Data in Brief 57 (2024) 111138

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

IndoWaveSentiment: Indonesian audio dataset for emotion classification

Anugrayani Bustamin[a,*], Andi M. Rizky[a], Elly Warni[a], Intan Sari Areni[b], Indrabayu[a]

[a] Informatics Department, Universitas Hasanuddin, Poros Malino Street Km 6, Gowa, South Sulawesi, Indonesia
[b] Electrical Engineering Department, Universitas Hasanuddin, Poros Malino Street Km 6, Gowa, South Sulawesi, Indonesia

ARTICLE INFO

Article history:
Received 3 July 2024
Revised 14 September 2024
Accepted 11 November 2024
Available online 16 November 2024

Dataset link: IndoWaveSentiment: Indonesian Audio Dataset for Emotion Classification (Original data)

Keywords:
Signal processing
Voice classification
Sentiment analysis
Emotion

ABSTRACT

Voice is a one of media for human communication and interaction. Emotions conveyed through voice, such as laughter or tears, can communicate messages more quickly than spoken or written language. In sentiment analysis, the emotional component is crucial for reflecting human perceptions and opinions. This paper introduces IndoWaveSentiment, a dataset of emotional voice recordings categorized into five classes: neutral, happy, surprised, disgusted, and disappointed. The data collection took place in a recording studio with ten actors, evenly split between men and women. Each actor repeated the same sentence in Bahasa Indonesia three times for each emotion class, and the recordings were saved in .wav format. The annotation process was manually conducted using Audacity and validated through a questionnaire-based sampling technique that supports audio data. This dataset is valuable for researchers in Signal Processing and Artificial Intelligence, aiding the development of classification models within Machine Learning.

* Corresponding author.
E-mail address: anugrayani@unhas.ac.id (A. Bustamin).
Social media: @uyyayauya (A. Bustamin)

https://doi.org/10.1016/j.dib.2024.111138

Specifications Table

Subject	Artificial Intelligence, Multimedia, Signal Processing.
Specific subject area	Dataset for signal processing, emotion and sentiment classification
Type of data	Raw with *.wav format, Analyzed, Annotate, Processed.
Data collection	The dataset collection was conducted using Cardioid Dynamic Vocal Microphone and Studio Monitoring Headphone in the recording studio room. The collected mono channel data will be visualized through Audacity which will then be annotated based on its emotional class.
Data source location	Prambors Radio Makassar and Engineering Faculty, Universitas Hasanuddin, Gowa, South Sulawesi, Indonesia.
Data accessibility	Repository name: https://data.mendeley.com IndoWaveSentiment: Indonesian Audio Dataset for Emotion Classification [1] Data identification number: 10.17632/j9ytfdzy27/1 Direct URL to data: https://data.mendeley.com/datasets/j9ytfdzy27/1
Related research article	

1. Value of the Data

- The IndoWaveSentiment Dataset consists of Indonesian language voice emotion data, publicly accessible and useful for developing audio sentiment analysis.
- Recording took place in controlled conditions, including high-quality rooms and equipment.
- The dataset includes ten respondents, equally divided between five male and five female. Each respondent uttered a sentence with 5 different emotional variations—neutral, happy, surprised, disgusted, and disappointed—at both normal and strong intensities, repeated three times. The selected sentence is lexical, meaning it conveys standalone information.

2. Background

Understanding human emotions is crucial for sentiment analysis, particularly in interpersonal communication [2]. Emotion can be measured by valence (positivity or negativity) and arousal (intensity). In detecting emotions from human communication, text alone is often insufficient due to the impact of sarcasm [3]. Therefore, incorporating speech data enhances the analysis of how emotions are expressed in different emotional contexts. Several studies have focused on comparing various Speech Emotion Recognition (SER) methods that analyze audio signals to identify emotions, utilizing existing speech emotion datasets such as the Berlin Emotional Database (EmoDB) and Surrey Audio-Visual Expressed Emotion (SAVEE) [3,4]. Voice emotion recognition technology is essential in various domains, including human-computer interaction, customer service, and, in this study, product review video analysis. Emotion-based sentiment analysis can enhance the interpretation of user reviews, addressing the limitations of text data in conveying emotions [5,6]. To date, research on Speech Emotion Recognition (SER) in Indonesian is sparse, primarily due to the lack of Indonesian-language corpora for Voice Emotion Recognition [7]. To address this gap, the author has developed an IndoWaveSentiment: Indonesian Audio Dataset for Emotion Classification.

