CHAPTER III
METHODOLOGY

3.1 Research Design

This study employs a *quantitative descriptive design* with an exploratory orientation. Creswell and Creswell (2023) describe quantitative research as an approach for testing objective theories by examining the relationship among variables that can be measured numerically and analyzed using statistical procedures. A descriptive design, specifically, is used when the purpose of a study is to systematically describe the characteristics of a phenomenon as it naturally occurs, rather than to test a hypothesis about cause-and-effect relationships between variables (Creswell & Creswell, 2023.) This design was selected because the primary data of this study consist of numerical acoustic measurements—fundamental frequency (F0), pitch range, and intensity—extracted from students' speech recordings using Praat. These measurements are quantitative by nature. Rather than testing a hypothesis or comparing groups statistically, this study describes the numerical patterns of these acoustic features across four emotional expressions (joy, sadness, calmness, and anxiety), consistent with the purpose of a descriptive quantitative study. The exploratory orientation reflects the fact that emotional intonation in EFL students' speech, examined through instrumental acoustic analysis, remains a comparatively underexplored area, as established in the research gap in Chapter II (2.9).

It is worth noting explicitly that this represents a refinement from a purely qualitative descriptive approach: because the core data (Hz values, dB values, and contour measurements produced by Praat) are numerical, a quantitative descriptive framework more accurately represents both the nature of the data and the analytical procedures used in Chapter IV, where acoustic

values are compared and summarized numerically (e.g., through mean pitch values and pitch range tables) across emotion categories.

Several alternative designs were considered and set aside before settling on a quantitative descriptive approach. A purely qualitative case-study design, relying on impressionistic listener judgments of a small number of recordings, was considered but rejected because it would not have permitted the kind of systematic, numerical comparison across emotion categories that this study's research questions require; impressionistic judgment is also more vulnerable to rater subjectivity than direct acoustic measurement. An experimental or quasi-experimental design—for example, comparing the same students' emotional intonation before and after a targeted instructional intervention—was also considered, since this would align well with the pedagogical implications discussed in Chapter II. This design was set aside for the present study, however, because establishing a reliable descriptive baseline of how EFL students currently realize emotional intonation acoustically is a logically prior step to designing and evaluating any instructional intervention; without first knowing what patterns currently exist and which emotions are hardest to differentiate acoustically, an intervention study would risk targeting the wrong emotions or acoustic features. A correlational design, relating acoustic measures to an external variable such as overall speaking proficiency scores, was likewise considered but set aside, since the research questions guiding this study concern how emotion is acoustically realized in itself, rather than what other variable it might statistically predict. The quantitative descriptive design was therefore judged the most appropriate fit between the study's purpose, its available data, and its position as an early, foundational investigation within a still-underexplored area of research (see Chapter II, 2.9).

3.2 Research Settings and Participants

This study was conducted at the English Education Study Program, Faculty of Teacher Training and Education, Universitas Islam Nusantara.

The participants of this study are EFL university students who are currently learning English as a foreign language. The researcher selects approximately 20–25 students using a convenience sampling technique, based on their availability and willingness to participate.

The population of this study consists of EFL university students enrolled in the English Education Study Program who are currently developing their English speaking skills. From this population, the researcher selected a sample of approximately 20–25 students using a *convenience sampling technique*, based on their availability and willingness to participate. Etikan, Musa, and Alkassim (2016) explain that convenience sampling is a non-probability sampling technique in which participants are selected based on their accessibility to the researcher; while it does not allow for statistical generalization to a wider population, it is appropriate for exploratory studies with practical constraints on time and resources, such as the present study. Because this study aims to describe patterns of emotional intonation rather than to generalize findings to the entire EFL student population, convenience sampling is an appropriate and sufficient technique for the scope of this research.

Participants were required to meet several inclusion criteria: (1) currently enrolled as an active student in the English Education Study Program; (2) having completed, or currently enrolled in, a speaking-focused course, to ensure a baseline level of English speaking experience; and (3) willingness to be audio-recorded and to have those recordings used for academic research purposes. Students who reported a diagnosed speech or hearing impairment were excluded, since such conditions could confound the acoustic realization of emotional intonation independently of language-related factors. No restriction was placed on participants' gender, since the acoustic

analysis procedure in this study (see 3.3.3) accounts for the typical pitch-range differences between male and female voices at the measurement stage rather than through participant exclusion.

