Perception of L2 speech fluency: The case of Brazilian Portuguese as an additional language

Pollianna Milan
Universidade Federal do Paraná
pmilan@ufpr.br

Alison Roberto Gonçalves
Universidade Federal do Paraná
arg@ufpr.br

Ubiratã Kickhöfel Alves
Universidade Federal do Rio Grande do Sul
ukalves@gmail.com

Laryssa Alves
Universidade Federal do Paraná
laryssaalves@ufpr.br

Luciana Varela Guerino
Universidade Federal do Paraná
luciana.guerino@ufpr.br

Abstract

This study investigated the perception of fluency in Brazilian Portuguese (BP) as a second language, focusing on speech produced by 10 learners, five native speakers of Spanish and five native speakers of French. Raters were 83 native Brazilian Portuguese speakers who were divided into two groups: monolingual speakers and bilinguals who spoke English as a second language (advanced proficiency level). Speech samples were obtained through simulated oral proficiency tests and rated on a 5-point Likert scale. Results indicated that bilingual raters consistently assigned more positive fluency ratings compared to monolinguals, regardless of the learners' native language. Statistical modeling confirmed that knowledge of an additional language by the raters significantly influenced fluency judgments, whereas the speakers' native language did not. These results suggest that bilingual raters may better recognize fluency-maintaining strategies or exhibit greater empathy towards

L2 learners. These findings highlight fluency perception as an emergent construct shaped by both speaker production and listener experience. The study has implications for L2 oral assessment practices, emphasizing the need for clear fluency criteria and adequate rater training.

Keywords: fluency perception; L2 Brazilian Portuguese; native speakers of Spanish and French; native Brazilian raters

1. Introduction

A notable gap in second language (L2) speech research in Brazil concerns the development of fluency. While empirical studies have more extensively addressed the acquisition of segmental and suprasegmental aspects (such as stress and intonation), as well as evaluations of L2 speech intelligibility and comprehensibility in various additional languages, to the best of our knowledge, few if any studies (Soares, 2015; Vieira, 2024) have addressed issues related to fluency, both in the production and in the perception domains of a second language.

Fluency, in Applied Linguistics, typically refers to the fluidity or ease with which the second language is spoken (Thomson, 2015). In efforts to support learners in developing fluent L2 speech, applied linguists and language educators have started to examine the development of L2 fluency across varying conditions, which include instructed versus naturalistic learning environments (e.g., Freed et al., 2004), differing amounts of input exposure to the target language (e.g., Saito & Hanzawa, 2016b), and individual learner characteristics, such as L2 proficiency level (e.g., Saito & Hanzawa, 2016a). In contrast, fluency is often used in everyday language to denote overall high L2 proficiency, encompassing advanced grammar, vocabulary, pronunciation, reading, and writing skills (Thomson, 2015). From this general understanding, learners at beginner or intermediate proficiency levels would not be considered fluent speakers. Thomson (2015) concedes that fluency should be understood as a cognitive skill that enables speakers, even those with limited proficiency, to communicate effectively by producing speech smoothly and with minimal hesitation, despite limited linguistic knowledge.

Segalowitz (2010) proposes that fluency can be conceptualized and measured in three distinct ways. *Cognitive fluency* refers to the speaker's ability to efficiently manage underlying mental processes involved in speech production, such as working memory, planning, lexical retrieval, and the suppression of competing languages. *Utterance fluency* encompasses the measurable aspects of speech, including speech rate (e.g., syllables per second), type, number and duration of pauses, mean length of run (average number of syllables between pauses), and frequency of self-repairs. Finally, *perceived fluency* involves listeners' subjective evaluations of a speaker's fluency based on speech samples, typically influenced by features of utterance fluency.

Within the realm of perceived fluency, studies have shown that comprehensibility and accentedness can significantly influence listeners' judgments (Derwing et al., 2007; Pinget et al., 2014). Comprehensibility is defined by the ease or difficulty with which a listener understands L2 accented speech (Derwing et al., 2007), whereas accentedness refers to the perceived deviation of a learner's pronunciation from a given pattern. Although these features are not equivalent to fluency, they frequently shape how fluency is assessed and may negatively affect the speaker's ability to produce speech automatically (Thomson, 2015). In this regard, Derwing (2017) argues that improving L2 fluency can positively impact communicative success by increasing other speakers' willingness to engage an L2 learner in conversation, thereby facilitating opportunities for input and interaction that support further language development.

With regard to studies on comprehensibility and accentedness judgments, several studies (Kennedy & Trofimovich, 2008; Huensch & Nagle, 2021, 2023) have shown that raters' experience with an L2, as well as different proficiency levels in these L2, may foster different evaluations of these constructs. These results indicate that these constructs do not depend solely on speech. These are actually emergent phenomena (Albuquerque, 2019), whose evaluation depends not only on the speech sample, but also on the rater's language experience.

Building on this perspective, in this paper we discuss the role of the listener in fluency ratings. The present study aimed to examine

whether knowledge of an additional language influences fluency judgments of speech produced by second language learners in the native language of the evaluators. It also sought to determine whether the speaker's native language affects these fluency evaluations. In order to achieve this goal, we collected data from two groups: (i) monolingual raters and (ii) bilingual raters. The bilingual raters spoke English as an L2 only (advanced proficiency level), which does not correspond to the speakers' L1. This allowed us to verify whether knowledge of an L2 in general (not knowledge of a specific L2, similar to the speakers' L1) plays a role in L2 fluency assessments. The following section presents our review of the literature, followed by the study method and the discussion of the results.

