Applied Linguistics, **2025, 00, 1–18**

https://doi.org/10.1093/applin/amaf075

Article

Acoustic analysis and perception ratings of first and second language speakers' Italian lexical stress

Seth Wiener[1,*], Adam A. Bramlett[1], Bianca Brown[1] and Jocelyn Dueck[2]

[1]Department of Languages, Cultures & Applied Linguistics, Carnegie Mellon University, 341 Posner Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States
[2]School of Music, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States

[*]Corresponding author. Department of Languages, Cultures & Applied Linguistics, Carnegie Mellon University, 341 Posner Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States. E-mail: sethw1@cmu.edu

This study examines the acquisition of Italian lexical stress by adult L2 learners. L1 Italian speakers and beginner L2 Italian speakers were recorded reading aloud trisyllabic Italian words, e.g. *COdice* with antepenultimate syllable stress ("code"), *moMENto* with penultimate syllable stress ("moment"). We analyzed four acoustic-phonetic cues: duration, fundamental frequency (pitch is the perceptual correlate), amplitude, and spectral tilt (a measure of energy change over frequencies). We corroborated previous findings: L1 speakers used all four cues to differentiate between antepenultimate (strong-weak-weak) and penultimate (weak-strong-weak) stressed words. We found evidence of L2 speakers producing inconsistent patterns for all four cues. We then played these L1 and L2 recordings for L1 Italian speakers ($N = 50$) and asked them to rate the utterances using a visual analog scale (VAS). As expected, the L1 speech was rated higher (more fluent stress) than the L2 speech (less fluent stress). We modeled how the acoustic cues predicted VAS responses. Our findings highlight the roles of duration and pitch for L2 learners. We conclude with implications for learners and teachers of Italian.

Introduction and aims

This paper serves as an attempt to link the acoustics of first (L1) and second (L2) language speech to L1 listener judgments in order to better inform L2 pedagogy. The speech we examine is Italian trisyllabic words. Spoken Italian has lexical stress, that is, the placement of stress varies depending on the word. This includes oxytone words ("parole tronche") with stress on the final syllable as *serviTU* ("servitude" note that capitalization indicates stressed syllable), paroxytone words ("parole piane") with stress on the second to last syllable as in *faMIGlia* ("family"), and proparoxytone words ("parole sdrucciole") with stress on the third to last syllable as in *TAvolo* ("table") (D'Imperio and Rosenthall 1999). Despite the relatively straightforward phonological rule—if the penultimate syllable ends in a consonant, it is stressed (Krämer 2009)—learning to produce Italian stress

patterns is difficult: five-year-old L1 Italian children are typically unable to produce adult-like L1 Italian stress patterns (Sulpizio and Colombo 2013; Arciuli and Colombo 2016; Bellocchi et al. 2016) and adult L2 learners similarly struggle in their acquisition (Primativo et al. 2013; Nicora et al. 2018; Spinelli et al. 2021).

Why is stress in Italian so hard to acquire? We begin this article with an acoustic analysis of Italian trisyllabic lexical stress. What makes a weak and strong syllable? We examine four phonetic correlates of stress—duration, fundamental frequency, amplitude, and spectral tilt—and demonstrate how L1 Italian and L1 English–L2 Italian speakers vary in their productions of these cues. Next, we explore how these phonetic cues affect crowd-sourced L1 Italian listener judgments about the perceived speech. We find considerable differences in ratings of the L1 and L2 speech and model how the L1 and L2 acoustics predict listener ratings. To conclude, we link these findings to best practices for Italian learners and instructors.

The acoustic correlates of Italian lexical stress

There are many acoustic-phonetic cues involved in the realization of stress; that is, there is no unitary cue (Lehiste and Peterson 1959; Lieberman 1960). Children and adults must learn which cues to use and over which syllables. In general, stressed syllables tend to be perceived as being louder than unstressed syllables as a result of stressed syllables' increased amplitude or intensity (McClean and Tiffany 1973). Stressed syllables also tend to have a higher fundamental frequency (F0) or perceived pitch and longer duration than unstressed syllables (Morton and Jassem 1965).

Yet, stress varies across languages. Cues must be learned and applied; stress realization and its perceptual correlates depend on the phonological system of a language. For example, stress in Arabic is cued by duration and F0 (de Jong and Zawaydeh 1999). Stress in Spanish is not cued by any one cue alone (Llisterri et al. 2003). Stress in Thai—a tonal language—is cued only by duration (Potisuk et al. 1996). For Italian, there does not seem to be one single cue that sufficiently defines stress placement (D'Imperio and Rosenthal 1999; Arciuli and Colombo 2016).

This paper is concerned with the general Italian stress patterns involving strong-weak-weak as in *AMbito* ("setting") or antepenultimate stress and weak-strong-weak as in *amBIto* ("coveted") or penultimate stress. Italian stressed vowels typically differ from unstressed vowels in terms of their F0, duration, amplitude (intensity), and spectral tilt, which is a measure of the change of energy over frequencies (Bertinetto 1981; Sulpizio and McQueen 2012; Albano Leoni and Maturi 2018; Bramlett and Wiener 2025a). The third to last syllable in Italian antepenultimate stressed words has longer durations, higher average pitch, higher average amplitude, and higher spectral tilt than the second to last syllable (Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). The third to last syllable in penultimate stressed words has shorter durations, higher average pitch, and higher average amplitude than the second to last syllables, and a relatively similar spectral tilt across the two syllables (Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). The first aim of our study is to collect L1 speech data using a web-based approach. This will allow us to confirm these expected acoustic-phonetic differences. To encourage future close replications and extensions, we provide all our materials and scripts through open science practices.

We know that adult L2 learners struggle to correctly assign stress in trisyllabic Italian words when reading aloud, and knowing more L2 words can lead to more accurate stress production (Primativo et al. 2013; Nicora et al. 2018; Spinelli et al. 2021). We also know that typological differences between a learner's L1 and L2 contribute to learners' errors (e.g. Missaglia 1999; De Meo et al. 2012). What we still do not know is where the breakdown occurs in the speech signal. Primativo et al.'s (2013) and Spinelli et al.'s (2021) studies are limited by the fact that L1 raters decided whether the L2 productions were correct or incorrect, that is, upon hearing the speech listeners made a binary decision. It is unclear which acoustic cues the L1 raters used to make those accuracy decisions. Our second aim is to therefore collect and report L1 English–L2 Italian acoustics data using our open methods. This will provide a fuller understanding of what makes L2 Italian stress difficult to acquire and what future cue(s) L2 pedagogy should target.

The perception of L1 and L2 Italian stress

We know that L1 Italian listeners are sensitive to stress (Gemelli 1950; Bertinetto 1980; Tagliapietra and Tabossi 2005). Duration is believed to be the most reliable and informative cue for stress (Ferrero 1972; Fava and Magno-Caldognetto 1976; D'Imperio and Rosenthal 1999; Alfano 2006; Alfano et al. 2007, 2009; Eriksson et al. 2016). Many studies, however, find other predictors such as amplitude (Albano Leoni and Maturi 2018), F0/pitch (Caccia et al. 2019), and spectral tilt (Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). These mixed findings—and recent exploratory two-way interactions between cues reported in Bramlett and Wiener (2025a)—suggest listeners may be using a combination of acoustic cues in tandem to perceive stress. That is, Italian may be like Spanish in that stress is not conveyed by one cue alone (Llisterri et al. 2003).

We also know that listeners are biased in a number of ways toward L2 speech. This finding has real-world implications and is itself a rich area within second language acquisition research (e.g. Reid et al. 2019; Cheng et al. 2021; Brown et al. 2023). Findings across languages indicate that listeners rate L1 speech as more fluent and less accented than L2 speech and that L2 or "non-native" speech is "flawed" in some way when compared to L1 or perceived "native" speech (e.g. Isaacs and Thomson 2013; Lindemann and Subtirelu 2013).

To understand perceptions of L2 Italian speech, Pellegrino (2012) had L1 Italian listeners rate L1 Chinese–L2 Italian speech and found that L1 listeners rated the L2 speech with longer vowels as having a stronger accent. Pellegrino (2012), however, did not carry out inferential statistics to test this claim. In another study, Pellegrino et al. (2021) played L1 Italian listeners speech from L1 Zurich German–L2 Italian speakers and L1 Italian speakers. Participants were asked which of the two utterances was more "native-like". Pellegrino et al. (2021) demonstrate subtle but statistically significant effects of duration and amplitude (envelope) on the listener behavior. Combei (2023) reports on L1 ratings of L2 Italian speakers from varying L1 backgrounds. L1 Italian listeners attributed L1 English–L2 Italian speakers' accent to stress misplacement. Yet, Combei (2023) did not directly test how the acoustics of the speech affected this behavior.

