facing the microphone to ensure uniform sound quality. Each test lasted approximately 5-7 minutes per participant. Recordings were exported in an uncompressed WAV format (44.1 kHz, 16-bit) to ensure optimal audio quality for subsequent perceptual rating and analysis.

The oral production data collected through the reading task underwent perceptual evaluation. All recordings were anonymized and presented in randomized order to minimize potential rater bias. The rater of the data was the author of the study, who listened to each recording individually and assigned scores on each of the following three dimensions. Then, the evaluation results were shared with four specialists from four different universities who approved these results. The rater listened to each recording individually and assigned scores on each of the three scales.

The recordings were evaluated using a 7-point Likert scale (1=very low, 7=very high) across three perceptual dimensions: comprehensibility, fluency, and accentedness (Munro and Derwing, 2015). Comprehensibility is defined as how easily the speech can be understood (1=extremely difficult to understand, 7=extremely easy to understand). Fluency refers to the smoothness and flow of speech (1=extremely disfluent,7=extremely fluent). Accentedness is defined as how much the accentual variety deviates from a native standard variety (1=extremely heavy accented, 7= sounds native).

3.5. Descriptive Overview

Statistical analyses were then conducted to examine the effects of the training on participants' oral performance. Descriptive statistics were first obtained for each group and variable. These included the Mean (*M*), representing the average score, and the Standard Deviation (*SD*), which indicates the degree of variability of the data around the mean. For each group, paired-sample t-tests were carried out to compare pre- and post-test scores across the three evaluated dimensions (comprehensibility, fluency, and accentedness). These tests determine whether the observed differences before and after the training are statistically significant within the same group. Subsequently, independent-samples of t-tests were performed to compare post-test outcomes between the control and experimental groups, to evaluate the effect of the implementation of gesture aligned with the rhythm-based instruction. Additionally, Cohen's *d* was also calculated to determine the effect size of the observed differences, with the significance level set at $p < .05$.

4. RESULTS

This section presents the outcomes of the statistical analysis conducted to evaluate the impact of rhythm-based instruction, with and without gesture, on learners' oral production. Results are

organized according to group (control vs. experimental), text type (poetry and narrative), and training conditions (trained vs. untrained). The results are interpreted in relation to the three rated variables –comprehensibility, fluency and accentedness– to determine which aspects of participants' oral performance benefited the most from the training. In addition to this, comparisons between the control and experimental groups are also discussed.

4.1. Overview

Descriptive statistics (means and standard deviations) are presented in Table 1 for the control group (n =20), and in Table 2 for the experimental group (n =15), along with t values, significance levels (p) and effect sizes (Cohen's d). Across nearly all conditions, post-test means were higher than pre-test means, and most of these gains reached statistical significance (p <.01).

Text type	Condition	Variable	M. Pre	SD.Pre	M. Post	SD.Post
poetry	trained	comprehensibility	4.05	0.76	5.05	1.00
poetry	trained	fluency	3.85	0.99	4.85	1.18
poetry	trained	accentedness	3.50	0.76	3.95	0.94
narrative	trained	comprehensibility	3.85	0.93	4.75	0.79
narrative	trained	fluency	3.50	0.95	4.35	0.88
narrative	trained	accentedness	3.55	1.04	3.75	0.79
poetry	untrained	comprehensibility	3.60	1.27	4.65	1.09
poetry	untrained	fluency	3.15	1.23	4.10	1.12
poetry	untrained	accentedness	3.30	1.22	3.70	1.26

Table 1: Descriptive analysis for the control group (*M.Pre*= Means of the pre-test. *SD.Pre*=Standard deviation of the pre-test. *M.Post*= Means of the post-test. *SD.Post*= Standard Deviation of the post-test)

