

**THE ACCURACY OF AUTO-GENERATED SUBTITLES OF FOUR
INDONESIAN ENGLISH-SPEAKING YOUTUBERS**

A THESIS



By:

Saviera Christina Devika
1213018004

ENGLISH EDUCATION STUDY PROGRAM
FACULTY OF TEACHER EDUCATION
WIDYA MANDALA SURABAYA CATHOLIC UNIVERSITY
APRIL 2022

**THE ACCURACY OF AUTO-GENERATED SUBTITLES OF FOUR
INDONESIAN ENGLISH-SPEAKING YOUTUBERS**

A THESIS

Presented to Faculty of Teacher Training and Education
Widya Mandala Catholic University Surabaya
in partial fulfillment of the requirement for the Degree of
Sarjana Pendidikan in Teaching English as a Foreign Language



By:

Saviera Christina Devika
1213018004

ENGLISH EDUCATION STUDY PROGRAM
FACULTY OF TEACHER TRAINING AND EDUCATION
WIDYA MANDALA CATHOLIC UNIVERSITY SURABAYA
APRIL 2022

SURAT PERNYATAAN

Jalur Skripsi

Bersama ini saya:

Nama : SAVIERA CHRISTINA DEUTKA
Nomor Pokok : 1213018004
Program Studi : PENDIDIKAN BAHASA INGGRIS
Jurusan : BAHASA INGGRIS
Fakultas : KEGURUAN DAN ILMU PENDIDIKAN

Menyatakan dengan sesungguhnya bahwa skripsi saya yang berjudul:

THE ACCURACY OF AUTO-GENERATED SUBTITLES OF INDONESIAN
ENGLISH-SPEAKING YOUTUBERS

benar-benar merupakan hasil karya saya sendiri. Apabila Skripsi ini ternyata merupakan hasil *plagiarisme*, maka saya bersedia menerima sanksi berupa pembatalan kelulusan dan/ataupencabutan gelar yang telah saya peroleh.

Demikian surat pernyataan ini saya buat dengan sesungguhnya dan dengan penuh kesadaran.

Surabaya, 11 April 2022
Yang membuat pernyataan,



Saviera Christina Deutka

Mengetahui,
Dosen Pembimbing I,

Mateus Purnamanto

NIK: 121 97 0290

Dosen Pembimbing II,

NIK: _____

APPROVAL SHEET (I)

This thesis entitled “The Accuracy of Auto-Generated Subtitles of Four Indonesian English-Speaking Youtubers” prepared and submitted by Saviera Christina Devika (1213018004) has been approved to be examined by the Thesis Board of Examiners.



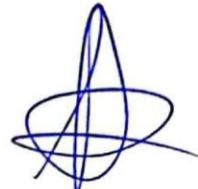
Mateus Yumarnamto, Ph.D

Thesis Advisor



Dr. Ruruh Mindari, M. Pd

Examiner 1



P. Hady Sutris Winarlim, S. Pd., M. Sc

Examiner 2

APPROVAL SHEET (II)

This thesis entitled "**The Accuracy of Auto-Generated Subtitles of Four Indonesian English-Speaking Youtubers**" prepared and submitted by Saviera Christina Devika (1213018004) has been approved to be examined by the Thesis Board of Examiners.

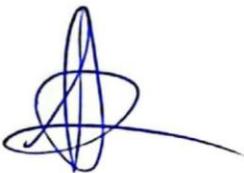


Dr. Riruh Mindari, M. Pd
Chair



Mateus Yumarnamto, Ph.D

Secretary



P. Hady Sutris Winarlim, S. Pd., M. Sc

Member



Faculty of Teacher Training and
Education



English Language Education Study
Program

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Demi Perkembangan Ilmu Pengetahuan, saya sebagai mahasiswa Universitas Katolik Widya Mandala Surabaya.