2. Literature Review

The studies reviewed below were conducted with adult participants, second language learners (L2) in immersive and/or instructional contexts, such as study abroad programs, intensive language courses, or university classrooms, which are settings that parallel the environment of the present. These studies explored speech fluency as one of their main constructs, whether through temporal measures (such as pause length and frequency, speech rate and self-repairs) or through perceptual evaluations by native and non-native raters. Some also explored the relationship between utterance fluency and perceived fluency. As there are currently no published articles or book chapters focusing on the perceived fluency of L2 Brazilian Portuguese, this review includes research on other target languages, namely Dutch, English and Spanish. In addition, some of the included studies also addressed related constructs, such as accentendness and comprehensibility.

Derwing et al. (2004) investigated how inexperienced raters would assess the fluency of low-level L2 learners of English. These students were Mandarin speakers who were learning English as an L2 in an immersion environment. The raters were native English speakers with no experience with speech assessment tasks. The researchers instructed raters to pay attention to temporal fluency variables such as filled and unfilled pauses, false starts, and repetitions. Assessment was conducted with a scale ranging from extremely fluent to extremely disfluent. The tasks for data collection consisted of

narrating a story with pictures, followed by a task that consisted of a monologue in which participants needed to talk about the happiest moment in their lives and, in task 3, participate in an interaction with one of the researchers in which they answered questions about the moment described in task 2.

Results confirmed that fluency varied according to the task performed, as the fluency assessments for describing images were considerably low compared to the assessment of the monologue and the interaction, which, in turn, did not differ. Derwing et al. (2004) argued that both the monologue and interaction tasks employed a relatively easy theme, which would be explored in past tenses in either of the learner's languages. Furthermore, raters noticed that the participants were more relaxed during the interaction, as they perhaps did not feel that they were being tested. The study also assessed the learners' comprehensibility, that is, how easy it was to understand them. Results demonstrated that the perception of fluency and comprehensibility are strongly related, as disfluent speech can be considered disruptive for listeners, resulting in a lack of focus or attention.

Derwing et al. (2007) conducted a longitudinal study over a two-year period to assess the fluency of L2 learners of English in an immersion environment. The learners were native speakers of Mandarin, Russian and Ukrainian, who performed picture narration tasks to generate the data for the study. To conduct fluency assessments, native speakers of English were recruited. The study reported a slight improvement in the fluency of speakers of Slavic languages. No improvement in the fluency of Mandarin speakers was reported. When evaluating speech comprehensibility, the study showed that the most fluent speakers were the easiest for raters to understand, as in Derwing et al. (2004). Derwing et al. (2007) inferred that the small progress in fluency development was due to the absence of formal instruction for oral production and fluency in the course they were taking, as it focused on reading and grammar instruction more frequently. Furthermore, the authors argued that the opportunities participants had for interaction with other English speakers in their social context were limited, which resulted in little exposure to English outside the classroom.

Rossiter (2009) investigated fluency perception with adult learners of L2 English who narrated stories with the use of pictures in two different data collection sessions on a 10-week interval. The learners of English were in an immersion environment, while also receiving formal instruction in L2 English courses. The group of raters included an experienced group of native ESL teachers, a native novice group, and a group of non-native English speakers. In general, fluency assessments among groups demonstrated similar patterns, which included the perception of pauses and the perception of reduced speech rate. The non-native listener group assigned more rigorous ratings to the samples, possibly due to greater awareness of L2 oral production that resulted from explicit instruction and feedback from their language instructors. Filled pauses, unfilled pauses, repetitions, and speaking rate accounted for the majority of temporal fluency ratings assigned. Other non-temporal variables such as pronunciation accuracy, grammar, and vocabulary were also salient negative influences on the assessment provided by the study raters.

Pinget et al. (2014) explored how native Dutch raters perceived fluency and foreign accent in L2 Dutch. Learners of L2 Dutch were recruited through a project at the University of Amsterdam. Moreover, the researchers also included eight native speakers of Dutch as a control group to generate the stimuli. All 38 speakers (30 L2 Dutch learners and eight native speakers of Dutch) performed a series of different computer-administered speaking tasks to provide spontaneous speech in the form of conversational monologue. Then, untrained native speakers of Dutch rated fluency (speed, breakdown and repair). Results showed that temporal L2 fluency measures (related to speed, pausing, and repairs) explained 84% of the variance in fluency ratings. Measures of speed and breakdown fluency (pauses) were the strongest predictors, while repair fluency contributed less.

In addition, Pinget et al. (2014) had 20 native Dutch raters evaluate accentedness using the same speech tokens from the former experiment. Raters were instructed to base their judgments on the pronunciation of sounds, word stress and intonation patterns. Results demonstrated that accent ratings were significantly, but weakly predicted by phonemic error and pitch range, accounting for 23% of the variance. The relationship between perceived fluency and accent was weak, which indicates that listeners can distinguish the two

constructs when provided with clear rating criteria. Interestingly, speech tokens that were rated as less fluent also tended to be rated as more accented. The study concluded that fluency and accent should be treated as separate dimensions in speech assessment. While fluency judgments align closely with temporal features, accent judgments involve more complex and possibly qualitative assessment not captured by the study.

