THE USE OF AUTOMATIC SPEECH RECOGNITION (ASR) TOWARD LEARNING PRONUNCIATION SKILL AT THE ELEVENTH GRADE STUDENTS OF SMK PEMBANGUNAN BUKITTINGGI

A Thesis

Submitted as a Partial Fulfilment of the Research to Obtain Strata One (S1)

Degree on English Language Education Study Program

By:

REVI YOLLA HERMAN
2319.159

Advisor:

Dr. Absharini Kardena, M. Pd.
NIP. 198803192020122004

ENGLISH LANGUAGE EDUCATION STUDY PROGRAM

FACULTY OF TARBIYAH AND TEACHER TRAINING

STATE ISLAMIC UNIVERSITY OF SJECH M. DJAMIL DJAMBEK

BUKITTINGGI

2026 M / 1447 H

SURAT PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama	: Revi Yolla Herman
Nim	: 2319159
Fakultas	: Tarbiyah dan Ilmu Keguruan
Program Studi	: Pendidikan Bahasa Inggris
Judul Skripsi	: The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi

Menyatakan dengan sesungguhnya bahwa skripsi saya dengan judul di atas adalah asli karya saya sendiri. Apabila dikemudian hari terbukti bahwa skripsi ini bukan karya sendiri, maka saya bersedia diproses sesuai hukum yang berlaku dan gelar kesarjanaan saya dicabut sampai batas waktu yang tidak ditentukan.

Demikian surat pernyataan ini saya buat dengan sesungguhnya untuk dipergunakan sebagaimana mestinya.

Bukittinggi, 11 Juli 2026
Saya yang menyatakan,

Revi Yolla Herman
Reg No. 2319159

LETTER OF ADVISOR APPROVAL

The thesis entitled " The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi", written by Revi Yolla Herman, Student Identification Number (NIM) 2319159, a student of the English Education Study Program, Faculty of Tarbiyah and Teacher Training, has fulfilled the requirements to be examined and has been reviewed and approved to be presented in the Thesis Examination.

Bukittinggi, 11 July 2026
Advisor,

Dr. Absharini Kardena, M. Pd.
NIP. 198803192020122004

ABSTRACT

Revi Yolla Herman (2319.159): The Use of Automatic Speech Recognition (ASR) Toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi. A Thesis, English Education Study Program, Faculty of Tarbiyah and Teacher Training, State Islamic University of Sjech M. Djamil Djambek Bukittinggi, 2026.

Pronunciation is one of the essential components of English communication because it affects the clarity and accuracy of spoken language. However, many students still experience difficulties in learning English pronunciation due to limited opportunities for practice, lack of confidence, and the use of conventional teaching methods that provide limited immediate feedback. Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, offers opportunities for students to practice pronunciation independently while receiving instant feedback on their pronunciation performance. Therefore, this research aimed to identify the kinds of Automatic Speech Recognition (ASR) applications used by the students and to analyze students' perceptions toward learning pronunciation skills by using Automatic Speech Recognition (ASR) at the eleventh-grade students of SMK Pembangunan Bukittinggi.

This research employed a quantitative research approach with a descriptive survey design. The research was conducted at SMK Pembangunan Bukittinggi during the 2025/2026 academic year. The population consisted of all eleventh-grade students, while the sample was 24 students from class XI Teknik Alat Berat (TAB), selected using a random sampling technique. The data were collected through a questionnaire questionnaire instrument. The collected data were analyzed using descriptive statistical analysis by used Likert Scale to identify the kinds of Automatic Speech Recognition (ASR) applications used by the students and to describe students' perceptions toward the use of ASR in learning pronunciation skills.