In the control group, as shown in Table 1, participants demonstrated measurable improvement from pre-test to post-test. Mean scores increased consistently in both the trained and untrained conditions, suggesting that rhythm-based instruction alone had a positive impact on learners' oral production. For the trained poetry task, *comprehensibility* rose from M=4.05 (SD=0.76) to M=5.05 (SD=1.00), and *fluency* from M=3.85 (SD=0.99) to M=4.85 (SD=1.18), representing gains of approximately one full scale point. Similarly, *accentedness* increased slightly from M=3.50 (SD=0.76) to M=3.95 (SD=0.94), showing a smaller but still positive improvement. Comparable progress was observed in the narrative task, where *comprehensibility* improved from M=3.85 (SD=0.93) to M=4.75 (SD=0.79) and *fluency* from M=3.50 (SD=0.95) to M=4.35 (SD=0.88), both with substantial increases. In contrast, *accentedness* showed a more limited

change, from M=3.55 (SD=1.04) to M=3.75 (SD=0.79). Interestingly, the untrained poetry task also showed notable gains, as *comprehensibility* increased from M=3.60 (SD=1.27) to M=4.35 (SD=1.09), and *fluency* from M=3.15 (SD=1.23) to M=4.10 (SD=1.12). However, the relatively small improvement in *accentedness* from M=3.30 (SD=1.22) to M=3.70 (SD=1.26) suggests that participants' oral performance slightly improved but still retained a perceptible foreign-accented oral production.

Text type	Condition	Variable	M. Pre	SD.Pre	M. Post	SD.Post
poetry	trained	comprehensibility	3.93	1.16	5.06	1.43
poetry	trained	fluency	4.06	1.03	5.13	1.19
poetry	trained	accentedness	3.60	0.98	4.26	1.22
narrative	trained	comprehensibility	4.06	1.09	5.20	1.08
narrative	trained	fluency	3.73	1.09	5.00	1.36
narrative	trained	accentedness	3.66	0.97	4.46	1.55
poetry	untrained	comprehensibility	3.60	1.35	4.93	1.75
poetry	untrained	fluency	3.13	1.19	4.26	1.53
poetry	untrained	accentedness	3.33	1.29	4.13	1.80

Table 2: Descriptive analysis for the experimental group (*M.Pre*= Means of the pre-test. *SD.Pre*=Standard deviation of the pre-test. *M.Post*= Means of the post-test. *SD.Post*= Standard Deviation of the post-test)

As observed in Table 2, the experimental group exhibited more consistent improvement across all measures compared to the control group, which may suggest the added value of gesture-enhanced rhythmic training. In the trained poetry task, *comprehensibility* rose from M=3.93 (SD=1.16) to M=5.06 (SD=1.43), and *fluency* from M=4.06 (SD=1.03) to M=5.13 (SD=1.19). Notably, *accentedness* improved markedly from M=3.60 (SD=0.98) to M=4.26 (SD=1.22), suggesting substantial gains in prosodic naturalness. A similar pattern was observed for the narrative task, where *comprehensibility* increased from M=4.06 (SD=1.09) to M=5.20 (SD=1.08), and *fluency* from M=3.73 (SD=1.09) to M=5.00 (SD=1.36), one of the largest means increases observed. Accentedness also improved from M=3.66 (SD=0.97) to M=4.46 (SD=1.55). The untrained poetry task revealed the most significant evidence of transfer: *comprehensibility* increased from M=3.60 (SD=1.35) to M=4.93 (SD=1.75), being the largest mean increase in the dataset. *Fluency* also showed improvements from M=3.13 (SD=1.19) to M=4.24 (SD=1.53). However, *accentedness* rose more modestly from M=3.33 (SD= 1.29) to M=4.13 (SD=1.80), although the improvement indicates that learners could successfully apply rhythmic principles to unseen materials. Overall, the descriptive statistics for the experimental group reveal substantial and transferable gains across all tasks and dimensions.

4.2. Inferential Statistics

To determine whether the observed improvements between pre- and post-test scores were statistically significant, a series of inferential analyses were conducted. On one hand, paired-sample *t*-tests were performed within each group to examine changes across the three dimensions of comprehensibility, fluency and accentedness. On the other hand, independent-sample *t*-tests compared post-test results between the control and experimental groups. In addition, effect sizes (Cohen's *d*) were calculated to estimate the magnitude of improvement.