Saito and Hanzawa (2016a) conducted an exploratory study to investigate how six years of foreign language instruction in Japanese secondary schools impacted the oral abilities of learners of L2 English. The participants were 56 Japanese university freshmen who had studied English only through classroom instruction, with no experience abroad. To measure their speaking proficiency, participants completed a timed picture description task designed with images that pushed them to use words with problematic segmental and syllable structure features. Their performances were evaluated in terms of global accentedness by trained native English-speaking raters and assessed for segmental accuracy, word stress, intonation, and speech rate (speed of utterance delivery).

Results showed that six years of instruction led to intermediate levels of pronunciation and fluency, and some degree of perceived accentedness improvement, though large individual variation was observed. When compared to a group of experienced learners in Canada, the Japanese learners scored significantly lower in all domains. None of them showed high proficiency in terms of speech rate. The most influential factors for oral proficiency included additional language practice outside regular classes, receiving pronunciation training during high school oral communication classes, and using the L2 in oral communication with nonnative speakers. These factors significantly predicted better performance in global accentedness, segmentals, and word stress, though not in intonation or speech rate. Saito and Hanzawa (2016a) emphasize that maximizing learning through extracurricular language practice and formal instruction, including pronunciation training, is crucial for enhancing oral proficiency in limited-input settings, such as foreign language contexts.

Saito and Hanzawa (2016b) carried out a longitudinal study over one academic year to examine how Japanese university students

developed L2 oral proficiency in relation to different types and amounts of input in a foreign language context. The study included 40 first-year university students, all of whom were native speakers of Japanese and had no prior study-abroad experience. Their spontaneous speech in English was recorded at three time points in the semester, and was evaluated for pronunciation (segmentals, word stress, intonation), fluency (speech rate), vocabulary (appropriateness, richness), and grammar (accuracy, complexity) by trained raters. The study found that after one semester of instruction, the only significant group-level improvement was observed in speech rate, with no other consistent gains in pronunciation, vocabulary or grammar. However, individual differences in the amount of L2 input received were predictive of improvement in several areas of L2 speech, which was quantified by the number of hours spent in form-focused and content-based classes, as well as reported L2 use outside the classroom.

The study also distinguished between the effects of different types of input. Content-based instruction, defined as classes in which English was used to teach subject matter, was more consistently predictive of gains in prosody, vocabulary, and grammar. Form-focused instruction showed more limited associations, primarily with prosodic features. Saito and Hanzawa (2016b) concluded that, in foreign language classroom settings, fluency and lexicogrammar may develop earlier and more rapidly than prosody and segmentals. Moreover, the authors argued that participants who proactively increased their exposure through content-based instruction were more likely to demonstrate substantial L2 oral development, even under conditions of limited overall input.

Leonard and Shea (2017) conducted a longitudinal study to examine the development of L2 Spanish fluency, accuracy, and complexity in a study abroad (SA) context. Thirty-nine native English-speaking undergraduate students from the United States, Canada, and Australia participated in a three-month SA program in Buenos Aires. Participants completed a series of monologic speaking tasks in Spanish and English at the beginning and end of the program. Their speech was analyzed for measures of syntactic complexity, lexical complexity, lexical variety, accuracy, and fluency. In addition, participants completed language knowledge tests (grammar and

vocabulary) and timed processing tasks (picture-naming and sentence-picture verification) to assess processing speed and automatization.

The study found significant improvements in several fluency measures, including articulation rate, mean length of run, and reduction in the rate of mid-clause pauses. However, no changes were observed in the rate of end-of-clause or short pauses. Repair rate remained stable. The authors argued that prior to the SA program, participants were often able to formulate and articulate full clauses without mid-clause pauses in their L1, while pausing more frequently in the middle of clauses in the L2, which suggested difficulty formulating and articulating full clauses without mid-clause pauses. Post-SA results demonstrated a tendency for their pattern of mid-clause and end-of-clause pauses in the L2 to approximate the L1 pattern, signaling an important change from pre- to post-SA.

Leonard and Shea (2017) found that although short-term study abroad programs can lead to some noticeable improvements in L2 speaking, these gains do not occur evenly across all aspects of L2 proficiency. Interestingly, increases in fluency were linked more to learners' processing speed before departure than to their existing language knowledge, which points to the importance of automaticity over explicit knowledge in fluency development. On the other hand, students with stronger vocabulary and grammar before going abroad tended to make more progress in accuracy and complexity, indicating that language knowledge provides a foundation for more target-like and complex language use. The authors also found higher rates of repair for participants with smaller fluency scores pre-SA. They suggest that repair may develop in a nonlinear manner and changes may differ for participants at different levels of L2 proficiency.

Hanzawa's (2021) study investigated the development of L2 speech fluency among Japanese University students learning English as an L2, who were enrolled in foreign language classrooms over one academic year. The study focused on three dimensions of L2 fluency: speed (articulation rate), breakdown (pause frequency and length), and repair (self-repair frequency). Fifty first-year students completed personal narrative tasks at three time points, and their performances were analyzed using six fluency measures derived from annotated speech samples. Results indicated a significant decline in mid-clause

pause[1] length during both semesters and in end-of-clause pause length over an academic year. However, no statistically significant changes were observed in articulation rate or pause frequency, which aligns with previous findings that pause duration may improve before other fluency measures in foreign language contexts (Leonard & Shea, 2017).

