Research on Spanish Beginners Correctly Acquire p/b and r/l? A Case Study of First-Year University Students in a Spanish Department

Linrui Wang

Sichuan International Studies University, Chongqing, China, 400031
Email: 3235428810@qq.com

Manuscript received July 4, 2026; accepted August 1, 2026; published September 23, 2026

Abstract—**This study investigates how 60 Chinese beginners acquire Spanish /p, b/ and /r, l/ phonemes. Drawing on acoustic analysis via Praat and subjective questionnaires, this study identifies two distinct mechanisms of phoneme acquisition. /p, b/ Contrast: Difficulty stems from perceptual category assimilation. Learners often mistake the Spanish voiced [b] for the Mandarin unaspirated voiceless [p]. A significant correlation (r = 0.52) exists between their perception and production accuracy. /r, l/ Contrast: Mastery is primarily a motor skill challenge. Perception and production are decoupled (r = 0.18), as many students identify the sound but fail at articulatory execution. Furthermore, dialects with "n/l" confusion significantly increase /r/ lateralization error rates (42.7%). Phonetic anxiety further impairs perceptual and production accuracy, especially for voiced consonants that demand precise articulatory control. The study advocates for tiered interventions: utilizing acoustic visualization for /p, b/ to restructure perceptual categories and focused muscle memory training for the /r/ trill.**

Keywords—**Spanish pronunciation acquisition, voiceless/voiced plosives, Alveolar Trill, perceptual assimilation, motor learning, dialectal transfer**

I. Introduction

In foreign language education, pronunciation is the foundational stone of professional competence. For first-year Spanish majors at Chinese universities, effective pronunciation acquisition plays a decisive role, as it underpins the development of subsequent language skills and profoundly shapes learners' professional identity and learning confidence. Within the Spanish phonological system, the voiceless/voiced plosives /p/ vs. /b/ and the alveolar trill /r/ vs. the lateral /l/ constitute primary challenges for Chinese learners. These phonemes are phonemic in Spanish, and their correct acquisition is key to achieving an authentic accent and accurate semantic transmission.

Existing research on Spanish pronunciation acquisition among Chinese learners has yielded preliminary findings, yet substantial methodological limitations still persist. Most studies rely on qualitative analysis and subjective auditory descriptions, which lack objective, quantitative data to reveal the acoustic nature of errors. For example, Zhang and Li [1] noted phenomena like "devoicing" or "aspiration" without employing acoustic software to measure specific parameters like Voice Onset Time (VOT). Similarly, Sun [2] emphasized L1 negative transfer from Mandarin but overlooked other significant variables such as dialectal backgrounds. Furthermore, affective factors like anxiety have long been marginalized in empirical models, leaving phenomena like "knowing the theory but being unable to produce the sound" inadequately explained. While some scholars advocate for visual feedback software, there is a scarcity of controlled, longitudinal experiments to confirm its efficacy for first-year students. Accordingly, a notable research gap remains in multidimensional quantitative analysis that incorporates language transfer, dialectal interference, and affective factors to elucidate phoneme acquisition mechanisms.

This study employs a mixed-methods design, integrating Praat-based acoustic analysis with subjective questionnaires to examine the acquisition of Spanish /p, b/and/r, l/among 60 Chinese beginner learners. The research content focuses on quantifying pronunciation error features through VOT distribution and spectrogram energy patterns. Additionally, it examines the influence of regional dialect backgrounds and psychological anxiety on production accuracy through structured questionnaires and in-depth interviews.

Accurately describe and quantify learners' pronunciation error features using objective acoustic data. Construct a multidimensional framework for analyzing influencing factors, including native language transfer, dialectal backgrounds, and psychological variables. Test and integrate relevant theories of phonetic acquisition, such as the Contrastive Analysis Hypothesis (CAH) and the Speech Learning Model (SLM) [3]. Propose pedagogically intervention strategies grounded in empirical evidence to inform teaching practice.

This study provides a scientific and systematic empirical analysis that precisely characterizes error patterns. By elucidating the mechanisms of various interfering factors, it provides a valid theoretical foundation for helping learners establish a solid pronunciation foundation, address fossilized errors, and boost their confidence in oral communication.