The recruitment procedure followed several steps. First, the researcher approached the English Education Study Program to identify students who met the inclusion criteria and were currently active in the program. Second, an open invitation was extended to eligible students, explaining the purpose and procedure of the study in general terms. Third, students who expressed willingness to participate were scheduled individually for their recording session, at a time and location convenient to them, in order to maximize both participation and recording quality. This process continued until approximately 25 participants had been recruited, consistent with the sample size typically considered adequate for an exploratory, descriptive acoustic study of this kind.

3.3 Research Instruments

3.3.1 Speaking Task (Emotional Script)

The main instrument is a speaking task consisting of *three semantically neutral, emotion-ambiguous sentences*:

1. “I didn’t expect that.”
2. “What are you doing here?”
3. “I can’t believe this.”

These sentences were deliberately selected because their lexical and grammatical content does not, by itself, indicate any particular emotion—each sentence can plausibly be uttered in a state of joy, sadness, calmness, or anxiety, depending only on how it is delivered. This design isolates intonation as the primary variable under investigation, consistent with the theoretical distinction between the grammatical and attitudinal/emotional functions of

intonation established in Chapter II (Crystal, 1969; Cruttenden, 1997): if the same sentence is held constant while only its prosodic delivery changes, any resulting acoustic differences across recordings can be attributed to the attitudinal/emotional function of intonation rather than to differences in wording.

Each participant performs all three sentences under each of the four target emotions—joy, sadness, calmness, and anxiety—adapted from the emotion categories used by Rodero (2011). To support natural emotional delivery, each sentence–emotion pairing is accompanied by a brief *situational cue* that contextualizes the intended emotional state without dictating specific wording, for example.

Table 3.1 *Situational Cues by Emotion and Sentence*

Emotions	Situational Cues
Joy	You just got good news.
Sadness	You lost something important.
Calmness	Today was just another day.
Anxiety	You're waiting on your test results

Each participant is asked to read or perform these scripts, and their speech is recorded for analysis. This produces 3 sentences × 4 emotions = 12 recordings per participant. With approximately 25 participants, this yields a total of approximately 300 recordings, consistent with the dataset analyzed in Chapter IV.

The three sentences and twelve situational cues were not adopted arbitrarily; they were developed through a short internal review process before being finalized for data collection. An initial pool of six candidate sentences was drafted, each screened against a single criterion: whether the sentence,

read without any specific emotional framing, could plausibly be interpreted as expressing any of the four target emotions. Sentences judged too semantically weighted toward one particular emotion (for example, sentences containing explicitly positive or negative vocabulary) were discarded at this stage, leaving the three sentences retained in the final instrument. The situational cues accompanying each sentence were then drafted to be concrete and easy to visualize, following the reasoning that a vaguely worded cue (e.g., simply "feel happy") would ask participants to generate their own scenario under time pressure, potentially increasing inconsistency in delivery, whereas a concrete, ready-made scenario allows participants to focus their attention on vocal delivery rather than on inventing a context. The cues were reviewed informally with a small number of peers prior to data collection to confirm that each cue was interpretable and unambiguous with respect to its intended emotion.

3.3.2 Audio Recording Tool

A recording device (e.g., a smartphone or an external microphone) is used to capture participants' speech in audio format. Recordings are made in a quiet room to minimize background noise, ensuring that the resulting audio files are suitable for acoustic processing in Praat. Where possible, recordings are captured at a minimum sample rate of 44.1 kHz, which comfortably exceeds the sampling rate required to accurately resolve the fundamental frequency range of human speech (typically 75–500 Hz for adult voices) according to the Nyquist sampling theorem, under which a signal must be sampled at a rate at least twice its highest frequency component to be accurately reconstructed. Recordings are saved in a standard compressed or uncompressed audio format (e.g., .wav or .ogg) compatible with Praat's built-in audio decoding capabilities. Each recording is saved individually per sentence–emotion combination and labeled using a consistent naming

convention (Participant–Emotion–Sentence, e.g., P01_Joy_S1) to support systematic organization during the analysis stage (see 3.5).