3. Data Description

The data for this study were collected directly from voice recordings of emotional speech in the Indonesian language. The respondents were selected based on their professions as radio announcers, voice actors, or singers, all of whom are native Indonesian speakers. The sample consisted of 10 respondents, evenly split between 5 males and 5 females. Respondents were required to have proficient Indonesian language skills and speak with a standard Indonesian accent. Each respondent's voice was recorded three times, with the resulting data—including gen-

Table 1
Data of actors.

No of Actor	Gender	Age	Profession
Actor 1	M	24	radio announcers
Actor 2	F	25	voice actors
Actor 3	M	26	radio announcers
Actor 4	F	23	singers
Actor 5	M	32	radio announcers
Actor 6	F	25	radio announcers
Actor 7	M	22	radio announcers
Actor 8	F	22	singers
Actor 9	M	25	radio announcers
Actor 10	F	24	radio announcers

Table 2
The rules of file naming indowavesentiment dataset.

Identifier	*Description of factor levels*
Actors	(01–10) actors, odd numbers are male actors, while even numbers are female actors
Emotion Class	(01) neutral, (02) happy, (03) surprise, (04) disgust, (05) disappointed
Intensity	(01) normal, (02) strong
Repetition	(01) first repetition, (02) second repetition, (03) third repetition

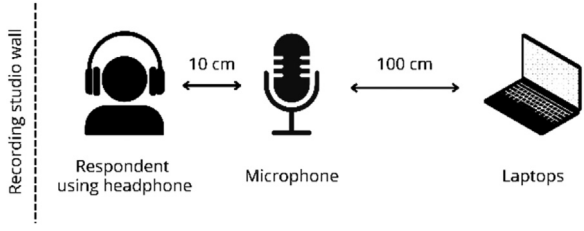

Fig. 1. Data collection scheme.

der, age, and background information—summarized in Table 1. Subsequently, the respondents' voices were recorded individually in a recording studio, as depicted in Fig. 1. The file naming conventions are detailed in Table 2.

Each respondent uttered the same sentence, namely "Kualitas HP ini cukup bagus," which means "The quality of this smartphone is quite good", expressing 5 different emotions: neutral, happy, surprised, disgusted, and disappointed. These emotions were conveyed at both normal and strong intensities, with each combination repeated 3 times. The chosen sentence is lexical, meaning it conveys complete and standalone information. This process results in 30 vocalizations per respondent. Each vocalization is exported in a single modality condition, which includes only audio. The primary dataset consists of 30 audio files per respondent, totalling 300 audio files (30 vocalizations $\times$ 10 respondents $=$ 300 audio files).

Having labelled each speech sample, multiple exports were performed to separate and save the audio based on the assigned labels. Each audio file was given a unique name consisting of four numbers or parts, separated by hyphens, to indicate its type. These identifiers were organized as 'Respondent - Emotion - Intensity - Repetition.wav.' Table 2 describes the rules of file naming in this research while Fig 2. Shows the organization of Audio File in IndoWaveSentiment [1]. For example, a file named '07-03-02-03.wav' signifies that the respondent is male (07), expressing surprise emotion (03), at strong intensity (02), on the third repetition (03).