A main finding was the correlation between overall L2 learning experience and changes in between-clause pause frequency and repair frequency, particularly during the second semester. Learners who engaged more extensively in productive activities outside the classroom (such as speaking and writing) reduced the number of pauses at the end of clauses and tended to produce more self-repairs, a pattern consistent with Leonard and Shea's (2017) research, which suggests increased monitoring activity as L2 proficiency develops. Hanzawa (2021) concluded that while foreign language classroom environments may foster fluency development, improvements occur gradually over time and are largely influenced by the quantity and quality of L2 learning experiences. The study emphasized that pause length may serve as an early indicator of fluency gains in contexts with limited foreign language input, and that productive practice outside the classroom plays a crucial role in fluency development, echoing previous research findings (Derwing et al. 2007; Saito & Hanzawa, 2016a, 2016b).

Kostromitina and Kang (2021) examined how one semester of L2 English immersion influenced L2 learners' pronunciation development across different proficiency levels. The study included 75 students (Arabic and Chinese L1) enrolled in an intensive English program in the U.S. Speech samples were collected at the beginning and end of the semester and analyzed for both segmental and suprasegmental features (fluency, stress, pitch). The findings showed no significant improvement in segmental pronunciation across all groups, with only small changes for intermediate learners. Most of the variance in segmental performance was explained by individual and L1-related factors, not immersion or proficiency. In contrast, some

[1] Hanzawa (2021) referred to these pauses as "within-clause" and "between-clause pauses", while Leonard and Shea (2017) referred to them as "mid-clause pauses" and "end-of-clause pauses", respectively. For the purpose of standardising terminology, we will adopt the terms used by Leonard and Shea (2017).

suprasegmental features showed significant development. Advanced learners significantly increased their speech rate (more syllables produced per second). As regards pause production, another significant change was a decreased number of filled pauses post-immersion in the speech samples produced by the beginner learners. However, while the beginner students used fewer filled pauses, both intermediate and advanced learners produced more filled pauses post-immersion. No significant changes were observed in silent pauses. The study concludes that short-term immersion alone does not lead to consistent improvement in pronunciation, especially at the segmental level. Improvements in suprasegmental features were more evident among higher-proficiency learners. This echoes previous results from Leonard and Shea's (2017) study, which suggests that learners with more advanced linguistic knowledge may exhibit more progress in fluency development. Kostromitina and Kang (2021) advocate for the need of explicit pronunciation instruction in L2 English and emphasize the need to consider individual learner factors, especially prior proficiency, when evaluating pronunciation development in immersion contexts.

The studies reviewed above offer varied insights into how L2 fluency develops across diverse learning contexts. A consistent finding across these studies (Derwing et al., 2004, 2007; Rossiter, 2009; Pinget et al., 2014) is that fluency is closely tied to temporal variables, such as pause length and frequency, repair frequency, and speech rate, which significantly shape listeners' perception of fluency. Notably, these studies also demonstrated that raters with different levels of experience associated disfluent speech with the occurrence of pauses and repairs. Furthermore, fluency and comprehensibility appear linked, as disfluent speech generally made listeners' comprehensibility rates harsher (Derwing et al., 2004, 2007), while accentedness and fluency emerged as perceptually distinct constructs (Pinget et al., 2014), even if disfluent speech was sometimes also judged as more accented.

As regards learning environments, the reviewed studies suggest that fluency gains are possible in both study abroad and foreign classroom contexts, but progress varies across different fluency dimensions, learner variables, and type of input received. For instance, Leonard and Shea (2017) observed gains in speech rate and

mid-clause pause reduction following a short-term study abroad program, with fluency development more closely tied to prior linguistic knowledge and processing ability. In contrast, accuracy and complexity gains were associated with pre-departure vocabulary and grammar knowledge, highlighting different developmental pathways for different proficiency levels. Similarly, Hanzawa (2021) found significant reduction in pause length, particularly among participants who engaged in productive use of the L2 outside the classroom, and suggested that pause length may serve as an early indicator of fluency gains in a foreign language context. Kostromitina and Kang (2021) also reported improvements in suprasegmental features (improved speech rate and reduction of filled pauses) during an immersion program, but findings were more prominent among participants with advanced L2 proficiency.

Other studies (Derwing et al., 2007; Saito & Hanzawa, 2016a, 2016b; Hanzawa, 2021) noted that the absence of explicit instruction or limited opportunities for naturalistic exposure and productive language use can constrain fluency development. Additionally, variability in gains was often linked to individual factors such as processing speed (Leonard & Shea, 2017), and linguistic knowledge and proficiency level (Leonard & Shea, 2017; Kostromitina & Kang, 2021), which overall suggest that while L2 fluency can develop over time, the trajectory of improvement is not uniform and may depend on variables such as proficiency level, type of instruction, and opportunities for interaction outside the classroom.

3. Method

The present study[2] investigated the perceived fluency of Brazilian Portuguese (BP) as an L2, focusing on speech produced by 10 learners, five native speakers of Spanish and five native speakers of French. The speech fluency of these learners was evaluated by 83 native speakers of Brazilian Portuguese, who listened to each audio sample once and rated their fluency using a 5-point Likert scale.