Task	Variable	*t*	*p*	*d*
Trained poetry	comprehensibility	6.85	<.001	1.53
Trained poetry	fluency	6.85	<.001	1.53
Trained poetry	accentedness	2.85	<.01	0.64
Trained narrative	comprehensibility	6.90	<.001	1.54
Trained narrative	fluency	6.10	<.001	1.37
Trained narrative	accentedness	4.04	<.001	0.90
Untrained poetry	comprehensibility	8.08	<.001	1.80
Untrained poetry	fluency	6.77	<.001	1.51
Untrained poetry	accentedness	4.04	<.001	0.90

Table 3: Inferential statistics for the control group

Note. p <.05=statistically significant; Cohen's *d*: 0.20=small, 0.50=medium, 0.80+= large effect size

As shown in Table 3, the control group exhibited statistically significant improvements across all three variables, confirming that the rhythmic-based instruction alone had a positive effect on learners' oral performance. The largest *t*-values were observed for comprehensibility and fluency across all tasks (p <.001), with large effect sizes (d=1.37-1.80), indicating that participants produced clearer and more fluent speech after training. However, although accentedness also improved significantly (p <.05), effect sizes were smaller (d=0.64-0.90), suggesting that pronunciation naturalness and stress timing remained challenging. These results imply that the training effectively enhanced rhythmic awareness and speech delivery, though the gains in prosodic accuracy were limited.

Task	Variable	*t*	*p*	*d*
Trained poetry	comprehensibility	4.12	<.001	1.07
Trained poetry	fluency	5.20	<.001	1.34
Trained poetry	accentedness	3.55	<.001	0.92
Trained narrative	comprehensibility	6.85	<.001	1.77
Trained narrative	fluency	5.55	<.001	1.43
Trained narrative	accentedness	3.06	<.01	0.79
Untrained poetry	comprehensibility	5.78	<.001	1.49
Untrained poetry	fluency	4.41	<.001	1.14
Untrained poetry	accentedness	2.87	<.01	0.74

Table 4: Inferential statistics for the experimental group

Note. p <.05=statistically significant; Cohen's d: 0.20=small, 0.50=medium, 0.80+= large effect size

As displayed in Table 4, the analyses revealed consistent and significant gains across all variables (p <.01). Improvements were particularly strong in comprehensibility, with effect sizes ranging from d= 1.07 (trained poetry) to d= 1.77 (trained narrative), indicating that the inclusion of gesture helped learners produce more intelligible speech patterns. Enhancements in fluency were also substantial, with d values between 1.14 and 1.43, reflecting a smoother natural speech flow. However, accentedness showed smaller effects (d=0.74-0.92), although the experimental group achieved a slightly higher improvement than their counterparts in the control group (d=0.64-0.90).

Both groups exhibited noticeable improvements in comprehensibility (Figure 1) from pre-test to post-test across all tasks. The increase was especially noticeable for the experimental group, whose post-test mean values slightly exceeded those of the control group. The largest gains appeared in the trained conditions, and although the control group improved noticeably as well, their post-test scores remained marginally lower. However, when observing the untrained conditions, both groups improved almost the same. These results align with the inferential analyses, which revealed slightly larger effect sizes for the experimental group in the trained conditions.

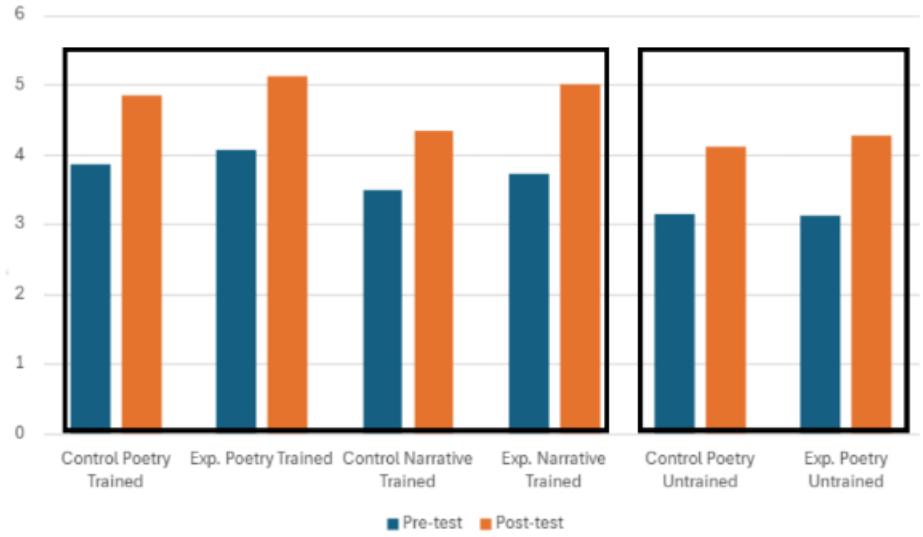

Figure 1. Mean pre- and post-test scores for comprehensibility across tasks and groups[2]

