COMPARATIVE ACOUSTIC ANALYSIS OF ENGLISH YES/NO QUESTION INTONATION, DURATION, AND INTENSITY FOR A NATIVE BRITISH ENGLISH SPEAKER AND A JAVANESE SPEAKER USING PRAAT

Naza Fanisa[1], Tengku Syarfina[2]
Department of English, Universitas Sumatera Utara
e-mail : naza.fanisa8@gmail.com

Abstract

Prosodic features such as intonation, duration, and intensity play a crucial role in spoken communication and are often influenced by a speaker's first language when using a second language. This study aims to acoustically compare the production of an English yes/no question by a native British English speaker and a Javanese speaker to examine potential prosodic transfer effects. The data consisted of one controlled utterance, *"Do you like noodles?"*, produced by two female speakers (n = 2), both aged 25, to minimize biological variation in pitch range and voice quality. The recordings were elicited in a quiet environment and analyzed using Praat. The analysis focused on three acoustic parameters: fundamental frequency (F0) to represent intonation, intensity (in decibels), and total utterance duration (in seconds). Acoustic measurements were taken at the beginning and end of the utterance, and the results were compared descriptively. The findings show that the native English speaker produced a stable rising intonation pattern with increased intensity toward the end of the utterance and a shorter duration, which aligns with typical English yes/no question prosody. In contrast, the Javanese speaker exhibited a wider pitch range, decreasing intensity, and a slightly longer duration, suggesting the influence of Javanese prosodic patterns on English speech production. Although no statistical testing was conducted due to the limited sample size, the acoustic comparison provides preliminary insights into individual level prosodic variation. Pedagogically, the study highlights the usefulness of Praat as a visual and analytical tool to raise learners' awareness of English intonation patterns and prosodic differences in second language pronunciation.

Keywords: *Intonation, Duration, Intensity, Javanese Speaker, Native English, English Prosody, Praat, Phonology*

INTRODUCTION

Spoken communication relies not only on the accurate articulation of segmental units but also on suprasegmental or prosodic features, including intonation, duration, and intensity. These features contribute crucially to grammatical meaning, discourse structure, and pragmatic interpretation (Cruttenden, 1997; Roach, 2009). In second language (L2) speech production, prosodic patterns are particularly susceptible to first language (L1) influence, often resulting in non native rhythmic and intonational realizations even when segmental accuracy is relatively high.

English is commonly described as a stress timed language, characterized by salient contrasts between stressed and unstressed syllables and by rhythmic compression of unstressed material (Roach, 2009; Wells, 2006). In contrast, many Austronesian languages, including Indonesian and Javanese, have been described as exhibiting more syllable timed rhythmic tendencies, in which syllables are produced with relatively even timing (Ladefoged, 2006). However, recent cross linguistic research cautions against rigid rhythmic typologies and emphasizes that rhythm should be understood as gradient rather than categorical (Jun, 2005). Accordingly, this study adopts a cautious position, treating Javanese prosody as *tending toward* syllable based timing rather than as strictly syllable timed.

Prosodic differences are particularly salient in interrogative structures. In English, yes/no questions are typically marked by a final rising intonation, which functions as a primary phonological cue signaling interrogativity (Wells, 2006). Failure to produce this rising contour may affect intelligibility or pragmatic interpretation, especially in conversational contexts. For learners whose L1 does not employ similar intonational marking, producing English question intonation can therefore be challenging.

Previous studies on English prosody produced by Indonesian and Javanese speakers have largely focused on classroom based observations or group level analyses (Mawarni et al., 2024; Ardini & Senowarsito, 2023). While these studies provide valuable insights, fewer investigations offer fine grained acoustic comparisons at the individual speaker level using controlled elicitation and detailed phonetic measurements. Moreover, studies that simultaneously examine multiple prosodic correlates such as pitch (F0), intensity, and duration within a single analysis remain limited.