The findings revealed that 91% of the students had positive perceptions toward the use of Google Gemini, while 80% had positive perceptions toward ChatGPT in learning pronunciation. In addition, 89% of the students showed positive perceptions in the cognitive aspect, 90% in the affective aspect, and 89% in the conative aspect. Overall, the questionnaire results showed an average percentage score of 88%, which was categorized as "Positive." These findings indicate that Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, effectively supports students' pronunciation learning by improving pronunciation awareness, motivation, confidence, and independent learning.

Keywords: Automatic Speech Recognition (ASR), Pronunciation Skill, Students' Perceptions.

ABSTRAK

Revi Yolla Herman (2319.159): Penggunaan Automatic Speech Recognition (ASR) Terhadap Pembelajaran Keterampilan Pronunciation Pada Siswa Kelas XI SMK Pembangunan Bukittinggi. Tesis, Program Studi Pendidikan Bahasa Inggris, Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, 2026.

Pengucapan (pronunciation) merupakan salah satu komponen penting dalam komunikasi bahasa Inggris karena memengaruhi kejelasan dan ketepatan penyampaian ujaran. Namun, banyak siswa masih mengalami kesulitan dalam mempelajari pengucapan bahasa Inggris akibat terbatasnya kesempatan untuk berlatih, kurangnya rasa percaya diri, serta penggunaan metode pembelajaran konvensional yang belum memberikan umpan balik secara langsung. Automatic Speech Recognition (ASR), khususnya Google Gemini dan ChatGPT, memberikan kesempatan kepada siswa untuk berlatih pengucapan secara mandiri sekaligus memperoleh umpan balik secara langsung terhadap performa pengucapan mereka. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi jenis aplikasi Automatic Speech Recognition (ASR) yang digunakan oleh siswa serta menganalisis persepsi siswa terhadap pembelajaran keterampilan pengucapan menggunakan Automatic Speech Recognition (ASR) pada siswa kelas XI SMK Pembangunan Bukittinggi.

Penelitian ini menggunakan pendekatan penelitian kuantitatif dengan desain survei deskriptif. Penelitian ini dilakukan di SMK Pembangunan Bukittinggi pada tahun ajaran 2025/2026. Populasi penelitian terdiri dari seluruh siswa kelas XI, sedangkan sampelnya adalah 24 siswa dari kelas XI Teknik Alat Berat (TAB), yang dipilih menggunakan teknik pengambilan sampel acak. Data dikumpulkan melalui instrumen kuesioner. Data yang terkumpul dianalisis menggunakan analisis statistik deskriptif dengan Skala Likert untuk mengidentifikasi jenis-jenis aplikasi Pengenalan Suara Otomatis (ASR) yang digunakan oleh siswa serta untuk mendeskripsikan persepsi siswa terhadap penggunaan ASR dalam pembelajaran keterampilan pengucapan.

Hasil penelitian menunjukkan bahwa 91% siswa memiliki persepsi positif terhadap penggunaan Google Gemini, sedangkan 80% siswa memiliki persepsi positif terhadap penggunaan ChatGPT dalam pembelajaran pengucapan. Selain itu, 89% siswa menunjukkan persepsi positif pada aspek kognitif, 90% pada aspek afektif, dan 89% pada aspek konatif. Secara keseluruhan, hasil angket menunjukkan skor rata-rata persentase sebesar 88%, yang termasuk dalam kategori "Positif." Temuan ini menunjukkan bahwa Automatic Speech Recognition (ASR), khususnya Google Gemini dan ChatGPT, secara efektif mendukung pembelajaran keterampilan pengucapan dengan meningkatkan kesadaran pengucapan, motivasi, rasa percaya diri, serta kemandirian belajar siswa.

Kata Kunci: Automatic Speech Recognition (ASR), Keterampilan Pengucapan, Persepsi Siswa.