Figure 2 displays the mean scores for fluency. Overall, both groups benefited from rhythmic-based training, but the improvement was more consistent and noteworthy in the experimental group. The graph shows a post-test increase (M.Pre=3.93, M.Post=4.06 for the poetry task, and M.Pre=4.06, M.Post=5.20 for the narrative task), particularly in the trained conditions, where participants who used gestures during rhythmic training reached almost the maximum fluency scores. The control group also improved their fluency levels, but the difference between pre- and post-test scores was smaller, especially in the trained narrative and untrained poetry condition. The most noticeable difference between the experimental and the control group in fluency can be observed in the trained narrative condition, where the experimental group clearly outperformed the control group.

[2] The black rectangles group together the trained tasks on the left side and the untrained tasks on the right side in all figures.

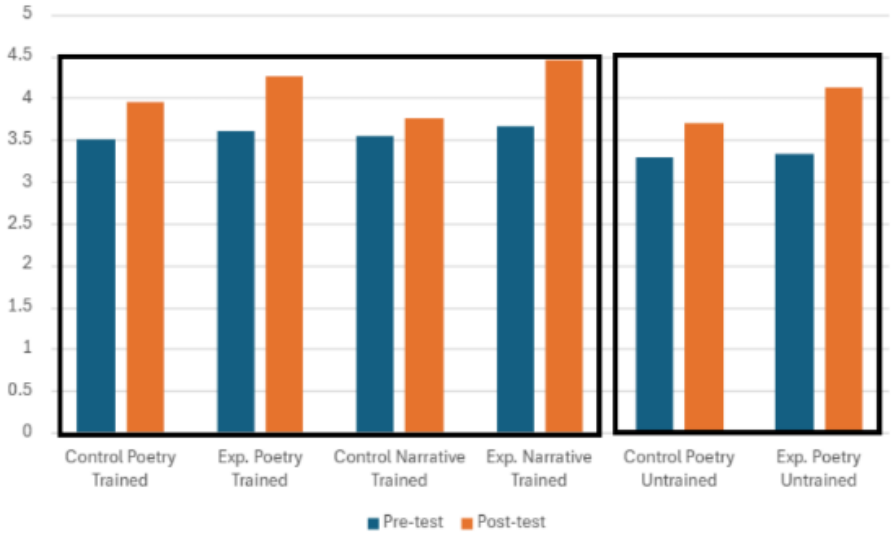

Figure 2. Mean pre- and post-test scores for fluency across tasks and groups

Lastly, as illustrated in Figure 3, accentedness scores improved modestly for both groups, though it was less pronounced than for the other two variables. Again, as observed in comprehensibility and fluency, post-test means were noticeably higher than pre-test means, indicating a general reduction in foreign accent after the training. However, differences between the control and experimental groups were minimal, with both groups showing similar gains in trained and untrained conditions. The most outstanding improvement was observed in the trained narrative condition, where the experimental group's mean score rose visibly after instruction. Nevertheless, both groups showed similar tendencies in the untrained poetry conditions.