Recent research has increasingly emphasized the importance of acoustic phonetic analysis in examining second language (L2) prosody. Studies using speech analysis software such as Praat demonstrate that L2 learners often display measurable differences in pitch movement, temporal organization, and intensity patterns when producing English intonation (Guo, 2022; Li & Post, 2022; Zhang & Levis, 2023). These acoustic differences are frequently associated with cross linguistic influence, where prosodic habits from a speaker's first language shape the realization of intonation in the second language (Huang, 2023; Garcia, 2023). Recent phonetic research also highlights that prosodic transfer may appear in subtle variations in pitch range, duration, and loudness distribution across utterances (Tanaka, 2023; Xu, 2024). Therefore, controlled acoustic analysis using tools such as Praat provides an effective approach to documenting individual variation in L2 prosody and examining how speakers from different linguistic backgrounds realize English question intonation (Lee, 2022; Pellegrino, 2024; Park, 2025).

To address this gap, the present study conducts a focused acoustic comparison of an English yes/no question produced by a native British English speaker and a Javanese speaker using Praat. The analysis operationalizes prosody through three measurable acoustic parameters: fundamental frequency (F0) as an index of intonation, intensity (in decibels) as an indicator of loudness variation, and total utterance duration (in seconds) as a temporal correlate of rhythm. Although the study employs a single elicitation sentence, this design choice allows for controlled comparison across speakers by minimizing lexical, syntactic, and contextual variability. At the same time, the study acknowledges this limitation and treats the findings as exploratory rather than generalizable.

By combining acoustic visualization with descriptive analysis, this research aims to illustrate how L1 prosodic patterns may influence English question intonation at the individual level. Pedagogically, the study demonstrates the potential of Praat as a tool for raising learners' awareness of prosodic features that are often difficult to perceive through auditory input alone.

METHOD

This study involved two female participants (n = 2), both aged 25. One participant was a native speaker of British English, while the other was a native speaker of Javanese from Medan, Indonesia, who uses Indonesian and Javanese in daily communication. The participants were selected using convenience sampling to allow for a controlled, individual level acoustic comparison. The small sample size reflects the exploratory and descriptive nature of

the study and does not aim to support statistical generalization. Both participants voluntarily agreed to take part in the study and provided informed consent prior to recording. The recordings were conducted for academic research purposes only, and the study followed standard ethical guidelines for linguistic research involving human subjects.

Participants were instructed to read and produce the sentence naturally, as if asking a question in casual conversation. No explicit instruction regarding intonation or emphasis was given in order to elicit spontaneous prosodic patterns. Each participant produced one token of the sentence, which was used for detailed acoustic analysis. The study acknowledges that the use of a single elicitation item limits phonetic variability and generalizability; therefore, the findings are interpreted as exploratory.

The recordings were conducted in a quiet indoor environment to minimize background noise. Speech data were recorded using a digital recording device with an external microphone placed at a consistent distance from the speaker's mouth. The recordings were saved in WAV format with a sampling rate of 44.1 kHz and 16 bit resolution. These settings ensured sufficient acoustic quality for pitch, intensity, and duration analysis. All recordings were analyzed using Praat (Boersma & Weenink, 2023). Three acoustic parameters were examined: intonation, intensity, and duration. Intonation was operationalized as fundamental frequency (F0, in Hz), intensity as vocal loudness (in decibels), and duration as the total length of the utterance (in seconds).

Pitch (F0) analysis was conducted using Praat's default pitch settings for female speakers. F0 values were measured at the beginning and end of the utterance to capture overall pitch movement and final rising contours characteristic of English yes/no questions. Intensity values were also measured at the beginning and end of the utterance using Praat's intensity display. Duration was calculated based on the total time span of the utterance, as indicated by the waveform. Each utterance was segmented using Praat TextGrid annotations. The segmentation was conducted manually by the author based on visual inspection of the waveform and spectrogram. Segments were labeled using the International Phonetic Alphabet (IPA) to represent the phonetic structure of the utterance. Due to the small dataset and descriptive aim of the study, inter rater reliability measures were not applied; this is acknowledged as a methodological limitation.