Here we contribute to better understanding the perception of L2 Italian stress by L1 listeners, which has not been isolated as a primary variable of interest. Specifically, there does not appear to be data on the L1 recognition of spoken trisyllabic words produced by L2 Italian speakers. In this study, we explore how variability in the L2 acoustic cues affects L1 listeners' ratings. It is reasonable to predict that L1 Italian listeners will rate L2 speech lower or less than L1 speech regardless of the construct we use. We are primarily interested in how the acoustic-phonetic correlates of stress are predictive of what we see as a gradient spectrum from positive to negative rather than binary "correct" and "incorrect". This left us with several options. Munro, Derwing, and colleagues (1995, 2006) have argued for at least four independent dimensions or constructs of speech utterances: intelligibility (actual understanding), comprehensibility (effort required for understanding), accentedness (extent to which speech deviates from expected norm), and fluency (temporal aspects of oral production). Because utterances that are rated as having strong accents can still be transcribed perfectly, the field has largely shifted away from accentedness research (Levis 2005; Nagle and Huensch 2020) and toward the other three constructs. We use the term stress fluency to emphasize temporal regularity and coordination of stress cues at the word level, rather than the broader construct of fluency in connected speech typically considered in the field. This framing highlights that our ratings target stress realization in isolated words—something participants could evaluate with confidence—and aligns with Derwing et al.'s (2009) emphasis on temporal aspects of speech.

Our third aim, therefore, is to solicit L1 responses to stress fluency in L1 and L2 speech. To do this, we collect continuous stress fluency scores using a visual analog scale (VAS). VAS allows for more fine-grained responses rather than binary fluent/not fluent categorization (cf. Pellegrino et al. 2021).

Linking acoustics to perceptual judgments

Second language acquisition research has long relied on perceptual evaluations such as intelligibility, comprehensibility, and fluency to assess learner speech (Derwing et al. 2009). These measures are valuable and offer insights into communicative success beyond native-likeness alone. Yet, relatively little work has directly linked these ratings to the specific acoustic features of learner speech. In order to understand variation in outcomes, we must move beyond broad categories of "native" and "non-native" and examine the structure of the speech signal itself (Xie et al. 2023; Bent et al. 2024). Some L2 speakers are rated as highly fluent or comprehensible, while others are not, even with similar learning experiences (Nagle and Huensch 2020). To explain this variability, researchers must examine what in the acoustic signal drives those perceptions. By connecting fine-grained acoustic measurements to listener ratings, we can better understand the dimensions of speech that matter for L2 evaluation, refine our models of L2 speech development, and ultimately inform pedagogical approaches that target the features most relevant to successful communication.

Italian lexical stress offers an ideal test case for linking acoustic structure to perceptual judgments. Unlike languages where stress distinctions are dominated by a single cue, Italian stress is conveyed through a combination of duration, pitch, amplitude, and spectral tilt (Bramlett and Wiener 2025a). This multi-cue structure presents a valuable opportunity: it is complex enough to require integration across cues, yet systematic enough to allow clear measurement of cue patterns. Critically, while duration is often a reliable indicator—especially for distinguishing antepenultimate and penultimate stress (Alfano et al. 2007, 2009)—successful production of duration patterns alone may not guarantee high stress fluency ratings. A learner could lengthen the appropriate syllable but fail to modulate pitch or spectral tilt in ways listeners expect, leading to lower perceived stress fluency. This makes Italian stress an ideal domain for capturing stimulus-specific variation: two learners might both acquire duration but struggle to acquire other cues, resulting in different ratings. By modeling how multiple acoustic features combine to influence gradient listener judgments, we aim to move beyond binary notions of correctness and better understand how L2 speech is perceived by L1 listeners.

Linking L1 listener behavior to pedagogy

The final two aims of our study tie together the data we collected in our first three aims. Our fourth aim concerns modeling the VAS ratings using a measure of contrastive stress cues: pairwise variability indexes (PVIs) of the four acoustic cues we measure. PVIs provide a measure of how the cue contrasts across two vowels. A positive PVI indicates greater stress on the first vowel whereas a negative PVI indicates greater stress on the second vowel. This yields fairly straightforward and interpretable results.

Our fifth and final aim concerns using our modeling results to, potentially, guide pedagogy. Pronunciation in Italian teaching, even with increased attention to phonology in recent textbooks, is not given enough focus in the classroom (Combei 2023). In at least three studies (De Meo et al. 2012, 2016; Pellegrino and Vigliano 2015), self-imitation techniques were used to test a relationship between pronunciation training and improvement in L2 Italian. All three studies found an improvement in prosody, but only two of those studies also saw a decrease in foreign accentedness. While these results are uneven and limited in scope, they support pronunciation training, specifically self-imitation techniques, as beneficial for learners' L2 speech development.

The study most closely related to our interest in pedagogical applications of a lexical stress experimental focus is Nicora et al. (2018). A group of Italian learners with the same teacher were divided into a control group ($N = 2$), who continued with their classes, and an experimental group ($N = 3$), who received intonation training in addition to their regular classes. The group who received training improved the most on a pre- and post-test that consisted of a reading task of words with varying lexical stress. While the authors contextualize these results with their limited sample size and some cases of improvement within the pretest itself (reducing the training effect), the study design offers an example of effective intonation training, which we aim to build on in our study.

Predictions

We predict that our L1 results (Aim 1) will be in line with previous research on L1 speech (e.g. Sulpizio and McQueen 2012): Italian antepenultimate stressed words will have third to last syllables with longer durations, higher average pitch, higher average amplitude, and higher spectral tilt than second to last syllables (Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). Penultimate stressed words will have third to last syllables with shorter durations, higher average pitch, and higher average amplitude than second to last syllables, and a relatively similar spectral tilt across the two syllables (Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). Although we make use of web-based recording and an uncontrolled participant environment, we have no reason to believe the overall results should differ given current advances in remote speech recording (see Zhang et al. 2021). This aim is largely meant to replicate previous findings and document our open methods. Prior research (Primativo et al. 2013; Spinelli et al. 2021) suggests early L2 learners will struggle to accurately produce the acoustic cues involved in Italian stress (Aim 2). This pattern is difficult for children acquiring their L1 (Arciuli and Colombo 2016) and L2 suprasegmental acquisition is seemingly difficult irrespective of the source and target language (e.g. Guion et al. 2004; Idemaru et al. 2019; Wiener et al. 2020). There is evidence on L2 acquisition of Italian geminates (Feng and Busa 2022) and stops (Feng and Busa 2023) which shows that even after three years of Italian language courses, L1 Mandarin Chinese speakers still struggled to acquire durational and voice onset time differences in speech. We also predict that listeners will rate the L1 speech and L2 speech at opposite ends of our VAS (Aim 3). We predict that duration will be a particularly informative cue in predicting responses (Bertinetto 1980; Alfano 2006) and may be easier for our L1 English–L2 Italian learners given that English syllable duration is far more variable than Italian syllable duration (Arciuli and Colombo 2016). We also predict interactions between the L2 group and the cues (Aim 4). These results will inform our pedagogical suggestions (Aim 5).

Methods

Acoustic analysis

Participants

Five adult L1 Italian speakers were recorded (mean age = 35; three female, two male). All five speakers were born in Italy and spoke Italian as their first language from birth. When asked to describe their variety of Italian, four speakers reported their accents as typical from the Emilia region in northern Italy and one speaker as from Sicily in southern Italy. All five speakers learned English as an L2 after puberty in Italy. Because there is considerable variation in regional Italian speech (e.g. Hajek 2003; Uguzzoni et al. 2003), this sample may not be representative of other L1 Italian speech. We return to this point in our limitations.

Twelve adult L2 Italian speakers took part in the recordings (mean age = 25; seven female, five male). All twelve speakers were born in the United States and spoke English as their first language from birth. All adults were enrolled in an Italian as a foreign language university class. At the time of recording, the L2 speakers had completed approximately eight weeks of structured classroom Italian lessons (roughly 200 minutes per week). Outside of class, the L2 speakers also had the opportunity to meet with a language assistant for additional speaking practice.

The study was approved by the authors' Institutional Review Board. The L1 participants volunteered their time. The L2 participants were paid as part of a larger study on language acquisition.