TABLE OF CONTENTS

SURAT PERNYATAAN ORISINALITAS ..... **i**
LETTER OF ADVISOR APPROVAL ..... **ii**
ABSTRACT ..... **iii**
ABSTRAK ..... **iv**
TABLE OF CONTENTS ..... **v**
LIST OF APPENDICES ..... **vii**
LIST OF TABLES ..... **viii**
LIST OF FIGURES ..... **ix**
CHAPTER I **1**
INTRODUCTION **1**
A. Background of the Problem 1
B. Identification of the Problems 6
C. Limitation of the Research 6
D. Formulation of the Problem 6
E. Purpose of the Research 7
F. Significance of Research 8
G. Definition of Key Terms 8
CHAPTER II **10**
REVIEW OF RELATED LITERATURES **10**
A. Review of Related Theories 10
1. Pronunciation 10
2. Students' Perception in Using ASR for Pronunciation Skill 20
3. Automatic Speech Recognition (ASR) 23
B. Review of Related Studies 37
C. Conceptual Framework 41
CHAPTER III **42**
RESEARCH METHOD **42**
A. Research Design 42
B. Research Setting 43
C. Population And Sample 44
D. Instrument and Technique of Data Collection 46
E. Validity Dan Reliability 47

F. Trustworthiness of Data 50
G. Technique of Data analysis 51
CHAPTER IV 54
FINDING AND DISCUSSION 54
A. Finding 54
1. Kinds of ASR Applications 54
2. Students' Perception in Using ASR for Pronunciation Skill 60
B. DISCUSSION 74
CHAPTER V 78
CONCLUSION AND SUGGESTION 78
A. Conclusion 78
B. Suggestion 79
REFERENCES 81
APPENDICES 84

LIST OF APPENDICES

Appendix 1: Research Guideline.....84

Appendix 2: Students' Questionnaire 86

Appendix 3: Validation Sheets.....89

Appendix 4: Documentations.....93

Appendix 5: Surat Izin Penelitian 96

LIST OF TABLES

Table 3.1 Population 45

Table 3.2 Score Criteria of Questionnaire 47

Table 3.3 Index Icon Scores Each Item 48

Table 3.4 Results of Index Icon Scores 49

Table 3.5 Score Criteria of Validity Test 49

Table 3.6 Interpretation of Using ASR Toward Learning Pronunciation Skill 53

Table 4.1 Description of the Data of Kinds of ASR Applications towards English learning 55

Table 4.2 Analysis of the Data of Kinds of ASR Applications towards English learning 56

Table 4.3 Description of the Data of Students' Perception in Using ASR for Pronunciation Skill 61

Table 4.4 Data Analysis of Students' Perception in Using ASR for Pronunciation Skill 62

Table 4.5 Data Analysis of Kinds of ASR Applications and Students' Perception in Using ASR for Pronunciation Skill 69

Table 4.6 Interpretation of the Questionnaire Result 73

LIST OF FIGURES

Figures 2.1 Conceptual Framework 41

Figure 4.1: Students' Perceptions Using Google Gemini 58

Figure 4.2: Students' Perceptions in Using ChatGPT 59

Figure 4.3: Students' Perceptions in Cognitive Aspect..... 66

Figure 4.4: Students' Perceptions in Affective Aspect 67

Figure 4.5: Students' Perceptions in Conative Aspect 68

CHAPTER I

INTRODUCTION

A. Background of the Problem

Pronunciation is one of the most important aspects of communication because it influences how messages are conveyed and understood by listeners. Pronunciation refers not only to the articulation of sounds, but also to stress, intonation, rhythm, voice quality, and other suprasegmental features. Good pronunciation helps learners communicate clearly and confidently in English. In contrast, inaccurate pronunciation may cause misunderstanding and make communication ineffective. Therefore, pronunciation plays a crucial role in mastering speaking skill and achieving successful communication in English (Prashant, 2018).

Pronunciation teaching in English language learning has developed from focusing only on segmental sounds to a broader approach that includes suprasegmental features such as stress, rhythm, and intonation. This development shows that pronunciation is an integral part of oral communication. Both segmental and suprasegmental aspects are important to improve students' speaking accuracy and fluency. Teachers are also expected to create learner-centered learning environments that support students in practicing pronunciation actively according to their individual learning needs and characteristics (Tergujeff, 2014).