[2] This study was submitted to the Human Research Ethics Committee of the Federal University of Paraná (69008123.6.0000.0214) and approved under the number 6.169.712 issued on July 7, 2023.

3.1. Speakers

Speakers were 10 (four women and six men) learners of Portuguese as an additional language. Their ages ranged from 19 to 28 years old. At the time of the study, all participants were learning Portuguese in an immersion context in Brazil, while also attending a formal language course at the Center for Languages and Interculturality at the Federal University of Paraná (Celin-UFPR).

Among the participants, five were native speakers of Spanish from Cuba, Guatemala, Equatorial Guinea (2) and Venezuela. The other five were speakers of French, coming from Haiti, Republic of Congo (3) and France. Eight of the 10 informants were foreign students admitted to Brazil through the Undergraduate Exchange Student Program (called PEC-G), an initiative of the Ministry of Foreign Affairs of Brazil that offers free enrollment in undergraduate courses in the country. The two additional participants (from France and Venezuela) were students at Intermediate levels B1 and B2 at Celin-UFPR, in accordance with the Common European Framework of Reference for Languages (CEFR).

The PEC-G students arrived in Brazil in February 2022 and began their Portuguese studies at Celin-UFPR. All of them started studying Portuguese at Basic Level (A1), due to their limited knowledge of the target language. At the time of data collection for this study, these students had already reached the B1 level. The other two participants, who had already been exposed to Portuguese for over 10 years (through friends or media), were classified at B1 and B2 levels from a placement test.

Additionally, in 2023, two Brazilian women, aged 20 and 34, were recorded, and their speech samples were included in the perception test as catch-trials for exclusion criteria. Both participants are native speakers of Brazilian Portuguese and had resided for most of their lives in Curitiba, Paraná, the same city of residence as the foreign learners. All participants who judged the fluency of the 10 foreign speakers also had to judge the fluency of the two Brazilian speakers. Only those participants who judged these two audio recordings as ‘quite fluent’ or ‘completely fluent’ were retained in the study.

3.2. Speech samples

Speech samples were collected in 2022, in a classroom without acoustic treatment, using a Sony ICD-PX240 recorder. Participants were recorded during a simulation of the oral exam for the Certificate of Proficiency in Portuguese for Foreigners (Celpe-Bras), Brazil's official proficiency test for Portuguese. The simulation was conducted by Celin-UFPR Portuguese language teachers, who also work as evaluators for the official exam. The oral component of the Celpe-Bras simulates a spontaneous conversation, lasting 20 minutes and divided into two parts: the first five minutes involve a discussion of personal topics based on a questionnaire filled by the applicant upon registration; the subsequent 15 minutes delve into several everyday subjects, introduced through the reading of three "stimulus materials". The materials used in the simulation were randomly selected from the twenty used in the 2021 edition of the exam. Topics covered included health, food, mobility, human rights, education, sustainability, work career, and technology, among others.

Given that the recorded speech was part of a conversational interaction and was not continuous, the audio clipping criterion prioritized the longest uninterrupted speaking time for each participant, standardizing 65 seconds of continuous speech for each of the twelve speakers.

3.3. Fluency perception test

The fluency perception test was run on the Survey Monkey platform[3]. The test was split in two different sections: (1) pre-task and (2) task. The pre-task section contained the Informed Consent Form and a profile questionnaire for the raters. At this initial stage, participants were asked to provide information on their nationality, age, gender, where they had lived most of their lives, which additional languages they spoke and their proficiency in each, whether they had ever lived outside Brazil, how often they usually traveled outside Brazil, how often they used additional languages, whether they could play musical instruments and, finally, whether they had any hearing problems. Participants who reported hearing problems were excluded from the study.

[3] https://pt.surveymonkey.com/

For this study, only two groups of participants were considered: those who were monolingual listeners and those who were listeners with knowledge of an additional language (specifically English, with a proficiency level of C1 or above).

The second stage consisted of the fluency perception task. This task section consisted of a 5-point Likert scale that ranged from 1 (not fluent) to 5 (completely fluent), according to Figure 1. Different from other studies which used a 9-point Likert Scale (Taylor, 2018; Suzuki & Kormos, 2020), we opted for a 5-point scale to simplify the scores and avoid confusion by the judges who could be undecided which score to assign due to their similar descriptions.



Figure 1: 5-point Likert Scale: Score 1 - Not fluent; 2 - Not very fluent; 3 -More or less fluent; 4 - Quite fluent; 5 - Totally fluent.

Each evaluator listened to an audio of 65 seconds from each L2 Portuguese speaker and the two native participants and then used the Likert Scale to assign a score between 1-5 according to their fluency perception. The audios were presented randomly on the Survey Monkey Platform and, in order to keep the judges from overthinking and misleading their judgement, they could listen to the audios only once.

3.4. Raters

The study considered the data from 83 raters who had no hearing problem and answered the task appropriately. All of them rated the fluency of the two Brazilian participants consistently (as quite fluent or totally fluent). From the total of 83 listeners/judges, 39.76% (n = 33) did not know any additional languages and 60.24% (n = 50) knew English as their only additional language, with advanced proficiency in English (C1 on the European scale or above).