The acoustic measurements were analyzed descriptively and compared across the two speakers. No statistical tests, effect sizes, or normalization procedures were applied due to the limited sample size and exploratory focus of the study. The results are therefore interpreted as indicative patterns of individual prosodic variation rather than statistically generalizable findings.

FINDINGS

This section presents the acoustic findings of the study using a descriptive case comparison approach. Given the limited dataset (one utterance produced by each speaker), the results are reported as individual acoustic measurements rather than statistically generalizable outcomes. The findings focus on three prosodic parameters: intonation (fundamental frequency/F0), intensity, and duration. Numerical values are summarized in Table 1, while figures provide visual representations of the acoustic contours. Table 1 presents a summary of the acoustic measurements for both speakers, including duration, F0 values at the beginning and end of the utterance, and intensity values at the beginning and end of the utterance.

Table 1. Acoustic Measurements of the Native English Speaker and the Javanese Speaker

Feature	Native English Speaker	Javanese Speaker
Duration	1.347460 seconds	1.406054 seconds
Pitch (F0)	Start: 243.4 Hz → End: 250.1 Hz	Start: 229.7 Hz → End: 282.0 Hz
Intensity	Start: 55.02 dB → End: 60.07 dB	Start: 56.08 dB → End: 54.37 dB

As shown in Table 1, the native English speaker produced a shorter overall duration, a relatively narrow rising pitch movement, and increasing intensity toward the end of the utterance. In contrast, the Javanese speaker exhibited a longer duration, a wider pitch range, and decreasing intensity across the utterance. These values are reported as descriptive indicators of individual prosodic patterns rather than as population level statistics.

The native English speaker produced a gradual rise in fundamental frequency from 243.4 Hz at the beginning of the utterance to 250.1 Hz at the end, reflecting a smooth rising intonation typical of English yes/no questions. Figure 1 shows the waveform and TextGrid of the native English speaker's utterance, illustrating the segmentation used for pitch and duration analysis.

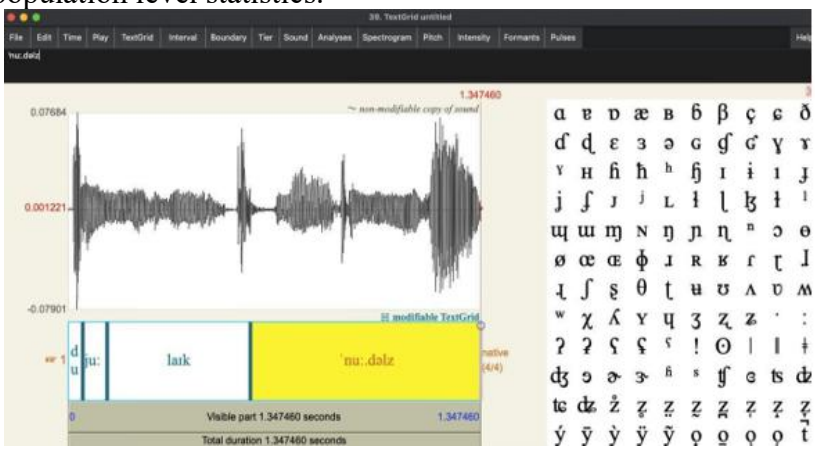

Figure 1. Waveform and TextGrid of the native English speaker

Figure 2 illustrates the pitch (F0) contour of the native English speaker, showing a controlled and gradual rise toward the final syllable.

Figure 2. Pitch contour (F0) of the native English speaker

In contrast, the Javanese speaker demonstrated a wider pitch movement across the utterance. As summarized in Table 1, the speaker's fundamental frequency increased from 229.7 Hz at the beginning to 282.0 Hz at the end.Figure 3 presents the waveform and intensity display of the Javanese speaker's utterance, providing an overview of the acoustic signal.