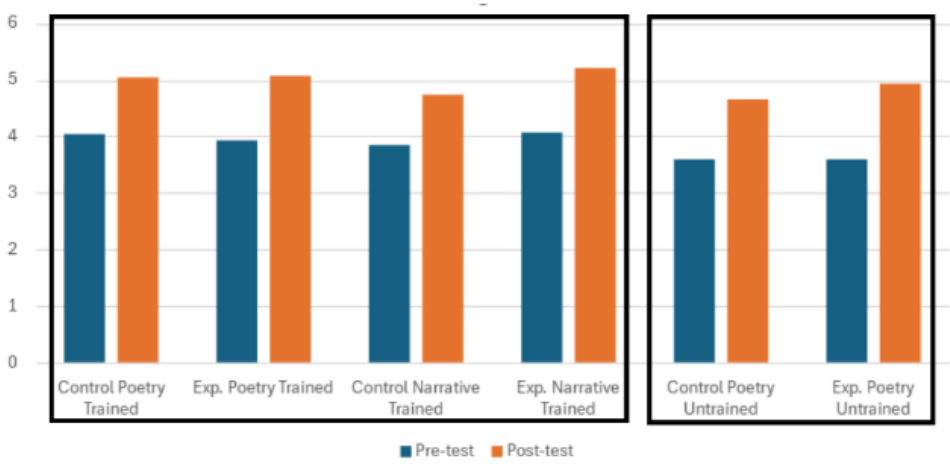

Figure 3. Mean pre- and post-test scores for accentedness across tasks and groups

Taken together, the results presented in Figures 1-3 and the statistical analyses performed show that both instructional conditions led to significant improvements in the learners' production of the English rhythm. However, the experimental group, which incorporated gesture into the rhythm-based instruction, achieved slightly greater gains, particularly in comprehensibility and fluency. The smaller improvement observed in accentedness may indicate that accent reduction remains a more gradual and complex process.

5. DISCUSSION

The present study investigated whether rhythm-based pronunciation training, supported by the implementation of gesture, could enhance Spanish learners' production of English rhythm, focusing on three variables: comprehensibility, fluency and accentedness. Overall, the results confirmed that both instructional groups benefited significantly from the training, as reflected in higher post-test means across all conditions. However, the slightly higher gains observed in the experimental group suggest that gesture served as a facilitator in helping learners to internalize and reproduce stress-timed rhythm patterns. The smaller improvement in accentedness compared to the other two variables indicates that short-term interventions may not be enough to produce reductions in foreign accent. A closer look at classroom behaviour offers an additional layer of interpretation. During training activities and even throughout the post-test recordings, 14 out of 20 participants in the control group and 13 out of 15 participants in the experimental group were observed spontaneously using hand or body movements while performing the rhythmic tasks. This suggests that gesture production emerged naturally, even without explicit instruction, as participants in the control group did not receive any gesture input. Such spontaneous use of gesture might explain why the outcomes between the two groups were not dramatically different. This implementation indicates that gesture can be also an intuitive and natural strategy that learners employ to coordinate stress and rhythm, reinforcing the idea that rhythm learning and production may be naturally embodied.

The outcomes of this study align with the previous findings on the positive impact of rhythm-based instruction on L2 pronunciation. Studies such as Estebas-Vilaplana and Soláns (2019) have shown that rhythm-focused training enhances learners' fluency by reinforcing their awareness of stress timing. In addition to this, the improvement observed in the experimental group supports the idea that gestures can function as a facilitator and an effective tool to make

prosodic features more perceptible to learners, especially for learners from syllable-timed language backgrounds such as Spanish.

However, several limitations may be acknowledged. First, the sample size was relatively small (*n*=35). Second, the intervention lasted only two sessions across two days, offering very limited exposure to the rhythm-based instruction. While notable improvement was observed within this short period, a longer program would likely produce stronger effects, particularly in accentedness, as these two sessions have demonstrated the potential of rhythm-based instruction. Future research could incorporate the remaining modules from Estebas-Vilaplana and Soláns' MOOC (2019), to expand the duration of the rhythmic training. Finally, although perceptual ratings provide valuable insights, future studies could combine them with acoustic analyses (e.g., measures of vowel reduction) to measure variations in phonetic segmental details.

6. CONCLUSION

This research provides empirical evidence that rhythm-based pronunciation training can lead to measurable improvement in the prosodic performance of Spanish learners of English, even within a short intervention period. By focusing on three variables –comprehensibility, fluency and accentedness–the results revealed significant gains across all measures for both instructional groups, suggesting that the exposure to rhythmic training contributes to a better grasp of speech rhythm and English stress timing. These findings also highlight the importance of prosodic awareness as a fundamental role in second language acquisition.

