Objectives: L2 acquisition of stress patterns is typically challenging for adults and often overlooked in structured classroom learning [1]. We asked three research questions concerning the acquisition of antepenultimate and penultimate stress patterns by adult L2 Italian learners. 1) To what degree do beginner L2 Italian speakers produce duration, amplitude, and fundamental frequency cues associated with different stress types? 2) How do L1 Italian listeners rate L2 stress patterns across four fluency dimensions: overall, duration, loudness, and pitch? 3) How do these acoustic cues affect the fluency ratings?

Method: Ten first and second semester L1 English-L2 Italian adult classroom learners (age 18-25) and five adult L1 Italian speakers from Italy read aloud 32 pairs of common, frequency controlled, trisyllabic words via Gorilla [2]. The words, taken from [3], were segmentally identical in their first two syllables but differed in their stress location, e.g., TOnaca ('habit'; capital letters indicate stress, though participants only saw lowercase letters) and toNAle ('tonal'). Files were cleaned, trimmed, and normalized for amplitude. 900 of the 960 recordings' vowels (1,800 vowels in total) were manually tagged in Praat [4]. Fundamental frequency (Bark transformed [5]), amplitude (dB), and duration (ms) were extracted. Next, these 900 files were played via Gorilla [2] for 50 L1 Italian listeners recruited through Prolific (mean age = 36; mean Italian proficiency = 57/60 using [5]). Participants took part in a headphone screening [7], and then heard 100 different L1 and L2 utterances. After each utterance, participants rated overall fluency, and duration, loudness, and pitch fluency using visual analog scales with the endpoints ranging from 0 to 1000.

Results: Outliers beyond 3 median absolute deviations [8] were removed (~3% of acoustic data). The L1 speakers demonstrated statistically significant duration and amplitude differences for both stress types and a pitch difference for penultimate stress in line with [3,9] (Table 1). The L2 speakers demonstrated a signification duration difference for both stress types. This difference was in the correct direction for penultimate stress but in the opposite direction for antepenultimate stress. Figure 1 and Table 2 presents the results from the four fluency ratings. In all dimensions, the L1 speakers were rated as 'more fluent' (higher rating) than the L2 speakers (lower rating). For the L2 speakers, slightly more variation was observed in the penultimate ratings than the antepenultimate ratings. Two mixed-effects linear regression models (one for each stress type) were built to predict the overall fluency rating for the L2 data. The three acoustic variables were scaled and used as predictors along with two-way interactions with vowel order (dummy coded with the first vowel as the reference level). For the antepenultimate model, a longer first vowel duration led to significantly higher fluency ratings ($\beta = 53.9$, $p < .001$) whereas a longer second vowel duration led to significantly lower fluency ratings ($\beta = -50.9$, $p < .001$). For the penultimate model, a longer duration on the second vowel led to significantly higher fluency ratings ($\beta = 32.1$, $p = .02$) while a higher pitch and a higher amplitude on the second vowel led to significantly lower fluency ratings (pitch: $\beta = -24.5$, $p = .02$; amplitude: $\beta = -49.6$, $p < .001$). Figure 2 plots the effect of duration on overall fluency. All other predictors in the models were null at a .05 alpha-level.

Conclusion: Unlike L1 speakers, L2 learners failed to produce consistently reliable antepenultimate and penultimate stress cues in terms of duration, fundamental frequency, and amplitude. The lack of salient acoustic cues resulted in the L2 utterances being deemed "less fluent" across all four ratings, particularly when compared to the L1 utterances. Regression analyses confirmed these acoustic cues drive fluency ratings, especially duration cues, which we found to be a reliable predictor in both our antepenultimate and penultimate model. We discuss potential L2 Italian pedagogy strategies, including ways to emphasize duration and pitch in pronunciation practice. We also examine learner variation by highlighting individual differences in L2 acquisition and hypothesize why some learners are "more fluent" than others.

References

[1] Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. In *Language Experience in Second Language Speech Learning: in honor of James Emil Flege. Amsterdam*. (pp. 281-298).
[2] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*, 388-407.
[3] Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*(1), 177-193.
[4] Boersma, P. & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.07, retrieved 17 March 2024 from http://www.praat.org/
[5] Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97-100.
[6] Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, *42*(2), 292-314.
[7] Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*, 1551-1562.
[8] Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766.
[9] Albano Leoni, F., Maturi, P. (1998). *Manuale di fonetica.* Roma, Italy: Carocci.
[10] Krämer, M. (2009). *The phonology of Italian.* New York: Oxford University Press

Table 1. Mean acoustic measures and t-test comparisons for the first and second vowel of the words with each stress pattern by speaker status. P-values adjusted for 12 comparisons.

	First vowel	Second vowel	p-value
Antepenultimate stress			
Duration (ms)	**L1**: 111 **L2**: 82	**L1**: 71 **L2**: 132	< .001 < .001
Pitch (Bark)	**L1**: 1.47 **L2**: 1.82	**L1**: 1.39 **L2**: 1.78	.15 .39
Amplitude (dB)	**L1**: 9.91 **L2**: 9.07	**L1**: 7.06 **L2**: 8.29	< .001 .14
Penultimate stress			
Duration (ms)	**L1**: 77 **L2**: 81	**L1**: 152 **L2**: 136	< .001 < .001
Pitch (Bark)	**L1**: 1.53 **L2**: 1.79	**L1**: 1.38 **L2**: 1.79	< .01 .92
Amplitude (dB)	**L1**: 10.05 **L2**: 9.06	**L1**: 7.26 **L2**: 8.70	< .001 .48

Table 2. Mean fluency ratings from 0-1000 and t-test comparisons by speaker status.

	L1	**L2**	p-value
Overall Fluency	908	264	< .001
Pitch Fluency	851	287	< .001
Loudness Fluency	842	382	< .001
Duration Fluency	823	358	< .001

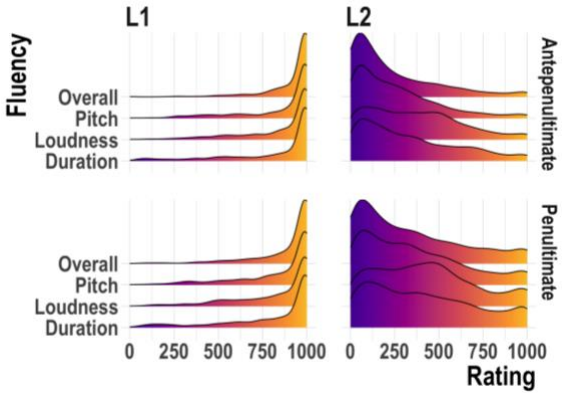

Figure 1. Density plots of fluency ratings by L1 (left) and L2 (right) speakers and stress pattern.

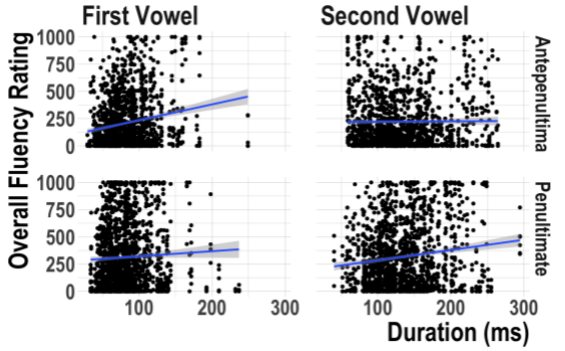

Figure 2. Regression lines (in blue) showing effects of L2 vowel duration on fluency ratings by vowel order and stress type.