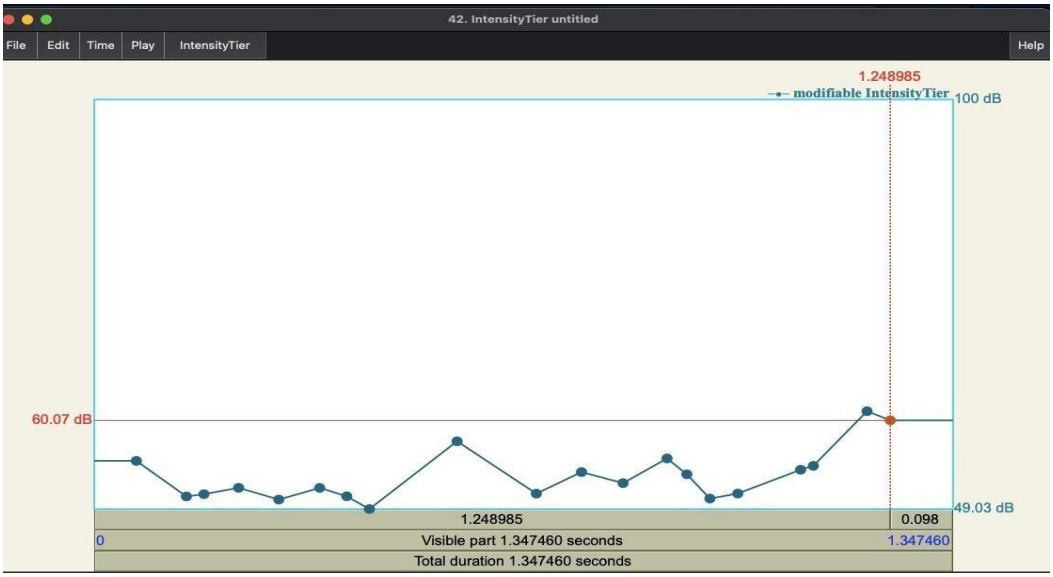

Figure 3. Intensity of the Javanese speaker

Figure 4 shows the TextGrid annotation of the Javanese speaker's utterance, including IPA based segmentation used for acoustic measurement.