II. Literature Review

A. First Language (L1) Transfer

Mandarin lacks genuine voiced plosives, and its unaspirated voiceless stops bear acoustic similarities to Spanish voiced plosives. This leads learners to assimilate target sounds into native categories, causing perceptual and productive confusion between /p/ and /b/ [3, 4]. Similarly, the absence of an apical trill /r/ in Mandarin often results in learners substituting it with /l/, the Mandarin Pinyin "r" or the English retroflex approximant. This dual negative transfer effect from Mandarin and English constitutes a primary source of Spanish pronunciation errors.

B. Regional Dialectal Influence

China's linguistic diversity exacerbates acquisition difficulties through dialectal transfer. In many southern dialect areas, such as Xiang and Southwestern Mandarin, the non-distinction between /n/ and /l/ is common [5, 6]. This "fossilized error" directly transfers to the acquisition of Spanish /r/ and /l/, creating an additional obstacle. Empirical data shows that learners from "n/l non-distinction" areas exhibit significantly higher /r/ lateralization error rates (42.7%) compared to those from "n/l distinction" areas (18.9%).

C. Psychological and Affective Factors

Pronunciation entails complex articulatory control and self-monitoring processes, both of which are highly sensitive to the affective filter [7, 8]. First-year students often experience anxiety and embarrassment in evaluative situations, leading to conservative production strategies. Research demonstrates that high levels of phonetic anxiety deplete cognitive resources and impede the automatization of pronunciation, particularly affecting voiced plosives that require precise vocal fold control.

D. Research Gap

Existing literature is constrained by qualitative dominance and simplistic explanatory frameworks. There is a lack of research that establishes reliable quantitative associations between dialectal features and error rates, or that incorporates affective variables into empirical models. Furthermore, the specific efficacy of visual feedback for different error types (perceptual vs. motor) remains inadequately explored [9, 10].

E. Research Questions

Based on the identified gaps, this study addresses the following four questions:

(1) What stage-specific characteristics are evident in pronunciation production? Do significant differences exist in the error rates and error types for /p, b/ and /r, l/ across three tasks of varying complexity—word reading, sentence reading, and impromptu expression? Do phonetic errors vary systematically with increases in task cognitive load?

(2) Do perceptual barriers constrain production levels? What is the relationship between learners' auditory discrimination ability for the /p, b/ voicing contrast and the /r, l/ quality difference, and their actual pronunciation accuracy? Is insufficient perceptual capacity a major prerequisite for phonetic production errors?

(3) What influence do dialect and psychological factors exert? Does a learner's dialect background (specifically, dialects exhibiting n/l confusion) significantly predict their level of difficulty in acquiring /r, l/? Is a self-reported high level of phonetic anxiety during phonetic production tasks statistically correlated with reduced pronunciation quality?

(4) Can visual feedback optimize the acquisition path? Compared with traditional mimicry and rote repetition methods, can spectrographic visual feedback via Praat, adopted as a pedagogical tool, more effectively help learners revise articulatory gestures of target phonemes and improve the accuracy and stability of their phonetic production?

III. RESEARCH DESIGN

A. Participants

This study recruited 60 participants from the 2025 intake of first-year Spanish majors at a mainland Chinese university. To ensure the representativeness and comparability of the sample, participants were selected based on the following criteria. First, all participants were true beginners with no prior formal study of Spanish. They possessed comparable scores on the National College Entrance Examination (Gaokao) English section. Furthermore, all had completed an identical four-week foundational Spanish pronunciation course with the same content and instructional hours upon enrollment. This controlled for variations in prior learning experience and instructional input. Second, participant selection considered the representativeness of regional dialect backgrounds. Based on their hometowns and self-reported dialect backgrounds, participants were purposively recruited to cover Northern Mandarin, Wu, Xiang, and Southwestern Mandarin dialect areas. This sampling strategy aimed to test the influence of different dialectal phonological features, particularly the presence or absence of the /n, l/ distinction and the status of voiced consonants, on the acquisition of the target Spanish sounds. All participants gave written informed consent prior to their involvement in this study.