Stimuli

Sixty-four trisyllabic words—32 pairs—were taken from Sulpizio and McQueen (2012). These pairs shared the same first two syllables, e.g. *coDIno* and *COdice*, but differed in the third syllable and in stress assignment. The words varied in lexical frequency (using the *CoLFIS* database; Bertinetto et al. 2005) and familiarity. In a previous pilot study, we found a lack of consensus among the

Italian language instructors and L2 learners regarding which of the 64 words they knew or were exposed to. On our OSF repository, we have additional corpus frequency information and word familiarity data from three students and two instructors from the same population as those we test. We return to this point in our limitations.

Procedure

Recordings were made using Gorilla (Anwyl-Irvine et al. 2020). Participants were asked to record themselves in a quiet room without background noise. The 64 words were randomly presented one at a time on screen. Participants were asked to read the word aloud clearly and accurately within two seconds. There was a 250 ms interstimulus interval in which a fixation cross was shown. Recordings were saved as web audio (.weba) files and converted to 16-bit WAV files with 48 kHz sampling rate using FFmpeg (Tomar 2006). The recording took approximately four minutes.

The first two authors manually checked the recordings for noise and audio quality. Two L2 speakers' recordings were removed from the study for having poor audio quality. Praat (Boersma and Weenink 2024) was used to analyze the remaining 15 speakers. For each audio file, first and second vowel boundaries were marked in Praat Textgrids. Boundaries were set by simultaneous consideration of waveform and spectrogram and placed at the nearest zero crossing. Of the maximal 960 records, 900 were considered sufficient for the rating task (see next section) and therefore chosen for acoustic analysis; the 60 removed files contained additional noise. These 900 files were analyzed using an in-house script that leverages Parselmouth (Jadoul et al. 2018), a Python interface for Praat (Boersma and Weenink 2024), integrating its acoustic analysis capabilities into the R environment (R Core Team 2023) via reticulate (Ushey et al. 2022; see OSF for additional data including comparisons with Dicanio's (2023) approach). Our approach enabled automated and reproducible extraction of the key acoustic cues including duration, F0, amplitude, and spectral tilt. Of the possible 1,800 measurements per cue (900 words × two syllables), 1,782 were used as 18 yielded no measurements.

Duration in milliseconds was extracted directly from TextGrid annotations, where the onset and offset of each speech segment defined its temporal boundary. Fundamental frequency (F0) was extracted using autocorrelation-based pitch estimation, which identifies the dominant periodicity in the speech signal. A harmonic analysis was applied within a frequency range of 80–300 Hz, allowing for robust pitch tracking. Only voiced frames were considered, with mean F0 calculated as the average across all nonzero pitch estimates. Amplitude was computed as root mean square intensity, capturing the average energy of the waveform over time and representing perceived loudness. Spectral tilt was derived from the long-term average spectrum, measuring the relative amplitude of different frequency bands. Energy values were computed for low (0–500 Hz) and high (2,000–4,000 Hz) frequency bands, and tilt was calculated as the logarithmic ratio of low-frequency to high-frequency power, reflecting the spectral balance of the signal. Additionally, the harmonic-to-harmonic difference (H1–H2) was measured by extracting spectral amplitudes at the first (H1) and second harmonics (H2), providing insight into glottal voice quality.

Data analysis

By-participant outliers beyond three absolute deviations from the median (Leys et al. 2013) were removed for each cue (duration ~1 per cent, F0 ~7 per cent, amplitude ~5 per cent, spectral tilt ~2 per cent). For plotting and analysis, F0 was Z-scored (Lobanov 1971). For each cue, we were interested in how measurements varied across vowels and whether this variation was similar across L1 and L2 speakers. Mixed effects linear regression models were built in R version 4.4.0 with a .05 alpha-level. Separate antepenultimate and penultimate models were built for each cue. Both predictor variables were two-level categorical variables with the first vowel, and L1 as the reference levels. Changes from the intercept reflect how the second vowel and L2 levels differ, respectively. A two-way interaction with vowel and group was also tested. A positive interaction indicates the

change between the two vowels is greater for L2 speakers whereas a negative interaction indicates that the change is reduced for L2 speakers. Participants and items were included as random intercepts. Given the relatively small number of observations, random slopes were not included as it would have overfit the model. Each model was releveled with L2 as the reference level to confirm any observed L2 changes from the first vowel to the second.

Rating task

Participants

Eighty-four participants were initially recruited on Prolific. These participants were required to have indicated in their Prolific profile that Italian was their first language from birth and their most fluent language. To maintain consistency across participants, we additionally required that all participants indicate being born and educated in Italy. Participants were also required to pass a basic hearing screening (with headphones) using a dichotic pitch task (Milne et al. 2021) and score above an 80 per cent accuracy on Amenta et al. (2021), which is a quick Italian lexical proficiency test used to gauge L1/L2 abilities. Bramlett and Wiener (2025a) found that 80 per cent serves as a reliable cutoff for estimating L1 abilities while still preserving sufficient data. Finally, participants had to pass an attention-check roughly halfway through the rating task. This task involved counting images on the screen within a time limit. After removing participants who failed any of the screening tasks, 50 participants remained. See OSF materials for additional participant information including biological sex, age, and education levels. The study was approved by the authors' Institutional Review Board and participants were paid for their time.

Stimuli

The 900 recordings used in the acoustic analysis (normalized for intensity) were used as stimuli for the rating task.

Procedure

The rating task was carried out on Gorilla (Anwyl-Irvine et al. 2020). Because the experiment was advertised in English, instructions for the task were also given in English. Participants were told that they would hear "native Italian speakers and learners of Italian" and we were interested in how "fluent" these spoken words' stress sounded to you "a native listener." Participants first did two practice trials to get used to the visual analog scale (VAS). A spoken word was played while a VAS slider was simultaneously shown on screen with "NOT fluent" and "VERY fluent" as endpoints. Participants were required to click anywhere on the slider (adjustments could be made after the first click) and then press the spacebar to continue to the next sound. Participants were encouraged to "use the whole scale." After the practice trials, participants began the main task. The 900 words were randomly split across 9 lists of 100 items. Participants were assigned one list in a counterbalanced order; that is, each list had different items and each listener rated only one list. Words were presented with a 500 ms interstimulus interval. The task took approximately 14 minutes and included a timed attention-check roughly half way through the task. Each word was rated by a minimum of five different participants. The task had high inter-rater reliability as measured by Cronbach's Alpha ($\alpha = .97$).

Data analysis

For each acoustic measure, a normalized pairwise variability index (PVI: Nolan and Asu 2009) was calculated: $100 * [(V_1 - V_2)/(V_1 + V_2)/2]$ where V_1 and V_2 are measures of duration, F0, amplitude, and spectral tilt on the first and second vowels, respectively. PVIs provide a measure of how the cue contrasts across the first and second vowel. A positive PVI indicates greater stress on the first syllable whereas a negative PVI indicates greater stress on the second syllable. Values beyond three absolute deviations from the median (Leys et al. 2013) were replaced with by-participant

condition means (see Arciuli and Colombo 2016). This was roughly 4 per cent for L1 and 8 per cent for L2 participants. Of the 900 possible sounds, 885 were analyzed. Fifteen sounds were removed for having background noise. From the 885 files, four PVIs were calculated. Roughly 13 per cent of the possible 3,540 were removed because our Praat scripts could not obtain reliable measurements (e.g. unreliable pitch tracking) leaving 3,067 PVI measurements.

There are many ways to model our data. Here we report two linear models and two generalized additive models (GAM). GAMs allow us to model more complex, non-linear relationships. Our online R code outlines our modeling approach in detail. In brief, both models included speaker group (L1/L2) as a categorical predictor with L1 as the reference level. In the linear model, we include the four cues' PVIs as predictors; in the GAM model, we use the four cues as smooths. Both models included participants and words as random effects.

Data availability

All stimuli, data, and code are available via the Open Science Framework: https://osf.io/72w8v/

Results

Acoustic analysis

Figure 1 plots the raw data of the four acoustic cues analyzed. Beneath each plot is the regression model output for that cue. The antepenultimate duration model (conditional $R^2 = 0.46$) revealed that the duration of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 94.65$, SE = 5.25, $P < .001$). The L1 second vowel was significantly shorter than the L1 first vowel ($\beta = -39.78$, SE = 4.22, $P < .001$) whereas the L2 second vowel was significantly longer than the L2 first vowel ($\beta = 54.88$, SE = 3.11, $P < .001$). The penultimate duration model (conditional $R^2 = 0.55$) revealed that the duration of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = -17.17$, SE = 5.31, $P = .001$). The L1 second vowel was significantly longer than the L1 first vowel ($\beta = 78.44$, SE = 4.25, $P < .001$). The L2 second vowel was significantly longer than the L2 first vowel ($\beta = 61.27$, SE = 3.18, $P < .001$), though this L2 difference was smaller than the L1 difference.