A key contribution of this study lies in the implementation of gesture as a potential facilitator of rhythmic learning. While both the control and experimental groups improved, the slightly higher gains in comprehensibility and fluency observed in the experimental group, suggest that gesture may assist learners in internalizing prosodic features. The integration of gesture not only aided participants in their tasks but also facilitated their production of stress-timed patterns, leading to a more comprehensive, fluent and rhythmically natural speech.

Moreover, the fact that gesture use emerged in both groups, underscores the point that learners appear to instinctively use physical movement as a strategy to manage stress timing. In this sense, the link between rhythmic movement and prosodic timing helps understand how learners from syllable-timed backgrounds can transition towards stress-timed production.

Furthermore, although this was a small-scale, short-term study, the consistency of improvement across both groups is remarkable. The fact that learners showed meaningful progress after only two sessions suggest that rhythm-based training can have an immediate impact on prosodic awareness, even when delivered in condensed classroom formats. A further innovative aspect of this study lies in its adaptation of digital resources to an in-person university context. The training materials were based on Estebas-Vilaplana and Soláns' MOOC (2019), originally designed for online delivery, but were formulated for classroom use in a controlled academic setting. The results of this adaptation, observed in both groups, indicate that these materials can be effectively adapted to face-to-face settings, maintaining their embodied practice.

Finally, the results of this research demonstrate that Spanish learners can improve their English rhythmic control through focused, multimodal practice that engages perception and movement. In addition to this, extending the training over a longer period would allow us to observe whether the gains found here persist or deepen, as well as exploring different implementation strategies of different gesture types. Moreover, incorporating acoustic analyses would further strengthen the understanding of these improvements and measure these changes in timing, duration and stress placement. In conclusion, this study highlights how rhythm and gesture converge to shape the acquisition of English prosody, offering new insights into how rhythmic awareness can be both taught and practiced in the L2 classroom.

ACKNOWLEDGEMENTS

This research has been possible thanks to the invaluable help and collaboration of Dr. Pilar Prieto Vives and Dr. Florence Baills during my PhD studies at the Universitat Pompeu Fabra.

REFERENCES

BAILLS, Florence, ALAZARD-GUIA, Charlotte, and Pilar PRIETO (2022): "Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy", *Applied Linguistics*, 43.4, pp. 776-804.

BÁRCENA, Elena, READ, Timothy, MARTÍN-MONJE, Elena, and María D. CASTRILLO (2014): "Analysing Student Participation in Foreign Language MOOCs: A Case Study",

EMOOCs 2014: European MOOCs Stakeholders Summit, eds. U. Cress and C. Delgado Kloos, Lausanne, P.A.U. Education S.L., pp. 11-17.

BARREIRO, Silvia, ESTEBAS-VILAPLANA, Eva, and Isabel SOTO (2005): "Teaching Phonetics through Singing and Reciting", *Proceedings of the Phonetics Teaching and Learning Conference 2005,* London, University College London (UCL), pp. 1-4.

--- (2007): "From Prosody to Sounds: A New Approach to Phonetics Teaching and Learning", *Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva. Perspectivas aplicadas entre disciplinas: actas del XXIV Congreso Internacional de AESLA. [Recurso electrónico],* p.38, Madrid, UNED.

--- (2009): *Antología Oral de la Literatura Inglesa: From Beowulf to Virginia Woolf*, Madrid, Universidad Nacional de Educación a Distancia.

CRUTTENDEN, Alan (2014): *Gimson's Pronunciation of English*, Oxon, Routledge.

CUTLER, Anne (2015): "Lexical Stress in English Pronunciation", *The Handbook of English Pronunciation*, Malden, Wiley-Blackwell, pp. 106-124.

DAUER, Rebecca M. (1983): "Stress-timing and Syllable-timing Reanalyzed", *Journal of Phonetics*, 11.1, pp. 51-62.

DOWNES, Stephen (2008): "Places to Go: Connectivism & Connective Knowledge", *Innovate: Journal of Online Education*, 5.1, pp. 2-6

ESTEBAS-VILAPLANA, Eva, and Mariángel SOLÁNS (2019): *The Acquisition of English Pronunciation through Songs and Literary Texts*. [MOOC]. UNED. https://iedra.uned.es/courses/course-v1:UNED+Proning+Permanente/about.