B. Speech Production Tasks

To comprehensively capture learners' productive abilities, the study designed three progressive recording tasks. All tasks were conducted in a quiet phonetics lab using the same model of high-fidelity microphone. Table 1 specifies the specific content, implementation procedures, and research objectives of each task.

Table 1. Speech production task design

Task Module	Sample Test Content	Research Purpose
Module A: Word Reading	Reading minimal pair words containing the target phonemes, e.g., "papa-baba", "pero-pelo".	To examine learners' static discrimination and production of the voicing and trill/lateral contrasts at the word level, isolating them from connected speech and semantic interference.
Module B: Sentence Reading	Reading specific sentences containing the target phonemes, e.g., "Quieres un poco de vino o de pino?"	To investigate the realization of target phonemes in connected speech, analyzing whether they are affected by coarticulation, intonation, speech rate, etc.
Module C: Impromptu Production	Viewing a picture containing items like "perro" (dog), "bala" (bullet), "todo" (all) and providing a one-minute oral description.	To examine learners' unconscious production habits of the target phonemes in a natural state with no written prompt and with attention focused on meaning, assessing the degree of pronunciation internalization.

C. Objective Data Processing and Acoustic Quantification

All recorded samples were imported into Praat speech analysis software (version 6.3.09) for acoustic parameter extraction and quantitative analysis, transforming speech production into objective measurements [9, 10].

For the voiceless/voiced plosives /p, b/, the analysis centered on Voice Onset Time (VOT) as the key metric [1, 11]. VOT is defined as the time interval between the release of the plosive closure and the onset of vocal fold vibration. On spectrograms, it was measured between the plosive burst's vertical line (release point) and the start of periodic voicing striations. For the context of this study, measurements were interpreted as follows: a VOT value greater than +20 ms was categorized as an "aspiration error", indicating the use of Mandarin aspirated voiceless stop articulation for Spanish unaspirated /p/ or voiced /b/. A VOT value between 0 and +10 ms, or a negative value, coupled with acoustic features of a voiceless consonant, was categorized as a "devoicing error", indicating a lack of vocal fold vibration for the intended voiced /b/.

For /r, l/, the analysis focused on energy distribution patterns in the spectrogram. The standard Spanish alveolar trill /r/ typically displays one or more very brief, clear energy interruptions (gaps) in the spectrogram, with clear formant structures before and after. If the target position for /r/ showed a sustained, smooth formant transition akin to the lateral /l/ with no energy gap, it was judged as a "lateralization error", meaning /l/ was substituted for /r/. Other attempted articulations for /r/, such as fricativization or substitution with an English-style /r/, were also noted.

D. Subjective Attribution Survey

To gain a deeper understanding of the causes behind the objective data, a subjective attribution survey was administered immediately after the speech recording tasks. This survey consisted of two parts: retrospective in-depth interviews and a structured questionnaire.

The in-depth interviews followed a semi-structured format and were conducted with 10-15 learners who exhibited typical and stable errors in the recording tasks. Interview questions centered on learners' metacognitive experiences during production, for example: "When pronouncing 'baba', were you consciously controlling your vocal folds? How would you describe the difference between the sound you produced and the model pronunciation?" "When attempting the trill in 'perro', how did the tip of your tongue feel? Did you deliberately change your articulation due to fear of failure?" The interviews aimed to gather firsthand accounts of learners' reflections on their own pronunciation process, their attributions for difficulties, and their affective experiences.

The questionnaire was administered uniformly to all 60 participants after the speech recordings. It employed a combination of 5-point Likert scale items and open-ended questions, covering three main dimensions. First, the Cognitive Interference dimension assessed learners' perceived clarity of the target phoneme contrasts and the strength of transfer from English pronunciation habits (e.g., "I find it difficult to distinguish between 'papa' and 'baba' in Spanish", "I often unconsciously use my English 'p' mouth position for Spanish 'p'"). Second, the Affective Filter dimension measured learners' anxiety levels in classroom and testing situations, as well as feelings of embarrassment or avoidance towards specific sounds like the trill (e.g., "I feel nervous when reading Spanish aloud in class") [7]. Third, the Dialect Background Self-Assessment dimension collected learners' self-evaluation of the influence of their local dialect on their Mandarin pronunciation (e.g., "My Mandarin is influenced by my dialect; my /n/ and /l/ are not clear") [5, 6]. Questionnaire data were cross-analyzed with the objective acoustic measurements to explore potential associations between psychological, cognitive factors, and pronunciation performance.