The antepenultimate F0 model (conditional $R^2 = 0.18$) revealed that the scaled F0 difference of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 0.51$, SE = 0.10, $P < .001$). The L1 second vowel had a significantly lower scaled F0 than the L1 first vowel ($\beta = -0.59$, SE = 0.08, $P < .001$) whereas the L2 second vowel had a statistically similar scaled F0 to the L2 first vowel ($\beta = -0.07$, SE = 0.06, $P = .19$). The penultimate F0 model (conditional $R^2 = 0.23$) revealed that the scaled F0 difference of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 0.75$, SE = 0.09, $P < .001$). The L1 second vowel had a significantly lower scaled F0 than the L1 first vowel ($\beta = -0.84$, SE = 0.07, $P < .001$) whereas the L2 second vowel had a statistically similar scaled F0 to the L2 first vowel ($\beta = -0.08$, SE = 0.05, $P = .12$).

The antepenultimate amplitude model (conditional $R^2 = 0.47$) revealed that the amplitude difference of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 3.27$, SE = 0.36, $P < .001$). The L1 second vowel had a significantly lower amplitude than the L1 first vowel ($\beta = -3.49$, SE = 0.29, $P < .001$) whereas the L2 second vowel had a statistically similar amplitude to the L2 first vowel ($\beta = -0.22$, SE = 0.21, $P = .31$). The penultimate amplitude model (conditional $R^2 = 0.44$) revealed that the amplitude difference of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 2.16$, SE = 0.37, $P < .001$). The L1 second vowel had a significantly lower amplitude than the L1 first vowel ($\beta = -2.73$, SE = 0.30, $P < .001$). The L2 second vowel had a significantly lower amplitude than the L2 first vowel ($\beta = -0.57$, SE = 0.23, $P = .01$), though this L2 difference was smaller than the L1 difference.

The antepenultimate spectral tilt model (conditional $R^2 = 0.69$) revealed that the spectral tilt difference of the first and second vowels varied, and this variation differed significantly between the L1 and L2 groups ($\beta = 1.30$, SE = 0.20, $P < .001$). The L1 second vowel had a significantly lower

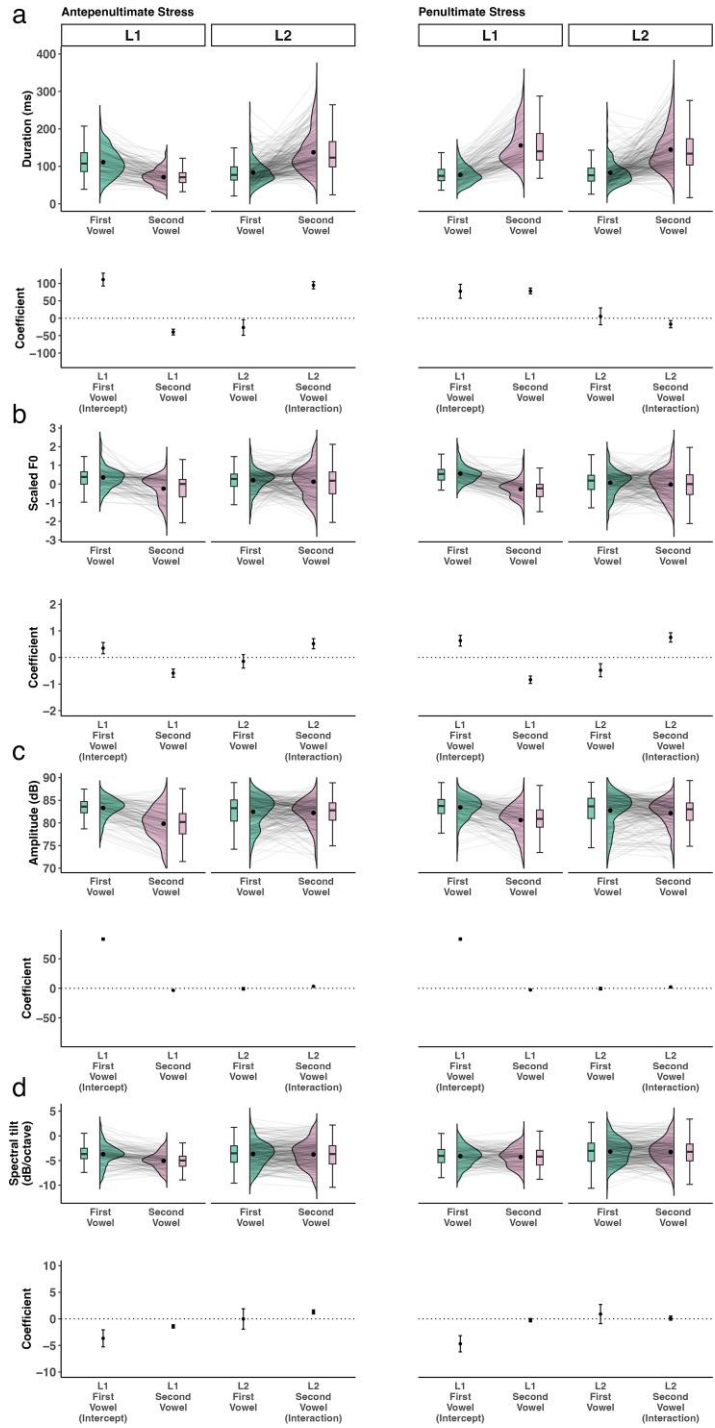

Figure 1. Acoustic results by cue (rows), speaker group (columns), and stress (columns). Raw data with individual measurements in a word connected by gray lines, group box plots, group means (black point), and density plots are shown for duration (a), scaled F0 (b), amplitude (c), and spectral tilt (d). Below each plot is the regression model coefficients with 95 per cent confidence intervals.

spectral tilt than the L1 first vowel ($\beta = -1.42$, SE = 0.16, $P < .001$). The L2 second vowel had a statistically similar spectral tilt to the L2 first vowel ($\beta = -0.12$, SE = 0.12, $P = .29$). The penultimate spectral tilt model (conditional $R^2 = 0.67$) revealed no significant results ($ps > .1$).

Rating task

Figure 2a plots the VAS responses by individual speaker and stress type. All participants—even the L1 speakers—showed variability and a range of perceived "very fluent" to "not fluent" responses. Nearly 20 per cent of our VAS responses were at the two ends of the scale while over 80 per cent of responses were between anchors. Figure 2b shows the normalized PVIs of the four cues and highlights how amplitude and F0 have smaller dispersions near 0 whereas duration and spectral tilt have larger dispersions. Figure 2c visualizes the data with linear fits and shows how duration PVI has slopes in opposite directions whereas the other PVIs have relatively similar slopes across the two stress types. Figure 2d visualizes the non-linear (i.e. generalized additive) smooth fit across all the cues and shows how antepenultimate stress has a relatively clear response contrast given the PVI whereas penultimate stress shows relatively similar response irrespective of the PVI. Table 1 reports the output for the four models.

In both the antepenultimate linear model (conditional $R^2 = 0.77$) and the penultimate linear model (conditional $R^2 = 0.60$), L1 speakers were rated higher than L2 speakers, as expected. Duration PVI was a significant predictor for antepenultimate responses whereas F0 PVI was a significant predictor for penultimate responses. All other PVI predictors in the linear models were null ($ps > 0.05$). In the antepenultimate GAM model (conditional $R^2 = 0.77$) and penultimate GAM model (conditional $R^2 = 0.62$), L1 speakers were once again rated higher than L2 speakers, as expected. Duration PVI and spectral tilt PVI were significant smooth terms in the antepenultimate model; F0 and amplitude were null ($ps > 0.05$). All four PVIs were significant smooth terms in the penultimate model.