--- (2020a): "The Role of a Pronunciation LMOOC in Higher Education Studies", *Journal of Interactive Media in Education*, 21.1, pp. 1-10. https://doi.org/10.5334/jime.589.

--- (2020b): "Los MOOC en el proceso de enseñanza y aprendizaje universitaria: El caso de un LMOOC de pronunciación inglesa a través de textos literarios", *Estrategias de aprendizaje en el entorno digital: propuestas desde el ámbito universitario,* Madrid, Egregius Ediciones, pp. 13-28.

ESTEVE-GILBERT, Nuria, and Pilar PRIETO (2018): *The Development of Prosody in First Language Acquisition*, Amsterdam, John Benjamins.

GOLDING-MEADOW, Susan (2005): *Hearing Gesture: How our Hands Help us Think*. Harvard University Press, Cambridge, Harvard University Press.

IVERSON, Jana M., and Susan, GOLDING-MEADOW (2005): "Gesture Paves the Way for Language Development", *Psychological Science*, 16.5, pp. 367-371.

JUSSLIN, Sofia, KORPINEN, Kaisa, LILJA, Niina, MARTIN, Rosemary K., LEHTINEN-SCHNABEL, Johanna, and Eva ANTTILA (2022): "Embodied Learning and Teaching Approaches in Language Education: A Mixed Studies Review" *Educational Research Review*, 37, pp 1-22.

KEHOE, Margaret, and Carol STOEL-GAMMON (1997): "The Acquisition of Prosodic Structure: An Investigation of Current Accounts of Children's Prosodic Development", *Language*, 2, pp. 113-144.

LAHOZ, Jose M. (2012): "Syllable, Accent, Rhythm: Typological and Methodological Considerations for Teaching Spanish as a Foreign Language", *Revista internacional de lenguas extranjeras / International Journal of Foreign Languages*, 1, pp. 129-150. https://doi.org/10.17345/rile1.13.

LI, Peng, BAILLS, Florence, BAQUÉ, Lorraine, and Pilar PRIETO (2023): "The Effectiveness of Embodied Prosodic Training in L2 Accentedness and Vowel Accuracy", *Second Language Research*, 39.4, pp. 1077-1105.

LOMOTEY, Charlotte F. (2013): "Does First Language Prosodic Transfer Affect Second Language Prosody?", *The Journal of the Acoustical Society of America*, 134.5, pp. 4246-4247.

MUNRO, Murray J., and Tracey M. DERWING (2015): "A Prospectus for Pronunciation Research in the 21st Century: A Point of View", *Journal of Second Language Pronunciation*, 1.1, pp. 11-42.

PIKE, Kenneth L. (1945): *The Intonation of American English*, Ann Arbor, University of Michigan Press.

ROACH, Peter (2009): *English Phonetics and Phonology Paperback with Audio CDs (2): A Practical Course*. Cambridge, Cambridge University Press.

TROFIMOVICH, Pavel, and Wendy BAKER (2006): "Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech", *Studies in second language acquisition*, 28.1, pp. 1-30.

WANG, Xue (2022): "Segmental Versus Suprasegmental: Which One Is More Important to Teach?", *RELC Journal*, 53.1, pp. 194-202.

ZHANG, Yuan, BAILLS, Florence, and Pilar PRIETO (2020): "Hand-clapping to the Rhythm of Newly Learned Words Improves L2 Pronunciation: Evidence from Training Chinese Adolescents with French Words", *Language Teaching Research*, 24.5, pp. 666-689.

APPENDIX 1. ADAPTED HANDOUTS FOR SESSIONS 1: "THE RHYTHM IN ENGLISH AND IN SPANISH", AND 2: "RHYTHM AND ACCENTUATION"

MÓDULO 1: El ritmo en inglés. CLASE 1: El ritmo en inglés y en español

1. *Práctica de percepción*. Escucha los dos audios y responde: ¿Son iguales o distintos?

   Érase una vez una niña pequeña que vivía en un pueblo cerca del bosque

2. *Práctica de percepción*. Escucha los audios y responde: ¿Son iguales o distintos?