IV. Comparative Analysis Pathways and Acquisition Mechanism Verification

A. Cross-Comparative Pathways

To comprehensively examine the effects of regional dialectal, affective, and cognitive factors on phonetic acquisition, this study performed cross-group comparisons and statistical analyses based on objective acoustic data, background questionnaires, and interview transcripts. The statistical methods employed included independent-samples t-tests, Mann-Whitney U tests, one-way ANOVA, and Pearson correlation analysis. All data analyses were conducted using SPSS, with the statistical significance level set at $\alpha = 0.05$.

1) Comparative analysis of objective errors and regional background

To test the dialectal negative transfer hypothesis, participants were first divided into two groups based on self-reported dialect background: the "n/l non-distinction dialect area" group (mainly including Xiang, some Wu, and Southwestern Mandarin dialects) and the "n/l distinction dialect area" group (mainly Northern Mandarin) [5, 6]. An independent-samples t-test was conducted to compare the overall rate of /r/-to-/l/ substitution errors across all phonetic production tasks between the two groups. The results are shown in Table 2.

Table 2. Independent samples t-test results for /r/ Lateralization error rates by dialect background

Variable	n/l Non-distinction Area (n = 28)	n/l Distinction Area (n = 32)	t	df	*p*	Cohen's d
/r/ Lateralization Error Rate (%)	42.68 (16.25)	18.91 (12.07)	6.24	58	<0.001	1.63

Data in table are mean (standard deviation). Table 2 shows that the mean /r/ lateralization error rate for the "n/l non-distinction" group was 42.68%, significantly higher than the 18.91% for the "n/l distinction" group ($t(58) = 6.24$, $p < 0.001$). The effect size, Cohen's d=1.63, is large, indicating the dialect background difference has high practical significance for /r/ acquisition. This result strongly supports the research hypothesis. The phonological processing habit of confusing /n/ and /l/ in the native dialect systematically transfers to the distinction between Spanish /r/ and /l/. Accordingly, learners from such dialectal regions tend to substitute the familiar /l/articulatory pattern for the target Spanish /r/phoneme.

However, a one-way ANOVA showed no significant difference in the "devoicing error rate" for the voiced plosive /b/ among learners from different dialect backgrounds

($F(3,56) = 1.57$, $p = 0.207$). This indicates that the difficulty in acquiring the /p, b/ voicing contrast is a universal issue, transcending specific dialect backgrounds. The root cause more likely lies in Mandarin Chinese's inherent lack of a true voiced plosive category, combined with interference from English voiced plosive pronunciation habits. Together, these factors constitute a shared acquisition challenge for all Mandarin native learners.

2) Comparative analysis of objective errors and emotional state

To examine the effect of the affective filter, participants were divided into a "High Anxiety Group" (top 30%, n = 18) and a "Low Anxiety Group" (bottom 30%, n = 18) based on their scores on the anxiety dimension from the "Psychological and External Influencing Factors" section of the questionnaire [7]. As the anxiety scores violated normality assumptions, the non-parametric Mann-Whitney U test was used to compare the overall pronunciation accuracy in the impromptu production task between the two groups. Results are in Table 3.

Table 3. Mann-Whitney U Test results for overall accuracy in the impromptu production task by anxiety level

Group	N	Mean Rank	Sum of Ranks	U	Z	p
Low Anxiety Group	18	23.61	425.00	285.00	-2.48	0.013
High Anxiety Group	18	13.39	241.00			

The test result ($U = 285$, $p = 0.013$) indicates that the Low Anxiety Group's overall pronunciation accuracy in the impromptu task (Mean Rank 23.61) was significantly higher than the High Anxiety Group's (Mean Rank 13.39). This confirms the Affective Filter Hypothesis [7]: heightened anxiety consumes cognitive resources, interferes with the automatization of phonetic production, and consequently impairs L2 oral performance quality.