Among the participants, 60% (n= 50) were aged between 17 and 30 years old; 30% (n = 25) were aged between 31 and 50 years old; and

only 10% (n = 8 were aged over 50 years old. Of this total, the majority (60%) were female (n= 50), 36% were male (n = 30) and 4% (n=3) were non-binary or preferred not to answer.

3.5. Recruitment and data collection procedures

All participants were recruited through informal invitations from the researchers, including invitation of students during undergraduate classes, students from the Intercultural and Language Center (Celin-UFPR), and invitation through social media. Upon accepting the invitation to participate in the study, the researchers explained the steps of the evaluation and the judges then received a Survey Monkey link with all the tasks that comprised the data collection. Initially, they had to sign the consent form before proceeding to inform background information. In case they refused to sign the consent form, they were immediately sent to the end of the survey and were not able to rate the audio samples. After filling out the form with their profile information, they had access to the fluency perception test. The fluency perception test lasted less than 30 minutes.

4. Results and Discussion

The present study aimed to examine whether knowledge of an additional language influences fluency judgments of speech produced by foreign learners in the native language of the evaluators. It also sought to determine whether the speaker's native language affects these fluency evaluations. As described in the Method section, the speech samples of five French-speaking and five Spanish-speaking learners were rated on five-point Likert scales by two groups of native Portuguese listeners: (i) monolingual listeners and (ii) listeners with knowledge of an additional language (English). English was chosen as the additional language because it differs from the learners' native language, being a Germanic language, as opposed to the two Romance languages spoken by the learners. Thus, any observed effects in this group of evaluators may be attributed to their bilingual status rather than familiarity with the learners' native languages.

For the present analysis, in order to facilitate data visualization and streamline inferential statistical procedures, the five-point Likert scale used by evaluators was collapsed into three categories: "Negative" (comprising the two lowest points), "Neutral" (the midpoint), and

"Positive" (the two highest points). Figure 2 presents the results, showing fluency judgments made by the two evaluator groups as a function of the learners' native languages (Spanish or French).

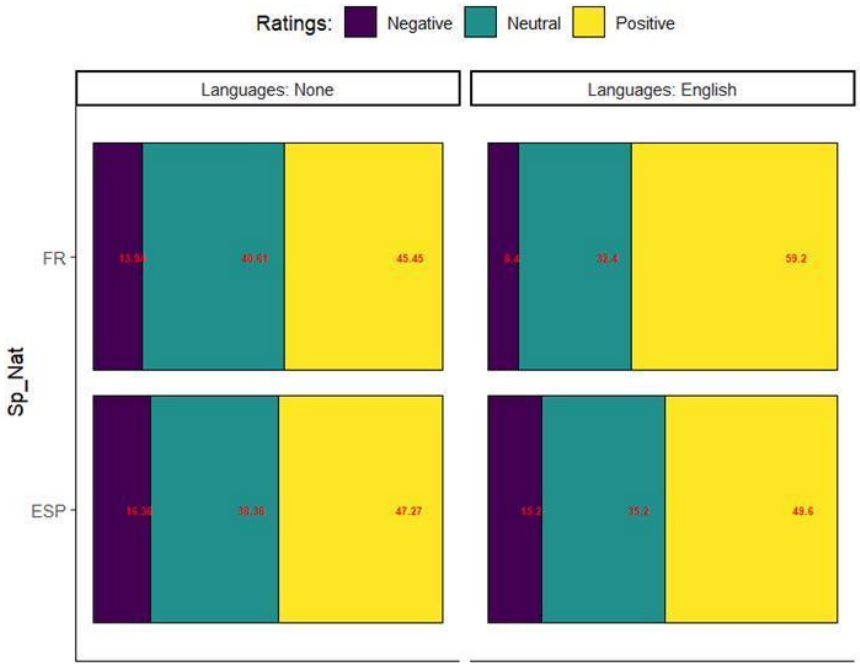


Figure 2: Descriptive data for the responses of the two evaluator groups as a function of the learners' native languages.

Note: [None] – no additional language. [English] – English as an additional language. [Sp_Nat] – Speaker's native language (ESP – Spanish; FR – French)

As for potential descriptive differences between the two evaluator groups, bilingual listeners, who were speakers of English as an additional language, tended to assign more positive ratings to both learner groups compared to monolingual listeners. This trend is particularly evident in the evaluations of French-speaking learners: while this group received 45.45% positive ratings (75 out of 165 responses) from monolingual listeners, it was rated positively in 59.2% of cases (148 out of 250 responses) by bilingual evaluators. Overall, this highlights that raters' experience (Rossiter, 2009), awareness of L2 production and familiarity with overall language

structure (Derwing et al., 2004; Pinget et al., 2014) can influence fluency judgments.

Regarding descriptive differences in fluency judgments between the two groups of speakers (French-speaking vs. Spanish-speaking), the graph suggests that monolingual listeners evaluated both groups similarly, assigning fewer than 50% positive ratings, 45.45% (75 out of 165) for the French-speaking group and 47.27% (78 out of 165) for the Spanish-speaking group. In contrast, among bilingual listeners, the proportion of positive evaluations was higher for the French-speaking learners (59.2% – 148 out of 250) than for the Spanish-speaking learners (49.6% – 124 out of 250), with both percentages outnumbering those given by monolingual listeners.