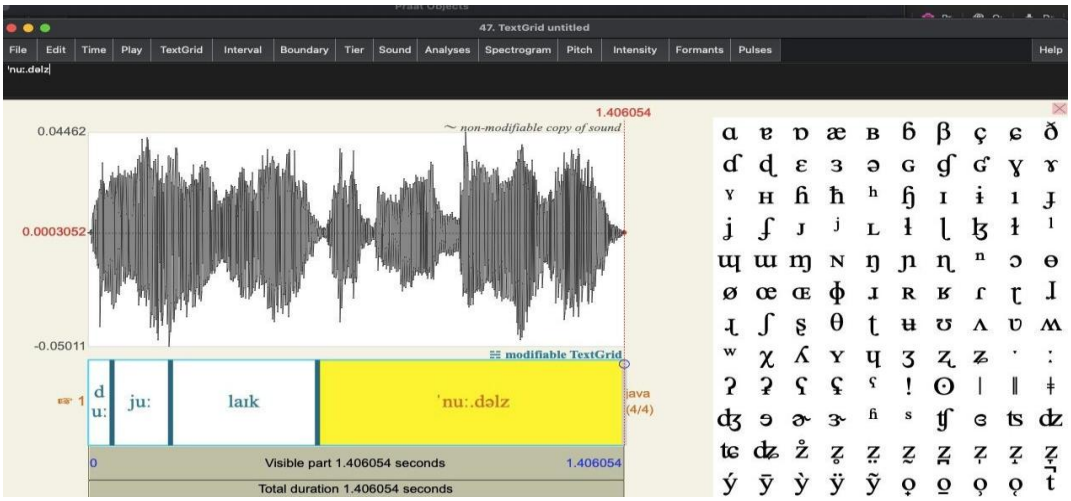

Figure 4. TextGrid segmentation of the Javanese speaker's utterance

The image provides a detailed view of a TextGrid generated using the Praat application, representing the acoustic analysis of the Javanese speaker's production of the English yes/no question, "Do you like noodles?". The TextGrid displays the segmentation of the utterance, which has a total measured duration of 1.406054 seconds. Crucially, the segments are labeled using the International Phonetic Alphabet (IPA), showing the linguistic breakdown of the utterance as: d, ju:, laik, and 'nu:dəlz. This visual segmentation, alongside the corresponding waveform and the adjacent IPA character palette, is a foundational step in the acoustic analysis, allowing researchers to precisely measure and compare prosodic features like duration and pitch across speakers

Figure 5 illustrates the pitch (F0) contour of the Javanese speaker, revealing a steeper and wider pitch rise compared to the native English speaker.

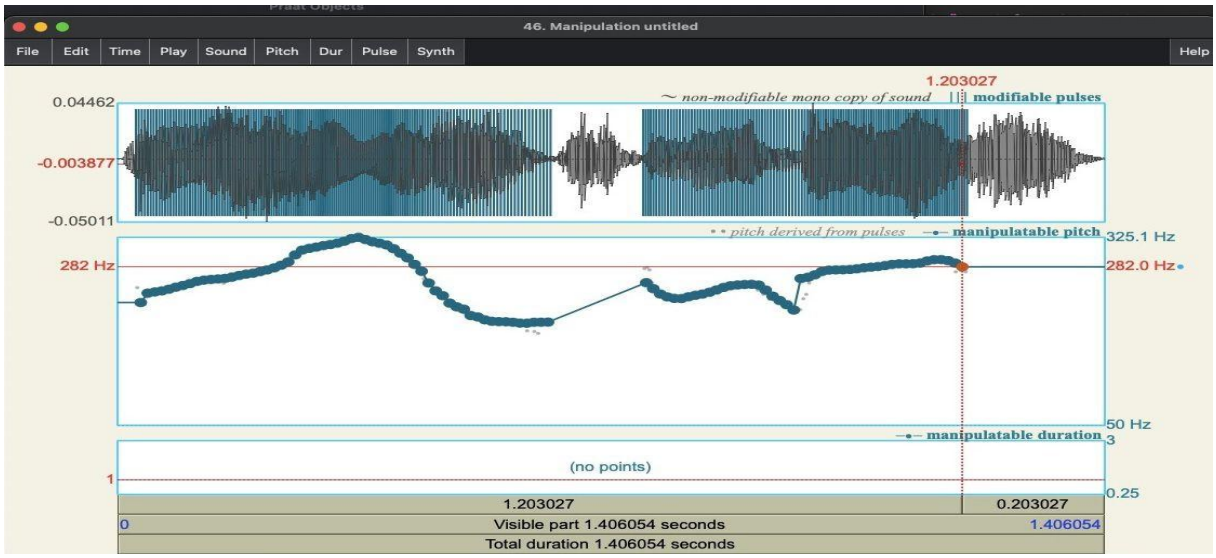

Figure 5. Pitch contour (F0) of the Javanese speaker

The figure above visualizes the acoustic manipulation and analysis of a Javanese speaker's utterance, as processed using the Praat software. Specifically, this analysis focuses on the fundamental frequency (F0) contour, a key acoustic correlate of pitch. The manipulation and subsequent measurement reveal a starting fundamental frequency of 229.8 Hz, which then ascends to a peak value of 282.0 Hz. This observed frequency range provides crucial insight into the speaker's prosodic features (e.g., intonation or emphasis) or the specific phonetic characteristics of the analyzed segment, which is highly relevant for studies concerning language specific tonal or pitch realization in Javanese phonetics. Intensity measurements further differentiate the two speakers' prosodic patterns. The native English speaker's intensity increased from 55.02 dB at the beginning of the utterance to 60.07 dB at the end, indicating greater loudness toward the final portion of the question. In contrast, the Javanese speaker showed a decrease in intensity across the utterance, from 56.08 dB at the beginning to 54.37 dB at the end, as summarized in Table 1.