Further analysis revealed that anxiety's effect is phoneme-specific. A Mann-Whitney U test showed that the High Anxiety Group produced significantly more "devoicing errors" on voiced plosives (/b, d, g/) than the Low Anxiety Group ($p = 0.022$). This aligns with learners' interview descriptions, such as "my throat tightens when I'm nervous, and I can't feel the vibration." It suggests anxiety particularly affects the production of voiced sounds requiring fine vocal fold control. In contrast, the two groups showed no significant difference in error rates for the /r/ sound ($p = 0.361$). This indicates that for /r/, a phoneme requiring a novel motor skill, the primary difficulty stems from the complexity of motor learning itself. Anxiety did not alter the fundamental error pattern of /r/, but rather undermined the production stability of sounds that are partially acquired yet not fully automatized, such as voiced plosives.

3) Comparative analysis of subjective perception and objective production

To investigate the relationship between perception and production, Pearson product-moment correlation coefficients were calculated between learners' subjective perceptual self-ratings and their objective production accuracy rates. Results are summarized in Table 4.

Table 4. Pearson correlation analysis between subjective perception self-ratings and objective production accuracy

Phoneme Contrast	Mean (M)	SD	Pearson's r	p (2-tailed)
/p/ vs /b/ Perception Self-Rating	3.45	0.89	0.52	<0.001
/p/ vs /b/ Production Accuracy (%)	71.33	18.42		
/r/ vs /l/ Perception Self-Rating	3.20	1.05	0.18	.174
/r/ vs /l/ Production Accuracy (%)	58.67	24.15		

As shown in Table 4, for the /p, b/ voicing contrast, a significant moderate positive correlation exists between learners' perception self-ratings and their production accuracy ($r = 0.52$, $p < 0.001$). This result supports the core tenet of the Speech Learning Model (SLM) [4]: accurate perception of a new phonemic category is a crucial prerequisite for successful production. Learners with greater auditory sensitivity to the voicing distinction demonstrated a stronger ability to articulate this contrast accurately in their own speech.

However, for the /r/ vs /l/ contrast, the correlation between perception self-rating and production accuracy was weak and not statistically significant ($r = 0.18$, $p = 0.174$). The data revealed a key phenomenon: 15 learners (25%) had high perception self-ratings (≥4), indicating they believed they could distinguish the sounds well, yet their /r/ production accuracy was below 50%. Retrospective interviews with these learners indicated their primary difficulty lay in motor execution. Typical statements included: "I can hear that my pronunciation differs from the teacher's. I know the teacher's tongue produces a trill, but my tongue cannot move that fast," and "I can mentally conceptualize the trill, but as soon as I attempt to produce it, the sound turns into an l." This clearly indicates that the bottleneck in /r/ acquisition lies not primarily in auditory perception, but in the motor learning and neuromuscular control process of translating an accurate auditory image into precise articulatory gestures. Perceptual ability is necessary but far from sufficient for accurate L2 sound production.

B. In-Depth Discussion of Core Acquisition Mechanisms

Building on the quantitative results from the cross-comparisons and integrating qualitative analysis of interview data, this study moves beyond surface-level error description to probe the underlying mechanisms of the acquisition difficulties Chinese Spanish beginners face with the /p, b/ and /r, l/ contrasts. The analysis indicates that these challenges stem from distinct cognitive and psychophysiological processes, thereby necessitating targeted theoretical explanations and pedagogical interventions.

1) Acquisition mechanism for /p, b/: Perceptual category restructuring

The findings strongly support the core argument of the Speech Learning Model (Flege, 1995) [4]: for L2 sounds that have a "similar" counterpart in the L1, the primary obstacle is perceptual assimilation. The Mandarin stop consonant system employs aspiration as its primary distinctive phonological feature. Its unaspirated voiceless stop [p] overlaps acoustically, particularly in the Voice Onset Time (VOT) range, with the Spanish voiced stop [b]. This acoustic similarity leads learners to assimilate the Spanish voiced /b/ into their native voiceless /p/ category, making it difficult to perceptually distinguish *papa*from *baba*.