Microphone placement and recording distance are kept as consistent as practically possible across participants, since substantial variation in distance from the microphone can influence measured intensity values independently of the participant's actual vocal effort. Where a built-in smartphone microphone is used, participants are asked to hold the device at a similar distance and orientation for each recording, and recordings are conducted in the same physical space where feasible, to minimize environmental sources of acoustic variability that are unrelated to the emotional content of the utterance itself.

3.3.3 Praat Software

Praat (Boersma & Weenink, 2021) is used to conduct the acoustic analysis of participants' speech recordings. The software enables extraction and visualization of the following acoustic parameters, which serve as the quantitative variables of this study:

a) Fundamental Frequency (F0), measured in Hertz (Hz)
b) Pitch range (the difference between maximum and minimum F0 within an utterance)
c) Intonation contour (the shape of pitch movement across an utterance)
d) Intensity, measured in decibels (dB)

To ensure that F0 measurements are extracted consistently and accurately across all 243 recordings, a fixed set of analysis parameters is applied uniformly. The pitch floor and pitch ceiling—the lower and upper bounds within which Praat searches for a fundamental frequency—are set to 75 Hz and 500 Hz respectively, a range wide enough to accommodate typical adult male and female speaking voices without needing separate settings per participant. Using a pitch floor that is too low, or a ceiling that

is too high, increases the risk of octave errors, in which Praat's autocorrelation-based pitch-tracking algorithm mistakenly detects a pitch at double or half the true F0; the 75–500 Hz range is a standard, widely recommended setting in phonetic research for this reason. The analysis window uses Praat's default time step (determined automatically based on the pitch floor, typically resulting in a step of approximately 0.01 seconds), which provides sufficiently fine-grained temporal resolution to capture pitch movement within short utterances such as the three- to six-word sentences used in this study's speaking task.

Table 3.2 *Operational Definitions of Variables*

Variable	Operational Definition	Unit	Measurement Tool
Fundamental Frequency (F0) — Mean	The average pitch of a recording, calculated across its entire voiced duration	Hertz (Hz)	Praat: Pitch → Get mean
Fundamental Frequency (F0) — Minimum	The lowest detected pitch value within a recording	Hertz (Hz)	Praat: Pitch → Get minimum
Fundamental Frequency (F0) — Maximum	The highest detected pitch value within a recording	Hertz (Hz)	Praat: Pitch → Get maximum
Pitch Range	The difference between the maximum and minimum F0 within a recording (Max F0 – Min F0)	Hertz (Hz)	Calculated from Praat pitch extraction
Intonation Contour	The visual shape of pitch movement across the duration of an utterance (e.g., rising, falling, fluctuating)	Descriptive/ visual	Praat: Pitch listing / pitch contour display
Intensity	The average acoustic energy (loudness) of a recording	Decibels (dB)	Praat: Intensity → Get mean
Emotion Category	The target emotion the participant was instructed to convey for a given recording	Categorical (Joy / Sadness / Calmness / Anxiety)	Assigned by task design (see Table 3.1)

3.4 Data Collection Procedure

Data collection follows a four-step procedure applied to every sentence–emotion combination:

1. *Read the situational cue***.** The participant reads the situational cue corresponding to the target emotion and sentence, in order to understand the intended emotional context (e.g., "You just got good news" for the joy condition).
2. *Choose the emotion and speak the sentence naturally***.** The participant delivers the target sentence aloud, attempting to naturally convey the emotion indicated by the cue, rather than reading it in a flat or neutral tone.
3. *Record the voice in a quiet place***.** Each performance is recorded using a recording device in a quiet room to minimize background noise, ensuring that the resulting audio files are suitable for acoustic processing in Praat.
4. *Submit the recording***.** The completed recording is submitted, labeled by participant, sentence, and emotion category, and organized for subsequent acoustic analysis.

This four-step procedure is repeated for all three sentences under all four target emotions, yielding 12 recordings per participant (see 3.3.1).

3.5 Data Analysis Technique

Data obtained from the recordings are analyzed quantitatively using Praat. Each recording is processed individually following a consistent sequence of operations, described in detail below to ensure the procedure is fully replicable.