In recent years, technology has become an important part of English language learning. The use of technology can help teachers create more

interactive, effective, and interesting learning activities. In pronunciation learning, technology provides opportunities for students to practice independently and receive direct feedback on their performance. Therefore, technology-based learning media are considered effective in supporting students' pronunciation learning process.

One of the technologies that can support pronunciation learning is Automatic Speech Recognition (ASR). ASR is a technology that can recognize spoken language and convert it into text automatically. In language learning, ASR can help students practice pronunciation and identify pronunciation errors. This technology also enables learners to receive immediate feedback regarding their pronunciation performance, making the learning process more effective and efficient (Elimat & AbuSeileek, 2014).

Automatic Speech Recognition (ASR) provides several advantages in English pronunciation learning. First, ASR helps teachers identify students' pronunciation difficulties and allows students to practice independently according to their individual needs. Second, ASR technology can store students' learning progress, making it easier for both teachers and students to monitor pronunciation development. Third, ASR-based learning programs can create interactive and enjoyable learning activities, such as speech-based exercises and speaking practices, which can increase students' motivation and participation in learning pronunciation (Elimat & AbuSeileek, 2014).

Furthermore, ASR technology works through several important phases in pronunciation learning. The first phase is speech recognition, in which the system recognizes and converts spoken language into text. The second phase is scoring, where learners receive direct evaluation and feedback regarding their pronunciation. The third phase is error detection, which identifies students' pronunciation mistakes. The fourth phase is error diagnosis, where the system explains the types of pronunciation errors and suggests improvements. Finally, the feedback presentation phase helps learners understand their progress and improve their pronunciation skill effectively (Elimat & AbuSeileek, 2014).

The use of ASR technology in pronunciation learning can also reduce students' anxiety in speaking English. Some students feel shy and afraid of making mistakes when speaking in front of their classmates. Through ASR technology, students can practice pronunciation independently without feeling embarrassed. In addition, ASR-based learning activities are considered more engaging because students can actively participate in interactive pronunciation exercises.

However, despite the advantages of ASR technology, its use in English language learning is still limited in many schools, especially in vocational high schools (SMK). Many teachers still use conventional methods in teaching pronunciation, such as repetition drills, reading aloud, and teacher explanation. As a result, students have limited opportunities to practice pronunciation independently and receive immediate feedback on their pronunciation

performance. This condition shows a gap between the potential of ASR technology and its implementation in English language learning (Elimat & AbuSeileek, 2014).

However, although many previous studies have discussed the use of technology in English language learning, studies specifically investigating the use of Automatic Speech Recognition (ASR) toward learning pronunciation skill at the vocational high school level are still limited. Most previous studies focused on speaking skill, general language learning, or were conducted at the university level. In addition, there is still limited research investigating how ASR technology can support pronunciation learning among eleventh-grade students in vocational schools, especially at SMK Pembangunan Bukittinggi. Therefore, this research is important to fill the gap in previous studies by examining the use of Automatic Speech Recognition (ASR) toward learning pronunciation skill at the eleventh grade students of SMK Pembangunan Bukittinggi.

Based on preliminary research conducted through classroom observation on January 23rd, 2026, at SMK Pembangunan Bukittinggi, especially among eleventh-grade students, most eleventh-grade students at SMK Pembangunan Bukittinggi struggle with pronouncing English words. Most students obtained pronunciation scores below the Minimum Completeness Criteria (KKM), while only a small number of students achieved satisfactory scores.

Based on the observation, it can be concluded that many students still have difficulties in learning pronunciation. These difficulties are influenced by several factors from both teachers and students. From the teacher's side, the teaching method used in the classroom is still conventional and does not optimally support students in improving their pronunciation skill. The learning activities also provide limited opportunities for students to practice pronunciation interactively and independently.