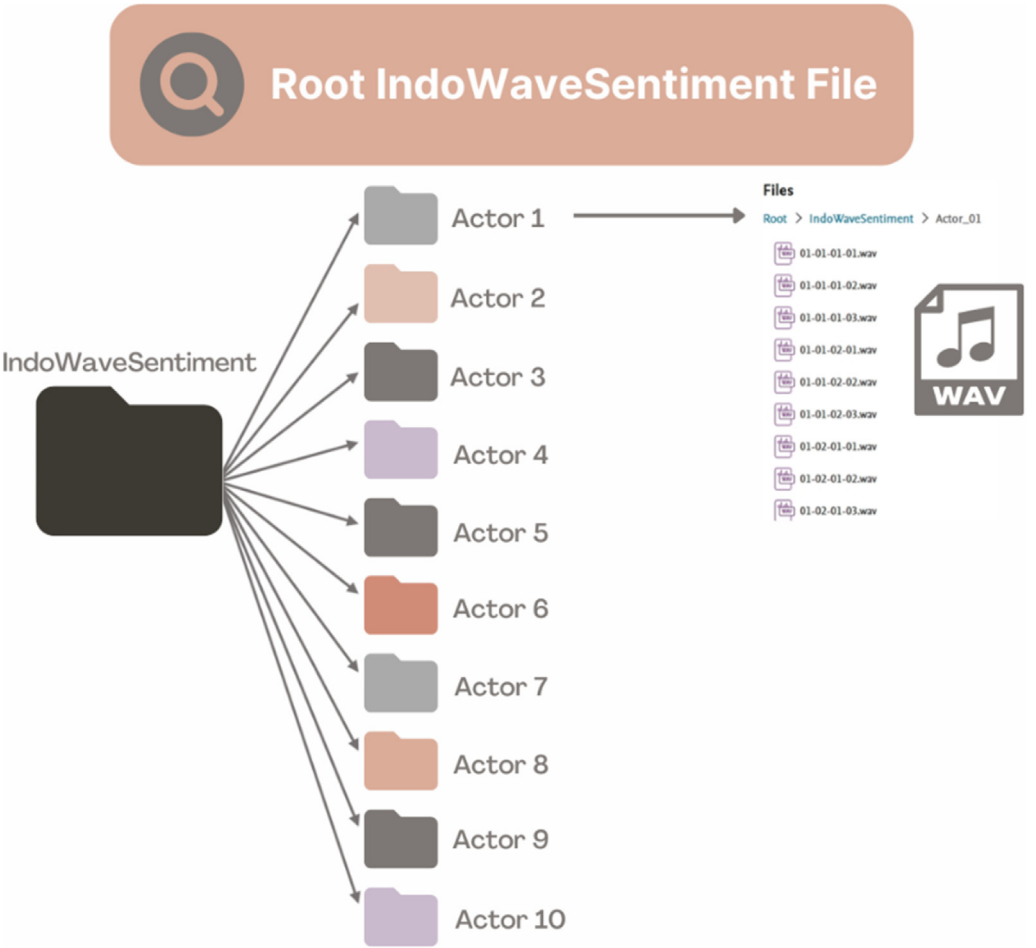

Fig. 2. Organization of IndoWaveSentiment files [1].

4. Experimental Design, Materials and Methods

4.1. Data recording

In this recording process, respondents are positioned in front of the studio wall, wearing headphones, and facing the microphone, as shown in Fig. 3. Their voices are captured using a cardioid vocal microphone with a mono channel, positioned no more than 10 cm away and connected to a laptop or computer. The recording process is visualized using the Audacity application. The voice recording settings are as follows:

a. Mono recording channel: This involves using a single sound channel to ensure better sound cohesion.
b. Sample rate: Indicates how many sound measurements are taken per second. Emotions, categorized as positive, negative, or neutral, offer valuable insights into individuals' attitudes and opinions Bit depth: Using 32 bits captures more sound details, resulting in higher quality audio.

4.2. Data pre-processing

After obtaining clean data recordings, the next steps involve volume normalization and noise reduction. This stage uses libraries such as librosa, pydub, and the Enhance Audio model, all of which are accessible for implementation.