1. Each recording is opened in Praat, and the corresponding pitch contour is extracted.
2. Numerical values for F0 (mean, minimum, and maximum), pitch range, and intensity are measured for each recording.
3. The extracted values are organized into descriptive tables, showing, for each emotion category, the mean and range of F0, pitch range, and intensity across all participants.
4. The four emotion categories are compared descriptively using these summary values (e.g., comparing the mean F0 of joy recordings against the mean F0 of sadness recordings) to identify patterns consistent with the theoretical expectations established in Chapter II (2.3: Ekman, 1992; Scherer, 2003).
5. Pitch contour shapes are also described qualitatively (e.g., rising, falling, fluctuating) to complement the numerical findings and support interpretation.

This five-steps sequence is repeated identically for every recording in the dataset, ensuring that all 243 usable recordings (see Chapter IV, 4.1) are processed under exactly the same analytical conditions. Once all individual recordings have been processed, the following aggregation and comparison steps are applied:

1. The extracted values are organized into descriptive tables, showing, for each emotion category, the mean and standard deviation of F0, pitch range, and intensity across all participants.
2. The four emotion categories are compared descriptively using these summary values (e.g., comparing the mean F0 of joy recordings against the mean F0 of sadness recordings) to identify patterns consistent with the theoretical expectations established in Chapter II (2.3: Ekman, 1992; Russell, 1980; Scherer, 2003).

3. Pitch contour shapes are also described qualitatively (e.g., rising, falling, fluctuating) to complement the numerical findings and support interpretation, in line with the triangulation approach described in 3.6.

The results are presented in Chapter IV in the form of descriptive statistics (means and standard deviations) supported by pitch contour visualizations generated in Praat, consistent with the quantitative descriptive design adopted in this study.

3.6 Data Validity and Reliability

To ensure the trustworthiness of the data, this study addresses validity and reliability separately, following the distinction commonly drawn in quantitative research between whether an instrument measures what it is intended to measure (validity) and whether it does so consistently (reliability).

a) *Content Validity.* The speaking task's content validity rests on the alignment between its design and the theoretical framework established in Chapter II. The four target emotions (joy, sadness, calmness, anxiety) were selected directly from Ekman's (1992) basic emotions framework and mapped onto Russell's (1980) valence-arousal space (see 2.10), ensuring that the emotions examined are theoretically motivated rather than arbitrarily chosen. The situational cues (Table 3.1) were designed to be concrete and unambiguous with respect to their intended emotion, and were informally reviewed prior to data collection (see 3.3.1) to confirm that each cue reliably suggested its intended emotion rather than a different or ambiguous one.
b) *Construct Validity.* The acoustic parameters measured in this study—F0, pitch range, and intensity—were selected because they correspond directly to the acoustic cues identified in Scherer's (2003) acoustic-encoding model as the primary carriers of emotional information in the

voice (see 2.3). Using Praat to extract these parameters, rather than relying on impressionistic listener judgment, strengthens construct validity by ensuring that "emotional intonation" is operationalized as a directly observable, numerically measurable acoustic pattern rather than as a subjective impression.

c) *Measurement Reliability.* All recordings are analyzed using the same Praat settings and procedures (e.g., pitch floor and ceiling parameters fixed at 75–500 Hz, described in 3.3.3) to ensure that acoustic measurements are obtained consistently across participants and recordings. Using a single, fixed analysis protocol—rather than adjusting settings on a case-by-case basis—removes a potential source of inconsistency that could otherwise arise from analyst judgment calls made differently across different recordings.

d) *Recording Consistency.* Recordings are conducted under similar conditions (e.g., quiet environment, consistent recording distance, as described in 3.3.2) to reduce sources of measurement error unrelated to the participants' emotional delivery. Where a recording is judged unsuitable for reliable analysis (for example, due to excessive background noise interfering with Praat's pitch-tracking algorithm), it is excluded from the dataset rather than analyzed under compromised conditions (see the discussion of recording attrition in Chapter IV, 4.1 and 4.3.4).

e) *Triangulation.* Numerical acoustic measurements from Praat are complemented by descriptive observation of pitch contour shapes, strengthening the credibility of the findings by combining quantitative and observational evidence. Where a numerical summary statistic (e.g., mean F0) might obscure an interesting qualitative pattern in how pitch moves across an utterance, the accompanying visual contour inspection provides a check against over-relying on any single numerical measure in isolation.