To determine whether the observed descriptive differences were statistically significant, we fitted a mixed-effects ordinal regression model, including a random intercept for Speaker and specifying an interaction between "Speaker's Native Language" (French, Spanish) and "Rater's Additional Languages" (None, English). The model revealed no significant interaction between these variables. In other words, from a statistical standpoint, the evaluations did not significantly vary as a function of the speakers' native language, despite the descriptive trend observed among bilingual raters.

Given the absence of significant interactions, we fitted an additive model without the interaction term. A model comparison via ANOVA revealed no significant improvement in model fit ($\chi^2(1) = 3.28$, $p = .070$). Therefore, we opted for the more parsimonious additive model. The results of the final mixed-effects ordinal regression model are presented in Table 1.

Predictors	*Log-Odds*	*CI*	*p*
Negative|Neutral	-1.92	-2.80 – -1.05	**<0.001**
Neutral|Positive	0.20	-0.66 – 1.06	0.650

Languages [English]	0.32	0.05 – 0.60	**0.023**
Sp Nat [FR]	0.15	-1.04 – 1.34	0.805
Random Effects			
σ^2	3.29		
$\tau_{00\ Speaker}$	0.87		
ICC	0.21		
$N_{\ Speaker}$	10		
Observations	830		
Marginal R^2 / Conditional R^2	0.007 / 0.215		

Table 1: Results of the Mixed-Effects Ordinal Regression Model.

The threshold separating the "Negative" category from the remaining response options was statistically significant (log-odds = –1.92, 95% CI [–2.80, –1.05], $p < .001$), indicating a clear distinction at this level of the ordinal scale. The same was not observed for the threshold between "Positive" and the remaining categories.

The ordinal model confirms that the raters' language background significantly influenced their fluency ratings. Bilingual raters were significantly more likely to assign positive evaluations (log-odds = 0.32, 95% CI [0.05, 0.60], $p = .023$). In contrast, no significant effect was found for the speaker's native language, as the slight increase in positive ratings for French-speaking learners was not statistically significant (log-odds = 0.15, 95% CI [–1.04, 1.34], $p = .805$). Notably, variability across speakers accounted for a substantial portion of the overall response variance, as reflected in the marginal and conditional R^2 values (R^2 marginal = 0.007; R^2 conditional = 0.215).

In sum, knowledge of an additional language appears to positively influence fluency evaluations of speech produced by foreign learners in the raters' native language. These evaluations are not affected by the speakers' native language, at least when it differs from the additional language known by the evaluators. Contrary to previous findings (Rossiter, 2009), the bilingual raters in the present study

assigned more favorable fluency ratings to L2 speakers, potentially reflecting a heightened awareness of the challenges learners encounter during the acquisition of an additional language. Interestingly, more lenient ratings may reflect heightened tolerance for disfluencies of L2 speech which is still in development, aligning with the relationship between fluency and comprehensibility highlighted in previous research (Derwing et al., 2004, 2007).

5. Conclusion

This study demonstrated that knowledge of an additional language positively influences fluency judgments of speech produced in the listener's native language by foreign learners. Given that the raters' additional language was English, a language which was unrelated to the speakers' native languages, these effects are unlikely to stem from mere linguistic similarity.

The findings suggest that the construct of "perceptual fluency in an additional language" does not depend solely on the speaker's production, as it emerges from speaker-listener interaction. These results are in line with previous findings related to other constructs in L2 speech, such as intelligibility and comprehensibility (Kennedy & Trofimovich, 2008; Albuquerque & Alves, 2017, 2020; Albuquerque, 2019; Huensch & Nagle, 2021, 2023; Trein et al., 2022, 2024). In the present study, raters with experience in an additional language were more lenient in their evaluations, assigning higher fluency scores. This may be due to their implicit ability to detect strategies used by speakers to maintain fluency, which raters then interpret as positive indicators of fluent speech. These raters may also be more empathetic toward the challenges faced by L2 learners and adjust their judgments accordingly.

One limitation of this study concerns the imbalance in the number of participants across rater groups. This reflects the inherent difficulty in recruiting self-declared monolingual participants. Future research should seek to increase the sample size for this group.

Despite this limitation, the findings have important pedagogical implications, particularly because "fluency" is a commonly used criterion in L2 oral assessments. The results underscore the need for institutions to establish clear guidelines on what characterizes fluency, especially in the case of L2 learners, who often differ from

native speakers in terms of utterance fluency markers such as speech rate, pause frequency, and repair rate. Moreover, specific fluency measures (speech rate, pause length and frequency, repairs) develop differently depending on context, input, and learner characteristics (Leonard & Shea, 2017; Hanzawa, 2021; Kostromitina & Kang, 2021). This variability highlights the importance of training raters to recognize the strategies L2 learners use to maintain fluency and to consider the learner's background when assigning ratings. Since bilingual raters appear to be more attuned to these challenges, it is essential that all raters and speech evaluators, regardless of language background, receive adequate training to understand the difficulties L2 learners face in maintaining fluency and to accurately recognize the strategies learners use to produce fluid speech. By doing so, assessments can more fairly reflect a learner's communicative ability rather than penalizing natural disfluencies that arise from the L2 acquisition process.