Nama Mahasiswa : Saviera Christina Devika
Nomor Pokok : 1213018004
Program Studi Pendidikan : Pendidikan Bahasa Inggris
Jurusan : Pendidikan Bahasa dan Seni
Fakultas : Keguruan dan Ilmu Pendidikan
Tanggal Lulus : _____

Dengan ini **SETUJU/TIDAK SETUJU** Skripsi atau Karya Ilmiah saya,

Judul :

The Accuracy of Auto-Generated Subtitles of Four Indonesian English-Speaking Youtubers

Untuk dipublikasikan/ditampilkan di Internet atau media lain (Digital Library Perpustakaan Universitas Katolik Widya Mandala Surabaya) untuk kepentingan akademik sebatas sesuai undang-undang Hak Cipta yang berlaku.

Demikian surat pernyataan **SETUJU/TIDAK SETUJU** publikasi Karya Ilmiah ini saya buat dengan sebenarnya

Surabaya, 12 Mei 2022

Yang menatakan



STATEMENT OF AUTHENTICITY

I declare that this thesis is my own writing, and it is true and correct that I did not take any scholarly ideas or work from others dishonestly. All the cited works were quoted in accordance with the ethical code of academic writing. I will take all the consequences if plagiarism is found in this thesis.

Surabaya, April 2022



**Saviera Christina Devika
(1213018004)**

ACKNOWLEDGEMENTS

Praise to be the Almighty God because of His grace I can finish this thesis entitled **“The Accuracy of Auto-Generated Subtitles of Four Indonesian English-Speaking Youtubers”** on time.

I am able to complete this thesis because of the helping of many people. I would like to express gratitude to those who have given their valuable time, love, guidance, patience, and support that encourage me to complete this thesis especially to:

1. Mr. Mateus Yumarnamto, Ph.D. as my advisor who always guides me during my thesis time.
2. Then, Mr. P. Hady Sutris Winarlim, S. Pd., M. Sc and Mrs. Dr. Ruruh Mindari, M.Pd as my thesis examiners who are willing to give me other advice for my thesis.
3. Mami, Papi, Ce Icha, and Max who always support me to finish this thesis.
4. Louisa Alexis who is willing to be my validator of this research.
5. My beloved best friends, Shelvia, Cici, Felicia, Safira, Mutiara, Grace, Louisa, and Mayvela who always support and courage me to finish this thesis.

Surabaya, April 2022

Saviera Christina Devika

ABSTRACT

Devika, S.C. (2022). *The Accuracy of Auto-Generated Subtitles of Four Indonesian English-Speaking YouTubers*. Thesis. English Education Study Program Faculty of Teacher Training and Education, Widya Mandala Catholic University, Surabaya, 2022.

Advisor: Mateus Yumarnamto, Ph.D.

Keywords: YouTube, YouTuber, Indonesian YouTuber, Pronunciation, Auto-generated subtitle, Inaccuracies in generating subtitles.

YouTube is well-known with almost 95% of internet users who use it today. Using its features, users are able to upload, watch and share videos freely. As it is easy to access, the video content must be well delivered. YouTube content creators (YouTubers) sometimes enjoy sharing their thoughts in a language other than their mother tongue, for instance, Indonesian using English. As YouTube provides Auto-Generated subtitles that have various subtitles quality, Indonesian YouTubers, as non-native speakers might make mistakes. This research aims to identify the inaccuracies in the subtitles of Indonesian YouTubers speaking English by generating two research questions; (i) Do the auto-generated subtitles represent the speakers' pronunciation accurately? (ii) What are the reasons underlying the inaccuracies if there are? Are there any sounds which are systematically misinterpreted by the auto-generated subtitle engine?

This research is descriptive qualitative. To collect the data, the researcher took the transcript of chosen two videos from each of four YouTube channels. The analysis was done by making 3 steps of analysis (Accuracy identification, Explanation of the inaccuracies, and Inaccuracies classification). The result of the analysis was written in narrative. The subjects were chosen based on these criteria; (1) Indonesian YouTuber and (2) Speak English in their videos.