Figure 6 displays the intensity contour of the Javanese speaker, showing a gradual reduction in loudness across the utterance.

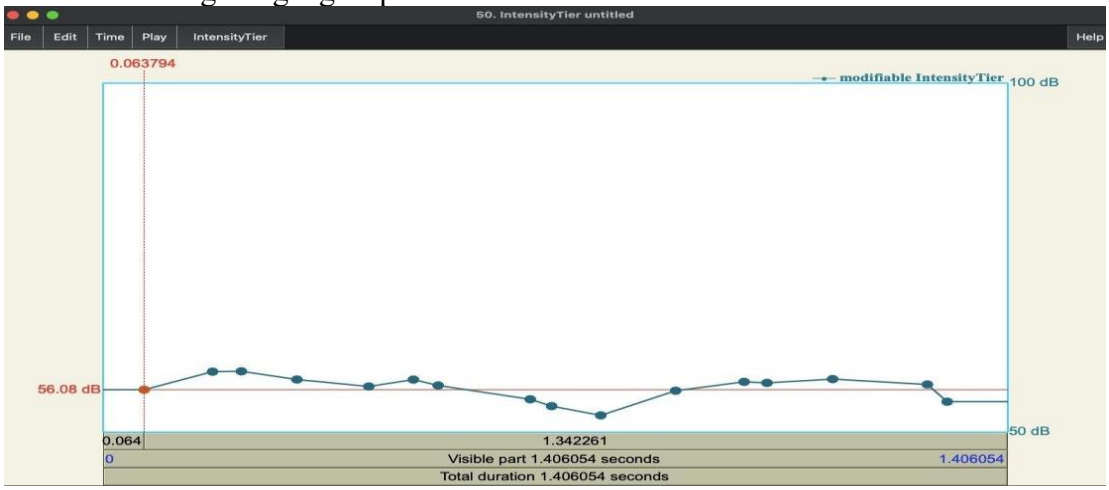

Figure 6. Intensity contour of the Javanese speaker

This figure presents the results of an acoustic analysis focusing on the intensity contour of a Javanese speaker, conducted using the Praat software. The analysis specifically tracked the dynamic range of vocal loudness across the measured utterance segment. The data reveals that the intensity started at 56.08 dB (decibels) and subsequently decreased to 54.37 dB. This specific decrease in intensity is a crucial finding, as the intensity contour contributes significantly to the speaker's stress and emphasis patterns. This measurable variation helps characterize the phonetic and prosodic features unique to the Javanese dialect being studied. With respect to temporal organization, the native English speaker produced the utterance in 1.347460 seconds, while the Javanese speaker produced the utterance in 1.406054 seconds. Although the difference in duration is relatively small, it indicates a slightly longer temporal realization by the Javanese speaker. Given the single token design, this difference is reported descriptively and interpreted cautiously. Overall, the findings reveal consistent differences across the three examined prosodic parameters. The native English speaker displayed a shorter duration, a controlled rising intonation, and increasing intensity toward the end of the utterance. The Javanese speaker, by contrast, exhibited a wider pitch range, decreasing intensity, and slightly longer duration. These results are presented as descriptive case comparisons and form the basis for the discussion of first language prosodic influence in the following section.

DISCUSSION

The present study examined prosodic differences in the production of an English yes/no question by a native British English speaker and a Javanese speaker through acoustic analysis of pitch (F0), intensity, and duration. The discussion below interprets these findings cautiously, acknowledging the exploratory nature of the study and its methodological limitations. Overall, the native English speaker exhibited a relatively narrow and controlled rising pitch contour toward the end of the utterance, increased intensity in the final portion, and a shorter overall duration. These characteristics are consistent with well documented descriptions of English yes/no question prosody, in which interrogativity is signaled primarily through a final rising intonation and reinforced by stress related pitch and loudness cues (Wells, 2006; Roach, 2009). The observed alignment of pitch movement and intensity increase toward the utterance final region reflects typical stress timed rhythmic organization, where prominence is concentrated on stressed syllables.