Finally, the present findings open several avenues for further research. One relevant direction involves investigating whether the rater's level of proficiency in their additional language affects their fluency judgments. In the current study, this variable was controlled by including only upper-intermediate and advanced L2 learners. Given that listener proficiency has been shown to influence comprehensibility and accent ratings (Trein et al., 2022, 2024), future studies should explore its role in fluency evaluations.

Another promising line of inquiry concerns the potential impact of the raters' familiarity with the speakers' native language. Such familiarity might increase fluency ratings, not only due to general bilingual experience as discussed here, but also due to increased tolerance or recognition of features associated with the speaker's foreign accent. In order to address this question, we are currently collecting data from listeners who speak Spanish or French as an additional language. However, we must highlight the methodological challenges involved in this regard, since most Brazilian learners of Spanish or French are also proficient in other languages, such as English.

In summary, this study contributes to the understanding of perceptual fluency as an emergent construct shaped by both speaker and listener. We hope the present findings help advance the discussion

on perceptual fluency as a shared construct and stimulate further research on a topic that remains underexplored in the South American academic context.

References

Albuquerque, J. I. A. de. (2019). *Caminhos dinâmicos em inteligibilidade e compreensibilidade de línguas adicionais: Um estudo longitudinal com dados de fala de haitianos aprendizes de Português Brasileiro* [Tese de doutorado, Universidade Federal do Rio Grande do Sul]. https://lume.ufrgs.br/handle/10183/207592.

Albuquerque, J. I. A., & Alves, U. K. (2017). Compreensibilidade em L2: Uma discussão sobre o efeito da experiência do ouvinte e do tipo de meio em excertos do Português Brasileiro produzidos por um falante haitiano. *Revista X*, *12*(2), 43–64. https://doi.org/10.5380/rvx.v12i2.51246.

Albuquerque, J. I. A., & Alves, U. K. (2020). Os construtos de 'inteligibilidade' e 'compreensibilidade' em dados do português brasileiro como língua adicional: Um olhar via Sistemas Dinâmicos Complexos. *Signótica*, *32*, e58214. https://doi.org/10.5216/sig.v32.58214.

Derwing, T. M. (2017). L2 fluency development. In: S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 246–259). Routledge.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380. https://doi.org/10.1093/applin/amm041.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(2), 275–301. https://doi.org/10.1017/S0272263104262064.

Hanzawa, K. (2021). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *28*(3), 816–838. https://doi.org/10.1177/13621688211008693.

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*(3), 626–668. https://doi.org/10.1111/lang.12451.

Huensch, A., & Nagle, C. (2023). Revisiting the moderating effect of speaker proficiency on the relationships among intelligibility, comprehensibility, and accentedness in L2 Spanish. *Studies in Second Language Acquisition*, *45*(2), 571–585. https://doi.org/10.1017/S0272263122000213.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, *64*(3), 459–483. https://doi.org/10.3138/cmlr.64.3.459.

Kostromitina, M., & Kang, O. (2021). The effects of ESL immersion and proficiency on learners' pronunciation development. *Frontiers in Communication*, *6*, 636122. https://doi.org/10.3389/fcomm.2021.636122.

Leonard, K. R., & Shea, C. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, *101*(1), 179–193. https://doi.org/10.1111/modl.12382.

Pinget, A.-F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365. https://doi.org/10.1177/0265532214526177.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review*, *65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395.

Saito, K., & Hanzawa, K. (2016a). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, *37*(4), 813–840. https://doi.org/10.1017/S0142716415000259.

Saito, K., & Hanzawa, K. (2016b). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *22*(4), 398–417. https://doi.org/10.1177/1362168816679.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge, New York.

Soares, L. A. A. (2015). O fenômeno pausa como pista para avaliação de fluência em Português do Brasil como segunda língua. *Brazilian English Language Teaching Journal*, *5*(2), 18-30. https://revistaseletronicas.pucrs.br/belt/article/view/19516.

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167. https://doi.org/10.1017/S0272263119000421.

Taylor, S. (2018). Correlating speech fluency measures with speech fluency perceptions. *Language Education and Research Center Journal*, *13*, 5–21.

Thomson, R. I. (2015). Fluency. In: M. Reed & J. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 209–226). Wiley. https://doi.org/10.1002/9781118346952.ch12.

Trein, A. D., Alves, U. K., & Luchini, P. L. (2022). Comprehensibility ratings in non-native speaker-listener binomials: Implications for NNL teaching. In: A. M. Souza, R. Garcia, & T. C. Santos (Orgs.), *Perspectivas Para o Ensino de Línguas* (Vol. 6, pp. 22–32). Edufac.

Trein, A. D., Alves, U. K., & Luchini, P. L. (2024). Accentedness perception ratings in non-native speaker-listener binomials: Analyzing the influence of proficiency. In: P. L. Luchini, U. K. Alves, & V. Innocentini (Orgs.), *Otras Cuestiones del Lenguaje: Traducción y Enseñanza de Lenguas Extranjeras* (pp. 369–386). Universidade Nacional de Mar del Plata.

Vieira, F. G. M. (2024). *Complexidade, Acurácia e Fluência no desenvolvimento do Português Brasileiro como Língua Adicional: Uma análise dinâmico-complexa de um aprendiz hondurenho* [Trabalho de conclusão de curso de graduação, Universidade Federal do Rio Grande do Sul]. https://lume.ufrgs.br/handle/10183/282457.