The data analysis provides two pieces of evidence. First, the significant positive correlation between perceptual self-rating and production accuracy ($r = 0.52$) shows that auditory sensitivity to phonetic quality differences directly impacts pronunciation accuracy. Learners who reported "not hearing a difference" exhibited the most widespread and stable devoicing in their production. Second, the lack of significant difference in /p, b/ error rates across dialect backgrounds rules out specific interference from dialectal phonology. It points to the root of the problem residing in the Mandarin-based perceptual system common to all native speakers. Therefore, acquiring the /p, b/ contrast is essentially a process of perceptual category restructuring. Learners must break free from the constraints of their L1 categories and establish a new mental representation based on the distinctive feature of [±voice] to provide a target for correct production.

2) Acquisition mechanism for /r, l/: Motor skill learning

In stark contrast, the learning difficulty associated with the Spanish /r/ phoneme manifests primarily as an articulatory motor skill challenge. The data shows an extremely low, non-significant correlation between /r/ production accuracy and perception self-rating ($r = 0.18$). Many learners could clearly distinguish the auditory difference between standard /r/ and /l/ but could not reproduce it. Frequent interview descriptions like "stiff tongue tip", "loss of control", and "it goes wrong as soon as I try too hard" precisely point to failure at the level of articulatory execution.

This phenomenon can be interpreted within the theoretical framework of articulatory motor skill learning [11]. Producing the Spanish alveolar trill /r/ relies on the tongue tip muscles, in a relaxed state, undergoing multiple rapid vibrations induced by airflow. This motor pattern does not exist in Mandarin and constitutes a novel, fine motor skill requiring deliberate practice for beginners. Its acquisition follows the "cognitive-associative-autonomous" three-stage model. Most first-year students remain in the "cognitive stage", understanding the principle but unable to execute it stably. Their errors—substituting [l], a tap [ɾ], or a fricative [z̩]—often result from over-controlling the tongue muscles in an attempt to "manufacture" the trill, leading to articulatory rigidity. Therefore, the key to /r/ acquisition lies not in auditory discrimination, but in establishing correct neuromuscular control patterns for the tongue through extensive, feedback-driven practice. The ultimate goal is to transform this articulatory movement from conscious, effortful attempts into an automatized and fluent phonetic skill.

3) Integrated model: Intertwining of multiple interferences

The acquisition mechanisms identified in this study do not function in isolation but intertwine dynamically within real-world learning processes. Dialectal negative transfer mainly operates during the initial perception and category establishment phase for the /r, l/ contrast [5, 6]. Learners from n/l non-distinction areas are more prone to psychologically confuse the distinctiveness of these two phonemes, exacerbating acquisition difficulty. The affective filter (anxiety) primarily affects the online execution of pronunciation, especially in complex tasks requiring simultaneous semantic processing and articulatory monitoring [7, 8]. It impairs stable control over acquired and partially acquired articulatory gestures, including vocal fold vibration regulation for voiced plosives. Cross-linguistic phonetic category assimilation (e.g., the influence of English aspirated stop habits on Spanish /p/) constitutes a pervasive background layer of interference, continuously affecting the acquisition of the /p, b/ contrast.

In summary, phonetic errors among beginner Chinese learners of Spanish stem from perceptual assimilation, articulatory motor learning challenges, dialectal interference, and affective factors. For /p, b/, the pedagogical core should be auditory training and category restructuring. For /r, l/, the focus should shift to articulatory demonstration and muscular training. Any effective teaching intervention must be grounded in an accurate understanding of these deep-seated mechanisms.

V. Pedagogical Implications

Acquisition and Pedagogical Recommendations for First-Year Students.

Drawing on the above findings, this study puts forward targeted recommendations for foundational Spanish pronunciation instruction, with the aim of facilitating the effective transformation from "theoretical understanding" to "practical mastery".