Discussion

L1 Italian acoustics replication

Our acoustic analysis results indicate that our five L1 speakers used duration and spectral tilt in ways that could be sufficiently informative to the listener (Fig. 1). Third to last syllables were

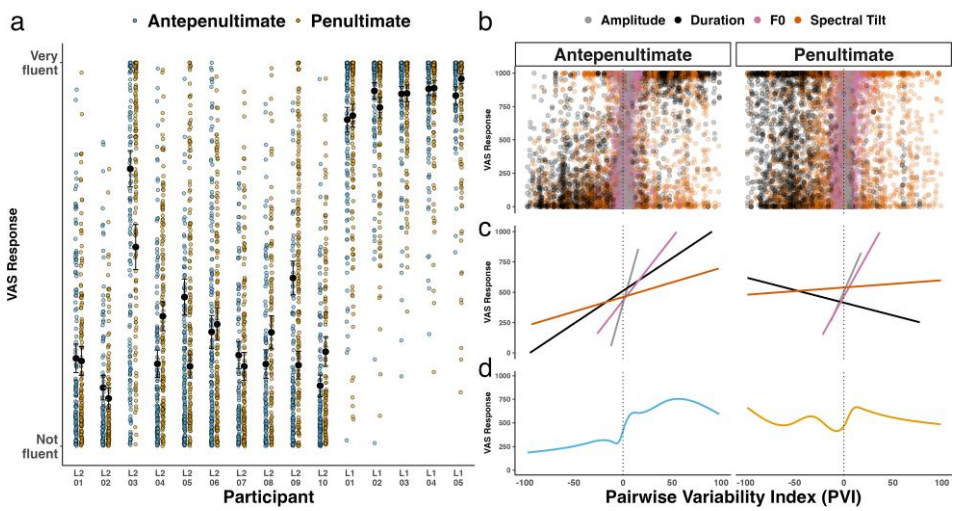

Figure 2. VAS responses (a) showing individual utterances by participant and stress type (color coded). Means and 95 per cent confidence intervals are shown in black. Normalized PVI responses color coded by acoustic cue (b). Linear fits of the four cues (c) and smoothed generalized additive fits (d).

Table 1. Model output for linear and generalized additive models.

Antepenultimate linear model					Penultimate linear model				
	Estimate	SE	t value	P-value		Estimate	SE	t value	P-value
L1 Group (Intercept)	881.06	16.55	53.24	< .001	**L1 Group (Intercept)**	874.87	20.85	41.96	<.001
F0 PVI	0.17	0.48	0.34	0.73	**F0 PVI**	1.77	0.63	2.82	.005
Duration PVI	0.87	0.12	7.12	< .001	**Duration PVI**	0.04	0.16	0.27	.79
Spectral Tilt PVI	−0.01	0.06	−0.02	0.98	**Spectral Tilt PVI**	0.09	0.08	1.2	.23
Amplitude PVI	0.66	1.03	0.64	0.53	**Amplitude PVI**	1.12	1.26	0.89	.37
L2 Group	−638.07	12.61	−50.62	< .001	**L2 Group**	−562.75	12.65	−44.49	<.001
Antepenultimate GAM model					**Penultimate GAM model**				
	Estimate	SE	t value	P-value		Estimate	SE	t value	P-value
L1 Group (Intercept)	850.93	16.97	50.16	< .001	**L1 Group (Intercept)**	854.27	18.5	46.18	< .001
L2 Group	−602.45	14.06	−42.84	< .001	**L2 Group**	−518.4	13.89	−37.33	<.001
	Smooth terms:					Smooth terms:			
	edf	Ref.df	F	P-value		edf	Ref.df	F	P-value
s(F0 PVI)	2.6	3.32	1.37	0.23	**s(F0 PVI)**	5.92	7.24	5.38	<.001
s(Duration PVI)	5.87	6.99	12.2	< .001	**s(Duration PVI)**	8.05	9.89	7.5	<.001
s(Spectral tilt PVI)	6.35	7.44	2.6	0.01	**s(Spectral tilt PVI)**	7.19	8.87	2.5	.01
s(Amplitude PVI)	1.01	1.02	1.05	0.3	**s(Amplitude PVI)**	5.23	6.47	4.05	<.001
s(Participant)	44.92	49	11.92	< .001	**s(Participant)**	42.92	49	8.23	<.001
s(Items)	17.25	31	1.34	0.001	**s(Items)**	23.66	31	3.49	<.001

longer and had higher spectral tilt for antepenultimate stress whereas second to last syllables were longer and had a similar (level) spectral tilt for penultimate stress. In both stress patterns, F0 and amplitude decreased from the third to last syllable to the second to last syllable. This overall pattern of duration, F0, amplitude, and spectral tilt corroborates a large number of studies on Italian stress production (D'Imperio and Rosenthal 1999; Tagliapietra and Tabossi 2005; Tagliapietra and McQueen 2010; Sulpizio and McQueen 2012; Arciuli and Colombo 2016; Albano Leoni and Maturi 2018; Bramlett and Wiener 2025a). Importantly, we demonstrated that we were able to replicate careful, in-person acoustic-phonetics research using entirely open and transparent web-based methods. We echo the calls to use web-based methods to access underdocumented populations and languages (see Hinton and Hale 2013).

L2 Italian acoustics findings

Our acoustic analysis results indicate that our sample of ten adult beginner L2 speakers struggled to produce the four acoustic cues in ways that could be informative to the listener (Fig. 1). Given the difficulty L1 Italian children have acquiring lexical stress (Arciuli and Colombo 2016), and the general challenge L2 prosody acquisition poses for adults (Guion et al. 2004), these results are exactly what predicted. L2 speakers produced longer second to last syllables than third to last syllables for both stress types. This corroborates findings that duration is a challenging cue to acquire

in an L2 (Primativo et al. 2013; Spinelli et al. 2021) and suggests an area of consideration for pedagogy, especially when producing antepenultimate stress. L2 speakers also failed to produce consistent F0 or spectral tilt differences; we did, however, find a decrease in amplitude only for penultimate stressed words. There is limited evidence that acquiring amplitude cues in an L2 is difficult (Wiener et al. 2022), though more research is needed to better understand this acquisition process. We are unaware of any research on acquisition of L2 spectral tilt cues. This highlights another potential area for teaching, which we return to below.

L1 ratings of L1 and L2 speech using visual analog scales

As expected, L1 speech was rated higher—or in our case, more fluent—than L2 speech (Fig. 2a). Moreover, our participants were in very high agreement about how to rate the utterances. Importantly, the VAS task provided more information beyond a binary "correct" or "incorrect" choice. Our L2 speakers were all recruited from the same in-person classes and yet there was considerable variation in their speech ratings. Even among the L1 speakers, there was variation that might otherwise be obscured in a correct/incorrect forced-choice task. These VAS results are in line with theoretical frameworks involving gradient speech perception (Apfelbaum et al. 2022). We present clear evidence that listeners are sensitive to relative acoustic differences (e.g. PVIs) that might otherwise be ignored in traditional categorization tasks (see Kutlu et al. 2022). Importantly, VAS tasks allow for more sensitive data that can better inform—and build—theories of L2 perception and production (see Apfelbaum et al. 2022 for discussion).

Modeling VAS responses using pairwise variability indices

Our PVI approach (Fig. 2b) allowed us to examine how each cue varied over the syllables, and test whether this measure could predict VAS responses. Our adult L1 PVI means are in the same direction as those reported in Arciuli and Colombo (2016; see Fig. 2b and OSF materials) whereas our adult L2 PVI means do not resemble L1 adult means or L1 children means. We built linear (Fig. 2c) and generalized additive (Fig. 2d) models to account for the VAS responses. Our linear models indicated that duration PVI—longer third to last syllable duration than second to last syllable duration—predicted antepenultimate responses whereas F0 PVI—larger F0 decrease from the third to last syllable to the second to last syllable—predicted penultimate responses. Our GAM models indicated that duration and spectral tilt PVIs were significant smooth terms across both stress types; F0 and amplitude PVIs were significant smooth terms for penultimate only. These results motivate the claim that Italian listeners use all four cues when perceiving—and rating—Italian stress. Duration appears to play a large role as shown in the GAM models and opposite PVI slopes (Fig. 2c; e.g. Alfano et al. 2007, 2009; Bertinetto 1980, 1981). Perceived pitch (decrease) plays a larger role for penultimate stress than for antepenultimate stress (see Caccia et al. 2019). Importantly, we found spectral tilt differences in both our GAM models, suggesting a small, but important role for higher frequencies in stressed vowels and the relative decrease for antepenultimate stressed words (Sluijter and Van Heuven 1996; Sulpizio and McQueen 2012; Bramlett and Wiener 2025a). The role of amplitude was much smaller and only found in our penultimate GAM model, suggesting the contribution may be minimal—or at least minimal to our population given L2 speakers only produced amplitude differences for penultimate stressed words (cf. Albano Leoni and Maturi 2018). In terms of our methodological contributions, our approach highlights how GAM with smooth terms can provide further insight by capturing non-linear relationships (see Coretta and Casillas 2025; Bramlett and Wiener 2025b).