From the students' side, several problems were identified. First, many eleventh-grade students at SMK Pembangunan Bukittinggi have difficulties in pronouncing English words correctly. This problem may be caused by limited knowledge of English phonetics, differences between English and Indonesian pronunciation systems, and lack of pronunciation practice. Second, many students show low motivation and interest in learning English pronunciation. Some students consider pronunciation difficult and less interesting, while others feel shy and lack confidence when speaking English. Third, the learning media and teaching methods used in the classroom have not effectively supported the development of students' pronunciation skill.

Considering these problems, the writer is interested in conducting research entitled *"The Use of Automatic Speech Recognition (ASR) Toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi."* This research is expected to identify how the use of ASR technology can support students in learning pronunciation and help improve their pronunciation skill in English learning.

B. Identification of the Problems

Based on the background of the problem above, the problems coul identified as follows:

1. Most eleventh-grade students at SMK Pembangunan Bukittinggi struggle with pronouncing English words.
2. Many students are interested in using Artificial Intelligence (AI) to improve their English pronunciation skills, but it is rarely integrated into classroom learning activities.
3. The curriculum and teaching methods at SMK Pembangunan Bukittinggi do not effectively support the development of English pronunciation skill.

C. Limitation of the Research

Based on the identification of the problem above, this research were limited to identifying the use of Automatic Speech Recognition (ASR) in learning pronunciation skill at the eleventh grade students of SMK Pembangunan Bukittinggi. This research focused on identifying the kinds of ASR used by the students and analyzing the students' perceptions toward learning pronunciation skill by using Automatic Speech Recognition (ASR).

D. Formulation of the Problem

Based on the limitation of the problem above, the problems of this research were formulated in the following questions:

1. What kind of Automatic Speech Recognition (ASR) is used by the students in learning pronunciation skill at the eleventh grade students of SMK Pembangunan Bukittinggi?
2. What are the students' perceptions of learning pronunciation skill by using Automatic Speech Recognition (ASR) at the eleventh grade students of SMK Pembangunan Bukittinggi?

E. Purpose of the Research

Based on the formulation of the research above, the purpose of this research were:

1. To identify the kinds of Automatic Speech Recognition (ASR) used by the students in learning pronunciation skill at the eleventh grade students of SMK Pembangunan Bukittinggi.
2. To find out the students' perceptions toward learning pronunciation skill by using Automatic Speech Recognition (ASR) at the eleventh grade students of SMK Pembangunan Bukittinggi.

F. Significance of Research

The results of this research were expected to provide the following practical benefits:

1. Teachers

   Teachers in senior high schools can use this research. They can discover the ASR in teaching senior high school students pronunciation skill.

2. Students

   Students of a senior high school can realize that pronunciation is a very important skill in English. Those who have difficulty pronouncing English words can be helped to improve their pronunciation through ASR.

3. The Researchers

   For other researchers, the research can be helpful for further research regarding ASR in teaching pronunciation skill to senior high school students. This research can be used as a reference for conducting a related research.

G. Definition of Key Terms

In this research, some definitions were put forward:

1. Automatic Speech Recognition (ASR)

   Automatic Speech Recognition is a technology that enables computers to recognize and transcribe spoken language into text. ASR systems use algorithms to analyze audio input, identify speech patterns,

and convert them into written text. In the context of the proposed research, ASR technology will likely be employed to analyze and provide feedback on students' pronunciation of English words.

2. Learning Pronunciation Skill

Learning pronunciation skill refers to the process of developing the ability to articulate sounds, stress, rhythm, and intonation accurately in a particular language, especially English. Learning pronunciation involves practicing the correct production of sounds and understanding how pronunciation influences communication. Effective pronunciation learning can be supported through various activities, exercises, and feedback that help learners improve their spoken language proficiency.