A. Implementing Tiered Intervention: Differentiating Perception Training from Motor Training

Instruction should first differentiate the nature of acquisition for the two types of phonemes. For voicing contrasts like /p, b/, /t, d/, /k, g/, the pedagogical core is perceptual category restructuring. Intensive, focused auditory discrimination training using minimal pairs should be scheduled early in the course, immediately supplemented with Praat visual feedback [9, 10]. Instructors should guide students to focus on the key acoustic cue of Voice Onset Time (VOT) [1]. Comparing spectrographic discrepancies between voiceless sounds lacking voicing bars and voiced sounds with distinct voicing bars converts the abstract phonological concept of voicing into a tangible visual cue, helping learners overcome the constraints imposed by L1 perceptual categories.

For the /r/ sound, the teaching focus should shift to motor skill learning [3]. Its acquisition must be framed as a physiological process moving from "conscious control" to "unconscious automation". Instruction should move beyond theoretical explanation to provide concrete, actionable physiological guidance. This involves emphasizing tongue-tip relaxation, employing consonant clusters such as tr- and dr- as facilitators to induce alveolar vibration, and consolidating muscle memory via intensive repetitive nonsensical syllable drills (e.g., dra-dre-dri-dro-dru). Practice for /r/ should be regularized and broken into manageable chunks to prevent anxiety arising from short-term frustration.

B. Designing Targeted Practice: Integrating Context and Mitigating Psychological Barriers

Practice design must balance scientific rigor with situational appropriateness. To address the /r, l/ difficulty for learners from specific dialect areas, listening and speaking exercises incorporating minimal pairs like "pero" (but) and "pelo" (hair), "caro" (expensive) and "calo" (I measure) can be designed to strengthen phonemic distinction awareness through contrast. To reduce the negative impact of the affective filter, a safe practice environment should be fostered. Early-stage practice can adopt more individualized private modes such as self-recording and software-assisted shadowing. Classroom activities can begin with low-pressure forms such as group mutual assistance or role-play, gradually transitioning to public performance. Instructors should adopt a developmental assessment attitude towards errors, explicitly acknowledging progress to transform anxiety into positive learning motivation.

C. Fostering Metacognitive Monitoring: Guiding Self-Awareness and Reflection

Instructors should guide students to foster metacognitive awareness in Spanish pronunciation learning. Students should be encouraged to keep a "pronunciation practice log". This log should not only record practice content but also describe the physical sensations during articulation (e.g., "when producing /b/, I feel vibration touching my throat with a finger"; "when attempting /r/, does my tongue tip lightly touch the alveolar ridge?") and the psychological state (e.g., "felt relaxed/tense when trying /r/ today"). This self-monitoring enhances students' awareness of their own production process, shifting their role from passive imitator to active investigator. Combined with regular playback of self-recorded audio and self-examination of spectrograms, learners can evaluate their progress more objectively, establish an internal feedback loop, and realize self-correction and sustained phonetic improvement [9, 10].

VI. Research Conclusions

This study employed a mixed-methods approach to conduct a multidimensional investigation into the acquisition process of the /p, b/ and /r, l/ phonemes by first-year Spanish majors in a Chinese university. Based on acoustic data analysis, questionnaire surveys, and in-depth interviews, the following core conclusions are drawn.

First, the acquisition difficulties Chinese learners encounter with Spanish /p, b/and/r, l/phonemes are rooted in distinct cognitive-psychological mechanisms. The challenge with the /p, b/ voicing contrast primarily stems from cross-linguistic perceptual category assimilation [4]. Because the Mandarin stop system lacks a true voiced category, learners tend to assimilate the Spanish voiced plosive [b] into their native unaspirated voiceless [p] category. This leads to perceptual confusion, subsequently triggering "devoicing errors" in production. This problem is universal and is not significantly influenced by specific Chinese dialect backgrounds. In contrast, the challenge of the /r/ sound is fundamentally a problem of novel motor skill learning [3]. Most learners' difficulty lies not in auditory discrimination but in the inability to execute the fine motor action of a rapid, relaxed tongue-tip trill. Their errors (e.g., lateralization) are manifestations of motor control failure. The discrepancy between perception and production ($r = 0.18$, $p = 0.174$) provides strong empirical evidence for this acquisition mechanism.