The result of this research shows that there were some inaccuracies from the system, YouTubers, and unknown reasons. The researcher then classified the reasons of the inaccuracies into 16 sub reasons: (1) weak or inappropriate pronunciation, (2) weak voiced-voiceless sounds, (3) weak plosive sounds, (4) weak vowel sounds, (5) unclear alveolar sounds, (6) unfamiliar people's names/name of places/event/brand, (7) homophones, (8) using foreign language, (9) mixing Indonesian and English, (10) mixing Korean and English, (11) bad video-audio editing, (12) background noises, (13) harsh words, (14) speak too fast/unclear/too soft, (15) simultaneous talk, and (16) the ASR failed to interpret. The researcher then sharpened the sub reasons classification into five main reasons: (1) phonological reasons, (2) morphological reasons, (3) code switching, (4) technical reasons, and (5) miscellaneous reasons. Based on the findings, more systematic inaccuracies related to phonology can be conducted further.

TABLE OF CONTENTS

TITLE PAGE	i
SURAT PERNYATAAN	ii
APPROVAL SHEET (I)	iii
APPROVAL SHEET (II)	iv
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
STATEMENT OF AUTHENTICITY	vi
ACKNOWLEDGEMENTS.....	vii
ABSTRACT	viii
TABLE OF CONTENT	ix
TABLE OF TABLES.....	xi
TABLE OF FIGURES	xii
CHAPTER I	1
INTRODUCTION	1
1.1 Background of the study	1
1.2 Statements of the Problems.....	4
1.3 Objectives of the Study.....	4
1.5 Theoretical framework.....	4
1.6 Scope and Limitation of the Study	5
1.8 Organization of the Thesis.....	7
CHAPTER II	8
RELATED LITELATURE	8
2.1 English Articulatory Phonetics	8
2.2 Indonesian Articulatory Phonetics	12
2.3 Pronunciation	14
2.4 Previous Studies	14
CHAPTER III.....	18
RESEARCH METHODS	18
3. 1 Research Design	18
3. 2 Site and Contexts	18

3. 3 Subjects.....	18
3. 4 Instruments.....	21
3. 5 Data Collection Procedure	21
3. 6 Sources of Data.....	22
3. 7 Data Analysis Procedure.....	23
CHAPTER IV	27
DATA PRESENTATION, FINDINGS, AND DISCUSSIONS	27
4.1 Data Presentations and Findings	27
4.1.1 Data Presentation.....	27
4.2 Discussion and Implication	30
4.2.1 Inaccuracies Caused by Phonological Process	31
4.2.2Inaccuracies Caused by Morphological Reasons	31
4.2.3 Inaccuracies Caused by Code Switching.....	31
4.2.4 Inaccuracies Caused by Technical Problems	32
4.2.5 Miscellaneous Reasons for Inaccuracies	32
CHAPTER V	33
CONCLUSION AND SUGGESTION	33
5.1 Conclusion	33
5.2 Suggestions	33
5.3 Agenda for Further Research	34
REFERENCES	35
APPENDIX	37
APPENDIX 1: Data Analysis	37
APPENDIX 2: Data Analysis Summary.....	213
APPENDIX 3: Data Analysis Classification	287
APPENDIX 4: Validator's Comment.....	378

TABLE OF TABLES

Table 2.1 Related Study	16
Table 3.1 Sources of Data	22
Table 3.2 Accuracy Identification	23
Table 3.3 Explanation of the inaccuracies	24
Table 3.4 Inaccuracy Classification	25
Table 4.1 Inaccuracies Data Source	27
Table 4.2 The Percentage of the Reasons	29

TABLE OF FIGURES

Figure 2.1 Charting Consonant Sounds	8
Figure 2.2 Places of Articulation	9
Figure 2.3 Vowel Chart	12
Figure 2.4 Bahasa Indonesia Consonant Chart	13
Figure 2.5 Bahasa Indonesia Vowel Chart	13
Figure 3.1 Research Design	18
Figure 3.2 Fathia Izzati's Channel	19
Figure 3.3 Titan Tyra's Channel	19
Figure 3.4 Fiki Naki's Channel	20
Figure 3.5 SkinnyIndonesian24's Channel	20