3. Students' Perception in Using ASR

Students' perception in using Automatic Speech Recognition (ASR) refers to students' views, opinions, feelings, and responses toward the use of ASR technology in learning pronunciation skill. It includes how students perceive the usefulness, effectiveness, advantages, and challenges of using ASR in the learning process. Students' perception is important because it can influence their motivation, interest, participation, and confidence in learning pronunciation through technology-based media.

CHAPTER II

REVIEW OF RELATED LITERATURES

A. Review of Related Theories

1. Pronunciation

This sub-section delves into the comprehensive understanding of pronunciation, encompassing its definition, key features, teaching methodologies and techniques, challenges faced by teachers, factors influencing students' difficulties, and the array of tools available for effective pronunciation instruction.

a. Definition of pronunciation skill

Clear and accurate pronunciation holds paramount importance in English communication, particularly in spoken discourse. Every intonation, stress pattern, and tone plays a pivotal role in conveying meaning effectively. For non-native English speakers, meticulous attention is required to pronounce words accurately to prevent misunderstandings. Thus, prioritizing clear pronunciation over native-like fluency is imperative for effective communication (Gilakjani, 2018).

Pronunciation is not merely the articulation of sounds but rather the utilization of sound systems integral to linguistic communication. It serves as a code for language, conveying meaning across diverse linguistic contexts. Crucially, pronunciation plays a

pivotal role in effective communication, serving as both the action and outcome of producing speech sounds, thereby facilitating linguistic expression and verbal communication. Understanding pronunciation involves not only the mechanics of sound production but also the nuanced interplay between sound, meaning, and context within language usage (Rosalia, Dayu, & Aprizani, 2021).

Furthermore, pronunciation is regarded as a vital aspect of the speaking component, encompassing the correction of word emphasis, rhythm, and intonation in spoken language. It serves as a means of conveying commonly recognized or understood expressions, significantly influencing the interpretation of utterance meaning. The importance of pronunciation in language utilization cannot be overstated, as it directly effects the comprehension and interpretation of the target language by listeners (Leong & Ahmadi, 2017).

In conclusion, clear and accurate pronunciation is indispensable for effective English communication, especially in spoken discourse. It involves not only the precise articulation of sounds but also the skillful use of intonation, stress patterns, and tones to convey meaning. Prioritizing clear pronunciation over native-like fluency is crucial for avoiding misunderstandings, particularly for non-native speakers. Pronunciation serves as a fundamental component of linguistic communication, acting as a code for language and facilitating effective verbal expression. Its role extends beyond sound production to

encompass word emphasis, rhythm, and intonation, all of which significantly influence the interpretation of utterance meaning. It can be concluded, pronunciation plays a vital role in language utilization, shaping how listeners comprehend and interpret spoken language.

b. Methods and Techniques in Teaching Pronunciation

Involves various techniques aimed at improving students' ability to articulate sounds accurately in the target language (Purwanto, 2019). Traditional pronunciation teaching techniques, include:

1) Listen and imitate

Students listen to model pronunciations provided by the teacher and replicate them.

2) Phonetic training

Utilization of articulatory descriptions, diagrams, and phonetic symbols to teach sound production.

3) Minimal pair drills

Exercises designed to help students distinguish between similar sounds that may pose difficulties.

4) Contextualized minimal pairs

Setting-based drills where students respond to specific linguistic contexts.

5) Visual aids

Use of visual media to enhance students' understanding of sound production.

6) Tongue twisters

Speech exercises aimed at improving articulation for native speakers.

7) Developmental approximation drills

Techniques that mirror the natural progression of sound acquisition observed in young English-speaking children.

8) Vowel-related affixation practice

Teaching vowel relationships based on principles of generative phonology.

9) Reading aloud

Students practice reading scripts aloud, with opportunities for feedback and self-evaluation provided through audio and video recordings. These techniques aim to enhance students' pronunciation skill by providing structured practice and feedback mechanisms.