Second, dialectal backgrounds and affective factors exert specific influences on the phoneme acquisition process. The negative transfer effect of dialect is selective, manifesting mainly in the /r, l/ contrast [5, 6]. Learners from n/l non-distinction dialect areas showed a significantly higher mean lateralization error rate for /r/ ($M = 42.7\%$) than learners from n/l distinction areas ($M = 18.9\%$). This indicates that the pre-existing /l/ allophone or the /n, l/ confusion habit in the native dialect phonology interferes with establishing /r/ as a distinct phoneme in the target language. Psychological anxiety primarily affects the online production quality of pronunciation [7, 8]. High-anxiety learners achieved significantly lower overall pronunciation accuracy in impromptu production tasks compared with their low-anxiety peers. This was especially evident in the decreased stability of vocal fold vibration control for voiced plosives. This confirms that the affective filter has a debilitating effect during stages where speech production is not yet fully automated.

Third, the visual acoustic feedback provided by Praat software demonstrates clear value in correcting specific errors [9, 10]. Preliminary teaching experiments indicate that comparing spectrographic differences between learners' own productions and model pronunciations (e.g., observing the presence/absence of voicing bars, VOT length) can effectively enhance learners' metacognitive awareness of the acoustic features of the /p, b/ contrast. This helps them establish correct perceptual targets, thereby guiding articulatory correction.However, for motor performance skills that rely heavily on articulatory coordination such as the Spanish /r/sound, the short-term effectiveness of visual feedback remains limited. Its mastery still depends on long-term, segmented motor training.

Through empirical methods, this study has illuminated the complex mechanisms affecting the acquisition of core pronunciation difficulties for beginning Chinese learners of Spanish. Future pronunciation teaching will become more effective if it adopts differentiated and targeted strategies according to the distinct acquisition patterns of voicing contrasts and rhotic–lateral contrasts, while fully taking into account learners' affective states and dialectal backgrounds. It will help learners build a solid, normative pronunciation foundation at the introductory stage, overcome fossilized errors, and enhance confidence and competence in oral communication.

Conflict of Interest

The author declares no conflict of interest.

References

[1] M. Zhang and X. Li, "A study on the application of visual feedback in second language phonetic acquisition—Taking Spanish voiceless/voiced plosives p/b as an example," *Foreign Language World*, no. 3, 2024.

[2] M. Sun, "The impact of cross-language interference on Spanish phonetic acquisition—A case study of learners with English/Chinese

Bilingual background," *Foreign Language Teaching and Research*, no. 6, 2022.
[3] J. E. Flege, "Second language speech learning: Theory, findings, and problems," 1995.
[4] M. L. Zampini, "The relationship between the perception and production of Spanish voicing contrasts by native English speakers," 1998.
[5] Y. Wang and Y. Chen, "The Impact of Dialect Negative Transfer on the Acquisition of Spanish r/l Sounds—A Case Study of Learners from Xiang and Wu Dialect Areas," *Foreign Language Teaching*, no. 5, 2023.
[6] J. Li, "A study on the acquisition errors of Spanish r/l sounds by Chinese learners in Southwest China—An empirical analysis based on Sichuan-Guizhou Dialect area," M.S. thesis, Sichuan International Studies University, 2022.
[7] E. K. Horwitz, M. B. Horwitz, and J. Cope, "Foreign language classroom anxiety," 1986.
[8] F. Liu and Y. Zhao, "An empirical test of the affective filter hypothesis in Spanish phonetic teaching," *Foreign Language and Literature*, no. 2, 2022.
[9] L. Zhou, "A study on the visual application of praat software in Spanish Phonetic teaching," *Research and Exploration in Laboratory*, no. 10, 2021.
[10] J. Wang, "A study on the effect of visual pronunciation correction on the acquisition of Spanish voiceless/voiced plosives—A teaching experiment based on praat and MOOC platform," M.S. thesis, Guangdong University of Foreign Studies, 2021.
[11] L. Colantoni, "Increasing the segmental inventory: The acquisition of Spanish rhotics by English-speaking learners," 2006.