We note that our finding is novel in that we assessed how acoustic cues were used to rate L1 and L2 speech as opposed to only recognize spoken words. That is, listeners not only needed to map the variable acoustic signal to a mental representation, but then also needed to evaluate that particular speech exemplar on a sliding scale. Crucially, we found evidence for a reliance on a variety of acoustic cues when evaluating speech. Italian stress is thus not conveyed by one cue alone (Llisterri et al. 2003) nor is it rated by one cue alone.

Contributions to pedagogy

In addition to learning information about each word's stress pattern, and recognizing less common stress patterns, L2 learners have to differentiate between stress types in their speech. This is a difficult task that takes children years to master (Arciuli and Colombo 2016). We found duration, F0, and spectral tilt play important roles in our L1 acoustic analysis and L1 rating task. We acknowledge that spectral tilt is a difficult concept to convey to the typical L2 learner, and set that finding aside for future follow-up studies. This leaves duration and pitch as teachable concepts and areas to focus on for instructors and learners. To effectively acquire penultimate stress production, a speaker must produce a longer second to last syllable duration and a sharper decrease in pitch from the third to last syllable to the second to last syllable. To effectively acquire antepenultimate stress production, a speaker must produce a longer third to last syllable duration and a less sharp pitch decrease across syllables.

When considering how to draw greater attention to these cues, instructors might guide learners in visualizing duration and pitch differences using waveform tools. Learners can use Praat (Boersma and Weenink 2024) to record themselves producing correct/incorrect stress for a single target word, or minimal pairs with different stress placement, and guess which waveform corresponds to each form. Such visual feedback is a common feature in web and mobile apps focusing on pronunciation and language learning. For example, Mango Languages presents the target waveform with an overlay of the app user's waveform for comparison and simultaneous playback. Such tools present a novel resource also for classroom use.

For learners of Italian, it helps that antepenultimate stressed Italian words are less common (roughly 18 per cent of trisyllabic words; Thornton et al. 1997) than penultimate stressed words (roughly 80 per cent of trisyllabic words). Penultimate stressed words are likely even more common for beginner learners, as elementary curriculum focuses on acquiring patterns such as the *-mente* adverb (*veloceMENte* "quickly") and infinitive forms of high frequency verbs (*anDAre* "to go," *parLAre* "to speak"). Exceptions are taught as grammatical deviations, i.e. verbs are categorized as regular/irregular based on the spelling of conjugated endings or changes to the word stem. Verbs are not grouped by "irregular stress" patterns, e.g. *Abito* "to inhabit/live" is a regular verb despite its antepenultimate stress in the present singular and third-person plural conjugations.

Limitations and conclusion

Our findings should be interpreted with caution given three shortcomings to our study: (1) we tested only two stress types (antepenultimate and penultimate), (2) our L1 Italian sample was small and regionally skewed toward Emilia–Romagna and Sicily; and (3) our L2 group consisted exclusively of L1 English speakers, so results cannot be generalized to learners from other language backgrounds. It is unclear to what degree our L2 results hold compared to more advanced students, students immersed in the L2 or even students with different L1 backgrounds. We know that the L1/L2 lexicon matters in terms of speech production (Primativo et al. 2013; Nicora et al. 2018; Spinelli et al. 2021). Our L2 participants obviously had a relatively small Italian lexicon and it is an open question whether these findings will hold once more words are learned. In our available R code we share equivalence test results showing equivalent acoustic results between known and unknown words. We also include lexical frequency data and show a null effect across all our models. Whether our results hold given a more complex speaking task (i.e. not list reading but dialog with an interlocutor), different stress types (i.e. not just antepenultimate and penultimate stress), varied L1 backgrounds (i.e. not L1 English speakers), and across participants with varied cognitive and perceptual abilities (see Saito 2023; Bramlett et al. 2024 for discussions) are all open questions. An additional limitation concerns the linguistic background of our L1 Italian speakers and raters. Although all participants were L1 Italian speakers born and educated in Italy, they necessarily had some exposure to English due to the study's recruitment and instructions being in English. Moreover, our L1 speakers undoubtedly had variation in their regional Italian speech (e.g. Hajek 2003; Uguzzoni et al. 2003); this sample may not be representative of other

L1 Italian speech. A more balanced and representative sample of L1 speakers, covering several regional backgrounds would yield clearer results. Additionally, to what degree our instructions involving rating the "stress fluency" (as opposed to intelligibility or comprehensibility) affected our results is an open question. Given our task and stimuli, we believe any construct would yield similar ratings such that L1 speech would be rated higher than L2 speech. Future work may explore this topic. Finally, we did not control for broader multilingual experience. It remains an open question whether monolingual Italian listeners, or Italian speakers with different degrees of exposure to other languages, might evaluate L2 Italian speech differently. Future research could directly compare listener groups to better understand how linguistic background shapes perceptions of L2 fluency.

To conclude, accurate production of Italian lexical stress requires coordination of multiple acoustic cues over multiple syllables. This is difficult for children acquiring their L1 and adults acquiring their L2. We found duration and pitch play important roles in antepenultimate and penultimate stress. We suggest ways these cues can be taught and encourage readers to make use of our open materials and methods to continue to research this topic.

Acknowledgements

We thank Giuseppina Gemboni for her help with the second study.

Funding

This work was supported by Carnegie Mellon University's Simon Initiative seed grant and Dietrich Faculty seed grant.

Notes on contributors

Seth Wiener is an Associate Professor at Carnegie Mellon University. His research interests are phonetics, psycholinguistics, and second language acquisition. He directs the Language Acquisition, Processing, and Pedagogy Lab, and the CMU Applied Linguistics and Second Language Acquisition doctoral program. Email: sethw1@cmu.edu

Adam A. Bramlett is a PhD candidate in Second Language Acquisition at Carnegie Mellon University. His research focuses on learning mechanisms, lexical and phonetic processing, and computational modeling of language acquisition. He has also taught Chinese language courses and contributed to assessment and psycholinguistics research initiatives at CMU. Email: abramlet@andrew.cmu.edu

Bianca Brown is a postdoctoral fellow at Carnegie Mellon University in Qatar. Bianca has taught elementary Italian courses, and focuses on interaction and learner identity in her research. Email: biancab@cmu.edu

Jocelyn Dueck is an Associate Professor of music at Carnegie Mellon, and a subject matter expert in prosody and related auditory coding. She has a distinguished record as a practitioner in music and language translation, training others to reproduce sound units in multiple languages and English dialects. Email: jdueck@andrew.cmu.edu

References

Albano Leoni, F., and Maturi, P. (2018) *Manuale di Fonetica; Nuova Edizione*. Rome: Carocci Editore.

Alfano, I. (2006) 'La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo [Lexical stress perception: comparing Italian and Spanish],' in R. Savy and C. Crocco (eds.) Proceedings of 2nd AISV (Associazione Italiana di Scienze della Voce), pp. 632-656. Salerno: EDK Editore.

Alfano, I., Llisterri, J., and Savy, R. (2007) 'The perception of Italian and Spanish lexical stress: a first cross-linguistic study', in Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, pp. 1793–6. International Phonetic Association.

Alfano, I., Savy, R., and Llisterri, J. (2009) 'Sulla realtà acustica dell'acento lessicale in italiano ed in spagnolo: La durata vocalica in produzione e percezione [On the acoustic reality of the lexical accent in Italian and Spanish: The vowel duration in production and perception]', in L. Romito, V. Galatà, and R. Lio (eds.) Proceedings of 4th AISV (Associazione Italiana di Scienze della Voce), pp. 22–39. Torriana: EDK Editore.