The differences observed between the native English speaker and the Javanese speaker are consistent with recent findings in second language phonetics research. Previous studies report that L2 speakers often demonstrate different pitch trajectories and prosodic timing patterns compared with native speakers due to cross linguistic influence and differences in speech rhythm (Guo, 2022; Zhang & Levis, 2023; Huang, 2023). Similar variations in pitch range, duration, and intensity have also been documented in acoustic analyses of English interrogative intonation produced by non native speakers (Li & Post, 2022; Xu, 2024). Furthermore, research on prosodic training indicates that learners can gradually approximate native like prosody when provided with acoustic feedback and visual representations of pitch contours (Pellegrino, 2024). These findings support the interpretation that the prosodic differences identified in the present study may reflect individual realizations influenced by first language prosodic patterns.

In contrast, the Javanese speaker produced a wider pitch range with a steeper overall rise, decreasing intensity toward the end of the utterance, and a slightly longer duration. While these patterns may suggest the influence of first language prosodic habits, the present findings should not be

interpreted as representative of Javanese prosody as a whole. Rather, they reflect individual level production patterns observed in a highly constrained dataset. The wider pitch excursion may be attributable not only to L1 prosodic transfer but also to individual speaking style, expressive strategy, or recording context.

From a rhythmic perspective, the Javanese speaker's longer duration and reduced intensity modulation may be tentatively related to descriptions of Javanese as exhibiting more syllable based timing tendencies (Ladefoged, 2006). In such systems, temporal organization and pitch movement are less tightly coupled to stress placement than in stress timed languages like English. However, following contemporary views of rhythm as a gradient rather than categorical phenomenon (Jun, 2005), the present study does not claim a strict syllable timed classification. Instead, the results suggest a possible tendency toward more evenly distributed prosodic cues in this individual speaker's English production.

Importantly, alternative explanations must be considered. The observed differences may reflect intra speaker variability, the influence of reading style, or the absence of pragmatic context beyond sentence reading. Additionally, the use of a single token per speaker limits the ability to distinguish stable prosodic patterns from momentary performance effects. As such, the findings should be interpreted as illustrative rather than confirmatory, providing insight into how prosodic features may vary across speakers with different linguistic backgrounds under controlled conditions. With respect to prosodic transfer, the results align with previous research suggesting that L2 speakers may rely on L1 prosodic strategies when producing L2 intonation patterns (Guo, 2022; Mawarni et al., 2024). However, given the extremely limited sample size, the present study does not claim evidence of systematic transfer at the language level. Instead, it demonstrates how acoustic analysis can reveal subtle individual differences in pitch, intensity, and duration that may reflect cross linguistic influence.

Finally, it is important to distinguish between production oriented findings and pedagogical implications. The discussion above focuses on describing and interpreting acoustic patterns in speech production. Pedagogical considerations such as the use of Praat to raise learners' awareness of English intonation are addressed separately in the following section to avoid conflating descriptive phonetic analysis with instructional recommendations. Future research should expand this line of inquiry by including larger and more diverse participant samples, multiple elicitation items, and repeated tokens per speaker. Such designs would allow for statistical analysis, assessment of variability, and stronger claims regarding cross linguistic prosodic influence. Despite its limitations, the present study contributes a preliminary, methodologically transparent case comparison that illustrates the potential of acoustic phonetic analysis in exploring L2 prosody.