   Once upon a time there was a little girl who lived in a village near the forest

3. *Práctica de percepción*. Escucha los audios y responde: ¿Es ritmo acentual o silábico?

   • ritmo acentual • ritmo silábico

   a.
   b.
   c.

4. *Práctica de producción*. Reproduce el texto con ritmo acentual.

   Once upon a time there was a little girl who lived in a village near the forest

5. *Práctica de percepción*. ¿Qué ritmo identificas?

   If all the world and love were young, • ritmo yámbico: _* _* _*
   And truth in every shepherd's tongue,
   These pretty pleasures might me move • ritmo trocaico: *_ *_ *_
   To live with thee and be thy love.

6. *Práctica de producción*. Reproduce el texto con ritmo yámbico.

   If all the world and love were young,
   And truth in every shepherd's tongue,
   These pretty pleasures might me move
   To live with thee and be thy love.

7. *Práctica de percepción*. ¿Qué ritmo identificas?

Merrily, merrily • ritmo yámbico: _* _* _*
Shall I live now
Under the blossom that • ritmo dactílico: *_ _ *_ _ *__
hangs on the bough.

8. *Práctica de producción*. Reproduce el texto con ritmo dactílico.

Merrily, merrily
Shall I live now
Under the blossom that • ritmo dactílico: *_ _ *_ _ *__
hangs on the bough.

9. *Práctica de percepción*. ¿Qué ritmo identificas?

• ritmo acentual • ritmo silábico

(There are) dark shadows on the earth but its
lights are stronger in the contrast.

10. *Práctica de producción*. Reproduce el texto con ritmo acentual.

(There are) dark shadows on the earth but its
lights are stronger in the contrast.

11. *Práctica de producción*. Reproduce el texto con ritmo trocaico.

Double, double toil and trouble;
Fire burn and cauldron bubble. • ritmo trocaico: *_ *_ *_
Scale of dragon; tooth of wolf;
Witches' mummy; maw and gulf

MÓDULO 1: El ritmo en inglés. CLASE 2: Ritmo y acentuación
10:43

1. *Práctica de producción.* Marca con un asterisco (*) las sílabas acentuadas.

   He was my North, my South, my East and West.

   My working week and my Sunday rest.

   My noon, my midnight, my talk, my song.

   I thought that love would last for ever: I was wrong.

2. *Práctica de percepción.* Identifica si es lectura neutra o contrastiva.

   a. **He was my North :**
   b. **He was my North :**

3. *Práctica de percepción*. Escribe 1, 2, 3 o 4 al lado de cada enunciado según escuches a Eva diciendo el enunciado número 1, 2, 3, 4.

   Clip número 1:
   Clip número 2:
   Clip número 3:
   Clip número 4:

4. *Práctica de producción en inglés y en español*. Repite las 3 formas distintas de cada enunciado que realiza Eva. Luego, elige una de ellas, asegúrate de reproducirla bien en voz baja, y luego reprodúcela en voz alta. Tu compañero ha de identificar cuál de las dos estás haciendo

   a. *Reader, I *married him
   b. *Reader, I *married *him
   c. *Reader, *I *married him

5. *Práctica de producción.* Repite a coro con Eva este ejemplo

6. *Práctica de producción.* Repite a coro con Eva este ejemplo. Luego, marcad por parejas las sílabas acentuadas con un asterisco (*) y reproducidlo en voz alta

It is a truth universally acknowledged

that a single man in possession of a good

fortune, must be in want of a wife

7. *Práctica de producción.*
   a. ¿Cómo sonarían estos ejemplos si se acentuaran "was" y "would"? Dilos en voz alta con tu compañero/a. ¿Hay algún cambio de significado ahora?

      1. The *old *man **was** *dreaming a*bout the *lions
      2. *Mrs. *Dalloway *said she **would** *buy the *flowers her*self

   b. ¿Están acentuados "wasn't", "wouldn't" y "was"? ¿Alguna sugerencia de por qué puede ser así?

      3. Everything is dangerous, my dear fellow. If it **wasn't** so, life **wouldn't** be worth living
      4. How wild it **was**, to let it be