4.3. Data annotation

After collecting the data, the next step is to assign an emotional label to each speech sample, with each label covering a duration of 3 to 5 s. Fig. 4 shows an example of an annotated primary data sample in the Audacity application. The aim of separating continuously recorded audio data is to improve the annotation process, reducing bias in cases where emotional intonations are similar across different classes. This ensures that class labels are more reliable and not based solely on a single individual's perception. At first, the annotation was performed by a single individual for all 300 audio samples. However, after reviewing the pairs of audio and labels, it became apparent that there was some ambiguity in the annotating process for certain classes. To address this issue, we implemented data validation to re-evaluate the labels.

4.4. IndoWaveSentiment data validation

To ensure that the annotation process is objective not solely based on the authors/researchers, we validated the IndoWaveSentiment dataset using an audio questionnaire administered through the QuestionPro tool. This method tests the reliability and validation of annotation data by sampling a wider range of respondents [8]. Data validation was carried out using a questionnaire involving 104 respondents who were asked to classify the sample emotion based on the audio clips they listened to**,** focusing on intonation rather than content. The validation survey results are detailed in Table 3. The survey results show that out of 20 questions, 12 have the highest scores matching the emotion class listed in the first column. If grouped into 3 sentiment classes, namely positive class (happy, surprise), negative class (disgust, disappointed), and neutral class, there are 18 questions with the highest score according to their class. However, the outcomes of this validation process are considered to enhance the earlier annotation work, particularly for certain classes that are often challenging to differentiate. For instance, the "happy" class was often annotated similarly to the "surprise" dataset as shown in the validation results for Q9 in Table 3. Therefore, we can re-evaluate and choose a more appropriate label.

Fig. 3. Recording situation.

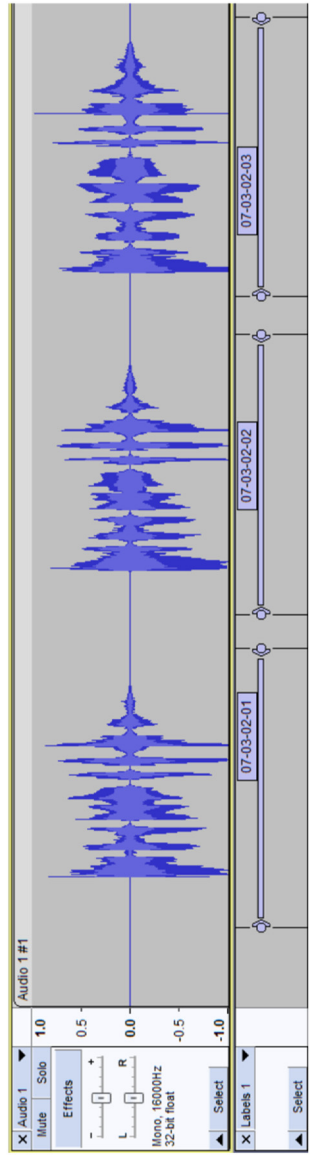

Fig. 4. Data annotating.

Table 3
Data validation based on survey's result.

Rater Selected Emotion	Emotion Expressed by Actor				
	Neutral	Happy	Surprised	Disgust	Disappointed
Q1 (neutral)	75.49%	6.86%	5.88%	5.88%	5.88%
Q2 (surprised)	2.94%	17.65%	65.69%	9.80%	3.92%
Q3 (surprised)	4.90%	62.75%	23.53%	8.82%	0.00%
Q4 (disgust)	10.78%	3.92%	2.94%	34.31%	48.04%
Q5 (disgust)	12.75%	0.00%	0.98%	34.31%	51.96%
Q6 (disappointed)	8.82%	0.00%	0.00%	13.73%	77.45%
Q7 (disappointed)	51.96%	4.90%	0.98%	4.90%	37.25%
Q8 (happy)	11.76%	61.76%	25.49%	0.98%	0.00%
Q9 (surprised)	2.94%	65.69%	25.49%	4.90%	0.98%
Q10 (happy)	19.61%	20.59%	53.92%	1.96%	3.92%
Q11 (disgust)	7.84%	0.00%	2.94%	13.73%	75.49%
Q12 (neutral)	84.31%	8.82%	1.96%	1.96%	2.94%
Q13 (disgust)	54.90%	5.88%	1.96%	16.67%	20.59%
Q14 (surprised)	10.78%	23.53%	58.82%	1.96%	4.90%
Q15 (happy)	11.76%	82.35%	4.90%	0.98%	0.00%
Q16 (neutral)	75.49%	12.75%	3.92%	2.94%	4.90%
Q17 (neutral)	77.45%	4.90%	0.00%	6.86%	10.78%
Q18 (disappointed)	17.65%	17.65%	20.59%	17.65%	26.47%
Q19 (disappointed)	26.47%	2.94%	1.96%	12.75%	55.88%
Q20 (happy)	8.82%	48.04%	38.24%	2.94%	1.96%