CONCLUSION

This study provided a preliminary acoustic comparison of an English yes/no question produced by a native British English speaker and a Javanese speaker, focusing on three prosodic parameters: intonation (fundamental frequency/F0), intensity, and duration. The findings showed that the native English speaker produced a relatively controlled rising intonation pattern, increased intensity toward the end of the utterance, and a shorter overall duration. In contrast, the Javanese speaker exhibited a wider pitch range, decreasing intensity, and a slightly longer duration. These results suggest potential individual level differences in prosodic realization that may be influenced by first language prosodic patterns. However, the conclusions of this study must be interpreted with caution. The

extremely limited sample size (n = 2), the use of a single elicitation sentence, and the reliance on one token per speaker constrain the scope of inference and preclude statistical generalization. Accordingly, the findings should be viewed as exploratory case comparisons rather than evidence of systematic prosodic transfer at the language or population level. From a phonetics research perspective, this study demonstrates how fine grained acoustic analysis using Praat can reveal subtle differences in pitch movement, intensity distribution, and temporal organization between speakers with different linguistic backgrounds. The study highlights the value of combining waveform visualization, TextGrid annotation, and acoustic measurements to document individual variation in L2 speech production under controlled conditions.

In terms of pedagogical implications, while the study does not provide direct empirical evidence of learning outcomes, it suggests that Praat may serve as a useful exploratory and awareness-raising tool in pronunciation instruction. Visual representations of pitch contours and intensity patterns can help learners notice prosodic features that are often difficult to perceive through auditory input alone. Nevertheless, claims about pedagogical effectiveness should be validated through classroom-based studies involving larger learner groups and instructional interventions. Future research should extend the present work by including larger and more diverse participant samples, multiple elicitation sentences, repeated tokens per speaker, and cross-dialectal or cross-linguistic comparisons. Such designs would enable statistical analysis, assessment of variability, and stronger conclusions regarding the role of first language prosody in second language intonation production. Despite its limitations, the present study contributes a transparent and methodologically grounded illustration of how acoustic phonetic analysis can be applied to the study of L2 prosody. Future research should involve larger datasets, multiple elicitation items, and repeated tokens per speaker in order to allow statistical analysis and stronger generalization regarding cross linguistic prosodic influence.

REFERENCES

Ardini, S. N., & Senowarsito, S. (2023). Prosodic features of Jonglish community: An effort toward glocalization. *Indonesian Journal of EFL and Linguistics, 8*(2), 379–394.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3) [Computer software]. University of Amsterdam. https://www.praat.org

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge University Press.

Garcia, M. (2023). Prosodic variation in second language speech production. *Applied Linguistics.*

Guo, X. (2022). Acoustic correlates of English lexical stress produced by Chinese dialect speakers compared to native English speakers. *Frontiers in Psychology,* https://doi.org/10.3389/fpsyg.2022.796252

Huang, B. (2023). Acoustic analysis of English intonation patterns in L2 speech. *Speech Communication.*

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*(pp. 430–458). Oxford University Press.

Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Thomson Wadsworth.

Lee, J. (2022). Pitch movement and prosodic transfer in L2 English. *Studies in Second Language Acquisition.*

Li, A., & Post, B. (2022). Intonation in second language speech: Cross-linguistic perspectives. *Journal of Phonetics.*

Mawarni, N., Syarfina, T., Zulhantiar, P., & Rangkuti, R. (2024). Analysis of Javanese language prosody: An acoustic phonetic study. *Fonologi: Jurnal Ilmuan Bahasa dan Sastra Inggris, 2*(4), 63–75.

Park, S. (2025). Acoustic realization of English interrogative intonation by Asian learners. *Journal of Phonetics.*

Pellegrino, E. (2024). After self-imitation prosodic training L2 learners converge prosodically to native speakers. *Languages, 9*(1), 33. https://doi.org/10.3390/languages9010033

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Tanaka, H. (2023). Intonation and discourse meaning in L2 speech. *Applied Linguistics Review.*

Wells, J. C. (2006). *English intonation: An introduction.* Cambridge University Press.

Xu, Y. (2024). Speech melody and prosody in second language acquisition. *Phonetica.*

Zhang, Y., & Levis, J. (2023). Prosodic transfer in second language pronunciation. *Language Teaching Research.*