Amenta, S., Badan, L., and Brysbaert, M. (2021) 'LexITA: A Quick and Reliable Assessment Tool for Italian L2 Receptive Vocabulary Size', *Applied Linguistics*, 42: 292–314. https://doi.org/10.1093/applin/amaa020

Anwyl-Irvine, A. L. et al. (2020) 'Gorilla in our Midst: An Online Behavioral Experiment Builder', *Behavior Research Methods*, 52: 388–407. https://doi.org/10.3758/s13428-019-01237-x

Apfelbaum, K. S. et al. (2022) 'Don't Force It! Gradient Speech Categorization Calls for Continuous Categorization Tasks', *The Journal of the Acoustical Society of America*, 152: 3728–45. https://doi.org/10.1121/10.0015201

Arciuli, J., and Colombo, L. (2016) 'An Acoustic Investigation of the Developmental Trajectory of Lexical Stress Contrastivity in Italian', *Speech Communication*, 80: 22–33. https://doi.org/10.1016/j.specom.2016.03.002

Bellocchi, S., Bonifacci, P., and Burani, C. (2016) 'Lexicality, Frequency and Stress Assignment Effects in Bilingual Children Reading Italian as A Second Language', *Bilingualism: Language and Cognition*, 19: 89–105. https://doi.org/10.1017/S1366728914000297

Bent, T. et al. (2024) 'Relating Pronunciation Distance Metrics to Intelligibility Across English Accents', *Journal of Phonetics*, 107: 101357. https://doi.org/10.1016/j.wocn.2024.101357

Bertinetto, P. M. (1980) 'The Perception of Stress by Italian Speakers', *Journal of Phonetics*, 8: 385–95. https://doi.org/10.1016/S0095-4470(19)31495-0

Bertinetto, P. M. (1981) *Strutture prosodiche dell'italiano*. Firenze: Accademia della Crusca.

Bertinetto, P. M. et al. (2005) Corpus e lessico di frequenza dell'Italiano scritto (CoLFIS). [Corpus and frequency lexicon of written Italian]. "Antonio Zampolli" Computer Linguistics Institute - ILC Istituto di Scienze e Tecnologie della Cognizione - ISTC.

Boersma, P. and Weenink, D. (2024) *Praat: doing phonetics by computer* [Computer program]. Version 6.4.23, http://www.praat.org/, accessed 27 Oct. 2024.

Bramlett, A. A. et al. (2024) 'Measuring music and prosody: accounting for variation in non-native speech discrimination with working memory, specialized music skills, and music background', in Proc. Speech Prosody 2024, pp. 482–6. International Speech Communication Association.

Bramlett, A. A., and Wiener, S. (2025a) 'Individual Differences Modulate Prediction of Italian Words Based on Lexical Stress: A Close Replication and LASSO Extension of Sulpizio and McQueen (2012)', *Journal of Cultural Cognitive Science*, 9: 55–81. https://doi.org/10.1007/s41809-024-00162-6

Bramlett, A. A., and Wiener, S. (2025b) 'The art of Wrangling: Working with web-Based Visual World Paradigm eye-Tracking Data in Language Research', *Linguistic Approaches to Bilingualism*, 15: 538–70. https://doi.org/10.1075/lab.23071.bra

Brown, B. et al. (2023) 'Searching for the "Native" Speaker: A Preregistered Conceptual Replication and Extension of Reid, Trofimovich, and O'Brien (2019)', *Applied Psycholinguistics*, 44: 475–94. https://doi.org/10.1017/S0142716423000127

Caccia, M. et al. (2019) 'Pitch as the Main Determiner of Italian Lexical Stress Perception Across the Lifespan: Evidence from Typical Development and Dyslexia', *Frontiers in Psychology*, 10: 1458. https://doi.org/10.3389/fpsyg.2019.01458

Cheng, L. S. et al. (2021) 'The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology with Inclusive and Accurate Measures', *Frontiers in Psychology*, 12: 715843. https://doi.org/10.3389/fpsyg.2021.715843

Combei, C. R. (2023) *Speaking Italian with a Twist : A Corpus Study of Perceived Foreign Accent*. Milano: Franco Angeli. https://digital.casalini.it/9788835154716

Coretta, S., and Casillas, J. V. (2025) 'A Tutorial on Generalised Additive Mixed Effects Models for Bilingualism Research', *Linguistic Approaches to Bilingualism*, 15: 429–52. https://doi.org/10.1075/lab.23076.cor

de Jong, K., and Zawaydeh, B. A. (1999) 'Stress, Duration, and Intonation in Arabic Word-Level Prosody', *Journal of Phonetics*, 27: 3–22. https://doi.org/10.1006/jpho.1998.0088

De Meo, A. et al. (2012) 'Imitation/self–imitation in computer–assisted prosody training for Chinese learners of L2 Italian', in J. Levis and K. LeVelle (eds.) Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, pp. 90–100. Vancouver, 24–25 August 2012. Iowa State University Digital Press.

De Meo, A., Vitale, M., and Pellegrino, E. (2016) 'Tecnologia della voce e miglioramento della pronuncia in una L2: imitazione e autoimitazione a confronto. Uno studio su sinofoni apprendenti di italiano L2', in F. Bianchi, and P. Leone (eds.) *Studi AItLA 4: Linguaggio e apprendimento linguistico*, pp. 13–25. Milano: AItLA.

Derwing, T. M. et al. (2009) 'The Relationship Between L1 Fluency and L2 Fluency Development', *Studies in Second Language Acquisition*, 31: 533–57. https://doi.org/10.1017/S0272263109990015

Dicanio, C. (2023) *Praat Scripts*. https://www.acsu.buffalo.edu/~cdicanio/scripts.html

D'Imperio, M., and Rosenthall, S. (1999) 'Phonetics and Phonology of Main Stress in Italian', *Phonology*, 16: 1–28. https://doi.org/10.1017/S0952675799003681

Eriksson, A. et al. (2016) 'The acoustics of lexical stress in Italian as a function of stress level and speaking style', in Proceedings of INTERSPEECH, pp. 1059–63, San Francisco, CA. International Speech Communication Association (ISCA).

Fava, E., and Magno-Caldognetto, E. (1976) 'Studio sperimentale delle caratteristiche elettroacustiche delle vocali toniche e atone in bisillabi italiani', in R. Simone, U. Vignuzzi, and G. Ruggiero (eds.) *Atti del Convegno Internazionale di Studi di Fonetica e Fonologia*, pp. 35–79. Rome: Bulzoni.

Feng, Q., and Busà, M. G. (2022) 'Mandarin Chinese-Speaking Learners' Acquisition of Italian Consonant Length Contrast', *System*, 111: 102938. https://doi.org/10.1016/j.system.2022.102938

Feng, Q., and Busà, M. G. (2023) 'Acquiring Italian Stop Consonants: A Challenge for Mandarin Chinese-Speaking Learners', *Second Language Research*, 39: 759–83. https://doi.org/10.1177/02676583221079147

Ferrero, F. (1972) 'Caratteristiche acustiche dei fonemi vocalici italiani', *Parole Metodi*, 3: 9–32. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7530025092

Gemelli, A. (1950) *La Strutturazione Psicologica del Linguaggio Studiata Mediante lanalisi Elettroacustica*. Città del Vaticano: Pontificiae Academiae Scientiarum Scripta Varia.

Guion, S. G., Harada, T., and Clark, J. J. (2004) 'Early and late Spanish–English bilinguals' acquisition of English word stress patterns', *Bilingualism: Language and Cognition*, 7: 207–26. https://doi.org/10.1017/S1366728904001592

Hajek, J. (2003) 'Patterns of vowel nasalisation in northern Italy: articulatory versus perceptual', in Proceedings of the 15th International Congress of the Phonetic Sciences, pp. 235–8. International Phonetic Association.

Hinton, L., and Hale, K. (2013) *The Green Book of Language Revitalization in Practice*. Leiden, The Netherlands: Brill.

Idemaru, K., Wei, P., and Gubbins, L. (2019) 'Acoustic Sources of Accent in Second Language Japanese Speech', *Language and Speech*, 62: 333–57. https://doi.org/10.1177/0023830918773118

Isaacs, T., and Thomson, R. I. (2013) 'Rater Experience, Rating Scale Length, and Judgments of L2 Pronunciation: Revisiting Research Conventions', *Language Assessment Quarterly*, 10: 135–59. https://doi.org/10.1080/15434303.2013.769545

Jadoul, Y., Thompson, B., and de Boer, B. (2018) 'Introducing Parselmouth: A Python Interface to Praat', *Journal of Phonetics*, 71: 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Krämer, M. (2009) *The Phonology of Italian*. New York: Oxford University Press.