Limitations

The dataset includes only five vocal emotional expressions: neutral, happy, surprised, disgusted, and disappointed. The audio has undergone a basic noise-cleaning process, which can be optimized using appropriate techniques.

Ethics Statement

Each respondent actor provided informed consent for releasing their speech-audio recordings, which was voluntary-based All data collected from these voluntary respondents were anonymized after the entire data collection process. It was ensured that all submitted information would be kept confidential and anonymous. Each participant approved the samples after recording. Hence, no ethical approval was required

CRediT Author Statement

Anugrayani Bustamin: Conceptualization, Methodology, Software, Validation, Supervision and Writing-drafting, editing and reviewing. **Andi M Rizky**: Software, investigation, writing – editing. **Elly Warni:** Supervision, Validation, Writing – Editing. **Intan Sari Areni**: Data curation, Investigation, Writing – Reviewing and Editing. **Indrabayu:** Resources, Writing- Reviewing and Editing.

Data Availability

IndoWaveSentiment: Indonesian Audio Dataset for Emotion Classification (Original data) (Mendeley Data).

Acknowledgments

This research was partially supported by Hasanuddin University with contract number 00310/UN4.22/PT.01.03/2024 via PDPU-LPPM Grant. We would also like to thank the Artificial Intelligence and Multimedia Processing (AIMP) Research Group and Artificial Intelligence Laboratory, Informatics Department, Universitas Hasanuddin.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Bustamin, A.M. Rizky, E. Warni, I.S. Areni, I. Indrabayu, IndoWaveSentiment: Indonesian audio dataset for emotion classification, Mendeley Data V1 (2024), doi:10.17632/j9ytfdzy27.1.
[2] S. Akinpelu, S. Viriri, Speech emotion classification using attention-based network and regularized feature selection, Sci. Rep. 13 (2023) 11990, doi:10.1038/s41598-023-38868-2.
[3] M.B. Akcay, K. Oguz, Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers, Speech Commun. 116 (2020), doi:10.1016/j.specom.2019.12.001.
[4] L.F. Parra-Gallelgo, Orozco-Arroyave, Classification of emotions and evaluation of customer satisfaction from speech in real world acoustic environments, Digit. Signal Proc. 120 (2022), doi:10.1016/j.dsp.2021.103286.
[5] M.C. Caschera, P. Grifoni, F. Ferri, Emotion classification from speech and text in videos using a multimodal approach, Multimodal Technol. Interact. 6 (4) (2022) 28, doi:10.3390/mti6040028.
[6] A.M. Alqahtani, Product sentiment analysis for amazon reviews, Int. J. Comput. Sci. Inf. Technol. doi:10.5121/ijcsit.2021.13302. 13 (3), (2021) 15–30.
[7] Y. Khoirotul Aini, T. Budi Santoso, T. Dutono, Pemodelan CNN untuk deteksi emosi berbasis speech Bahasa Indonesia, J. Komput. Terapan 7 (1) (2021) 143–152 Vol. 7 No, doi:10.35143/jkt.v7i1.4623.
[8] S.R. Livingstone, F.A. Russo, The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English, PloS One 13 (5) (2018) e0196391, doi:10.1371/journal.pone.0196391.