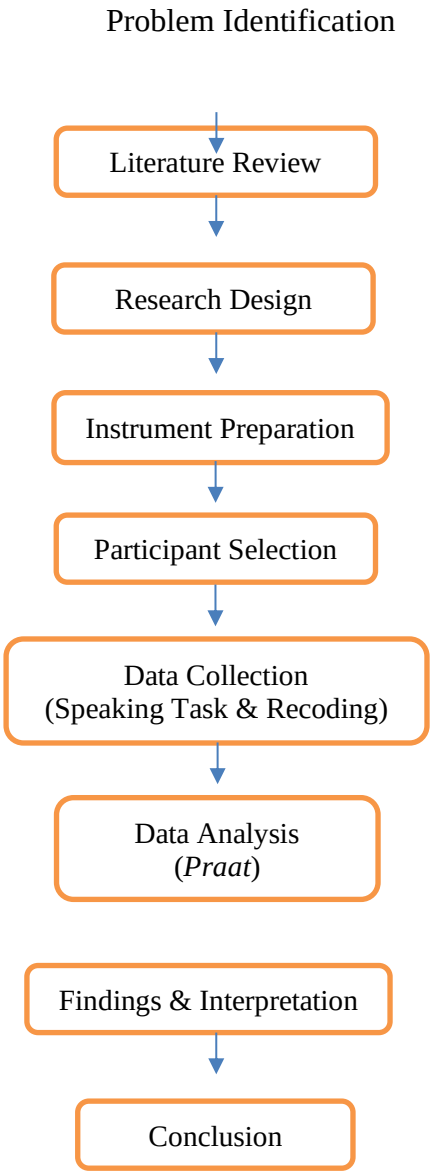
f) *Limitations Acknowledged at the Design Stage*. It is acknowledged that this study does not employ a second, independent rater to cross-verify Praat measurements (a limitation discussed further in Chapter IV, 4.3.4). This decision reflects a practical trade-off appropriate to an undergraduate thesis of this scope: because Praat's pitch- and intensity-extraction algorithms are deterministic (the same recording, analyzed with the same settings, will always yield the same output), the primary source of potential inconsistency in this study lies not in inter-rater disagreement but in ensuring that identical settings are applied uniformly across recordings—a concern directly addressed through the fixed-protocol approach described above.

3.7 Ethical Considerations

This study follows standard ethical research principles:

a) Participants are informed about the purpose of the research, what will be asked of them, and how their recordings will be used, prior to data collection. Consent is obtained verbally or in writing before the recording session begins.
b) Participation is voluntary, and participants may withdraw at any time without penalty, including after their recordings have already been collected, in which case their recordings are deleted and excluded from analysis upon request.
c) Participants' identities are kept confidential; recordings are labeled with anonymized participant codes (e.g., P01, P02) rather than names, and no personally identifying information is included in the dataset, spreadsheet, or final written thesis.
d) Audio recordings are used strictly for academic analysis related to this thesis and are not shared publicly or used for any other purpose without participants' separate, explicit consent.

3.8 Research Procedure Flowchart



3.9 Summary

This chapter has outlined the methodology used to investigate emotional intonation in EFL students' speech production. A quantitative descriptive design was adopted to match the numerical nature of the acoustic data produced by Praat, after considering and setting aside qualitative, experimental, and correlational alternatives (3.1). Approximately 20–25 students were selected through convenience sampling based on defined inclusion criteria (3.2), and data were collected

using a carefully developed emotional speaking task, consisting of three emotion-ambiguous sentences, twelve situational cues, an operationally defined set of acoustic variables, and a pilot-tested procedure, together with an audio recording tool and Praat software configured with fixed analysis settings (3.3). A structured, briefed data collection procedure (3.4) and a fully specified, replicable six-step Praat analysis sequence (3.5) were used to ensure consistency across all 243 usable recordings. Attention to content validity, construct validity, measurement reliability, recording consistency, and triangulation (3.6), together with standard ethical safeguards including a formal recruitment announcement and informed consent procedure (3.7), support the trustworthiness of the resulting data. The overall research procedure is summarized in the flowchart above (3.8). This methodological foundation provides the basis for the results and discussion presented in Chapter IV.