Kutlu, E., Chiu, S., and McMurray, B. (2022) 'Moving Away from Deficiency Models: Gradiency in Bilingual Speech Categorization', *Frontiers in Psychology*, 13: 1033825. https://doi.org/10.3389/fpsyg.2022.1033825

Lehiste, I., and Peterson, G. E. (1959) 'Vowel Amplitude and Phonemic Stress in American English', *The Journal of the Acoustical Society of America*, 31: 428–35. https://doi.org/10.1121/1.1907729

Levis, J. M. (2005) 'Changing Contexts and Shifting Paradigms in Pronunciation Teaching', *TESOL Quarterly*, 39: 369–77. https://doi.org/10.2307/3588485

Leys, C. et al. (2013) 'Detecting Outliers: Do not use Standard Deviation Around the Mean, use Absolute Deviation Around the Median', *Journal of Experimental Social Psychology*, 49: 764–6. https://doi.org/10.1016/j.jesp.2013.03.013

Lieberman, P. (1960) 'Some Acoustic Correlates of Word Stress in American English', *The Journal of the Acoustical Society of America*, 32: 451–4. https://doi.org/10.1121/1.1908095

Lindemann, S., and Subtirelu, N. (2013) 'Reliably Biased: The Role of Listener Expectation in the Perception of Second Language Speech', *Language Learning*, 63: 567–94. https://doi.org/10.1111/lang.12014

Llisterri, J. et al. (2003) 'The perception of lexical stress in Spanish', in Proceedings of the 15th International Congress of Phonetic Sciences, pp. 2023–6, Barcelona. International Phonetic Association.

Lobanov, B. M. (1971) 'Classification of Russian Vowels Spoken by Different Speakers', *The Journal of the Acoustical Society of America*, 49: 606–8. https://doi.org/10.1121/1.1912396

McClean, M. D., and Tiffany, W. R. (1973) 'The Acoustic Parameters of Stress in Relation to Syllable Position, Speech Loudness and Rate', *Language and Speech*, 16: 283–90. https://doi.org/10.1177/002383097301600310

Milne, A. E. et al. (2021) 'An Online Headphone Screening Test Based on Dichotic Pitch', *Behavior Research Methods*, 53: 1551–62. https://doi.org/10.3758/s13428-020-01514-0

Missaglia, F. (1999). 'Contrastive prosody in SLA: an empirical study with adult Italian learners of German', in Proceedings from: 14th International Congress of Phonetic Sciences, pp. 551–4, San Francisco. International Phonetic Association.

Morton, J., and Jassem, W. (1965) 'Acoustic Correlates of Stress', *Language and Speech*, 8: 159–81. https://doi.org/10.1177/002383096500800303

Munro, M. J., and Derwing, T. M. (1995) 'Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners', *Language Learning*, 45: 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., Derwing, T. M., and Morton, S. L. (2006) 'The Mutual Intelligibility of L2 Speech', *Studies in Second Language Acquisition*, 28: 111–31. https://doi.org/10.1017/S0272263106060049

Nagle, C. L., and Huensch, A. (2020) 'Expanding the Scope of L2 Intelligibility Research: Intelligibility, Comprehensibility, and Accentedness in L2 Spanish', *Journal of Second Language Pronunciation*, 6: 329–51. https://doi.org/10.1075/jslp.20009.nag

Nicora, F., Mcloughlin, L. I., and Gili Fivela B., (2018) 'Impact of prosodic training on Italian as L2 by hiberno–English speakers', in Speech Prosody 9, pp. 970–4, Poznań, 13–16 June 2018. International Speech Communication Association (ISCA).

Nolan, F., and Asu, E. L. (2009) 'The Pairwise Variability index and Coexisting Rhythms in Language', *Phonetica*, 66: 64–77. https://doi.org/10.1159/000208931

Pellegrino, E. (2012) 'The perception of foreign accented speech: segmental and suprasegmental features affecting degree of foreign accent in Italian L2', in H. Mello et al. (eds.) Proceedings of the VIIth GSCP International Conference Speech and Corpora, pp. 261–7, Leipzig, Germany. Firenze: Firenze University Press.

Pellegrino, E., Schwab, S., and Dellwo, V. (2021) 'Native Listeners Rely on Rhythmic Cues When Deciding on the Nativeness of Speech', *The Journal of the Acoustical Society of America*, 150: 2836–53. https://doi.org/10.1121/10.0006537

Pellegrino, E. and Vigliano, D. (2015) 'Self-imitation in prosody training: a study on Japanese learners of Italian', in Proceedings SLaTE 2015. Sixth Workshop on Speech and Language Technology in Education, pp. 53–7. ISCA Special Interest Group SLaTE.

Potisuk, S., Gandour, J., and Harper, M. P. (1996) 'Acoustic Correlates of Stress in Thai', *Phonetica*, 53: 200–20. https://doi.org/10.1159/000262201

Primativo, S. et al. (2013) 'Bilingual Vocabulary Size and Lexical Reading in Italian', *Acta Psychologica*, 144: 554–62. https://doi.org/10.1016/j.actpsy.2013.09.011

R Core Team. (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/, accessed 18 Aug. 2023.

Reid, K. T., Trofimovich, P., and O'Brien, M. G. (2019) 'Social Attitudes and Speech Ratings: Effects of Positive and Negative Bias on Multiage Listeners' Judgments of Second Language Speech', *Studies in Second Language Acquisition*, 41: 419–42. https://doi.org/10.1017/S0272263118000244

Saito, K. (2023) 'How Does Having a Good ear Promote Successful Second Language Speech Acquisition in Adulthood? Introducing Auditory Precision Hypothesis-L2', *Language Teaching*, 56: 522–38. https://doi.org/10.1017/S0261444822000453

Sluijter, A. M., and Van Heuven, V. J. (1996) 'Spectral Balance as an Acoustic Correlate of Linguistic Stress', *The Journal of the Acoustical Society of America*, 100: 2471–85. https://doi.org/10.1121/1.417955

Spinelli, G., Forti, L., and Jared, D. (2021) 'Learning to Assign Stress in a Second Language: The Role of Second-Language Vocabulary Size and Transfer from the Native Language in Second-Language Readers of Italian', *Bilingualism: Language and Cognition*, 24: 124–36. https://doi.org/10.1017/S1366728920000243

Sulpizio, S., and Colombo, L. (2013) 'Lexical Stress, Frequency, and Stress Neighbourhood Effects in the Early Stages of Italian Reading Development', *The Quarterly Journal of Experimental Psychology*, 66: 2073–84. https://doi.org/10.1080/17470218.2013.785577

Sulpizio, S., and McQueen, J. M. (2012) 'Italians use Abstract Knowledge About Lexical Stress During Spoken-Word Recognition', *Journal of Memory and Language*, 66: 177–93. https://doi.org/10.1016/j.jml.2011.08.001

Tagliapietra, L., and McQueen, J. M. (2010) 'What and Where in Speech Recognition: Geminates and Singletons in Spoken Italian', *Journal of Memory and Language*, 63: 306–23. https://doi.org/10.1016/j.jml.2010.05.001

Tagliapietra, L. and Tabossi, P. (2005) 'Lexical stress effects in Italian spoken word recognition', in Proceedings of the XXVII Annual Conference of the Cognitive Science Society, pp. 2140–4. Stresa, Italy: Lawrence Erlbaum.

Thornton, A. M., Iacobini, C., and Burani, C. (1997) *BDVBD: Una Base di Dati sul Vocabolario di Base Della lingua Italiana*. Roma: Bulzoni.

Tomar, S. (2006) 'Converting Video Formats with FFmpeg', *Linux Journal*, 2006: 10. https://doi.org/10.5555/1134782.1134792

Uguzzoni, A. et al. (2003). 'Short vs. Long and/or abruptly cut vowels. New perspectives on a debated question', in Proceedings of the International Congress of Phonetic Sciences, pp. 2717–20, Prague, Czech Republic. International Phonetic Association.

Ushey, K., Allaire, J., and Tang, Y. *reticulate: Interface to 'Python'*. R package version 1.26. https://rstudio.github.io/reticulate/. 2022.

Wiener, S. et al. (2022) 'Acquisition of non-Sibilant Anterior English Fricatives by Adult Second Language Learners', *Journal of Second Language Pronunciation*, 8: 68–94. https://doi.org/10.1075/jslp.20067.wie

Wiener, S., Chan, M. K. M., and Ito, K. (2020) 'Do Explicit Instruction and High Variability Phonetic Training Improve non-Native Speakers' Mandarin Tone Productions?', *The Modern Language Journal*, 104: 152–68. https://doi.org/10.1111/modl.12619

Xie, X., Jaeger, T. F., and Kurumada, C. (2023) 'What we do (not) Know About the Mechanisms Underlying Adaptive Speech Perception: A Computational Framework and Review', *Cortex*, 166: 377–424. https://doi.org/10.1016/j.cortex.2023.05.003

Zhang, C. et al. (2021) 'Comparing Acoustic Analyses of Speech Data Collected Remotely', *The Journal of the Acoustical Society of America*, 149: 3910–6. https://doi.org/10.1121/10.0005132