Another set of techniques for teaching pronunciation, according to (Yates, L., & Zielinski, 2009) includes:

1) Drilling

Pronunciation practice is predominantly conducted through drilling, where students repeat words or structures provided by the instructor. Drilling serves to help students pronounce and remember new vocabulary items. Often, drilling is followed by eliciting, where students recall previously studied words, phrases,

or structures with the assistance of prompts, images, or gestures provided by the instructor.

2) Chaining

Chaining allows for the practice of lengthy sentences containing difficult words and sounds. Instructors isolate specific words from sentences, demonstrate them individually for students to replicate, and progressively assemble the words into complete sentences. There are two types of chaining:

a) Rear Chain

Students practice pronouncing sentences and constructing parts of words from the end, gradually increasing in length. Instructors address students' errors in pronouncing specific word parts individually, and students replicate each component demonstrated by the instructor.

b) Front Chain

Students articulate complete sentences and construct sentence fragments, starting from the beginning and gradually increasing their length. Instructors correct students' mispronunciations of specific word parts individually, and students replicate each component demonstrated.

c) Substitution drilling

This variation involves drilling a structure while substituting different components into the sentence.

d) Open-pair drilling

It involves organizing class-wide question-and-answer sessions, where one student poses a question and another responds.

e) Accepting feedback

Instructors provide correction as feedback to help learners mitigate pronunciation errors, thereby enhancing language accuracy. Utilizing a variety of correction strategies or techniques is crucial when providing feedback to students.

In conclusion, traditional and contemporary techniques for teaching pronunciation encompass a range of methods aimed at improving student's ability to articulate sounds accurately in the target language. Traditional approaches include listen and imitating, phonetic training, minimal pair drills, contextualized minimal pairs, visual aids, tongue twisters, developmental approximation drills, vowel-related affixation practice, and reading aloud. These methods provide structured practice and feedback mechanisms to enhance pronunciation skill. Additionally, contemporary techniques such as drilling, chaining, substitution drilling, open-pair drilling, and accepting feedback offer further avenues for refining pronunciation through repetitive practice, sentence construction, and interactive sessions. By incorporating a combination of these techniques, instructors can effectively address

pronunciation challenges and facilitate language accuracy among learners.

c. The Tools for Teaching Pronunciation

Contemporary approaches to teaching pronunciation include listening and imitation exercises, the use of visual aids to aid understanding of sound production and practice with vowel and stress shifts related to affixes. Moreover, recording learners' output can provide valuable feedback and opportunities for self-assessment. By combining these traditional and modern methods, language teachers can create dynamic and comprehensive pronunciation lessons that cater to diverse learning styles and needs.

While traditional pronunciation teaching techniques have proven effective for many students, it's important to acknowledge that different learners have varying abilities and preferences when it comes to acquiring pronunciation skill in a new language. While some students may thrive with methods involving explicit instruction and practice, others may find it challenging to achieve desired outcomes solely through these approaches. Recognizing this diversity in learner needs, educators are continuously exploring and developing new techniques to complement the acquisition of English pronunciation. These emerging techniques aim to provide alternative pathways to pronunciation mastery, catering to a wider range of learning styles and preferences. By embracing innovative approaches alongside traditional

methods, language educators can better support the diverse needs of their students and enhance learning outcomes in pronunciation acquisition (Celce-Murcia, Brinton, & Goodwin, 2010).

The integration of various disciplines such as drama, psychology, and speech pathology has led to the emergence of innovative approaches in teaching and learning English pronunciation. These approaches emphasize the use of fluency-building activities alongside accuracy-oriented exercises, catering to diverse learning styles and preferences. Additionally, techniques appeal to multisensory modes of learning, incorporating authentic materials to enhance engagement and relevance. Moreover, the application of instructional technology plays a crucial role in facilitating pronunciation instruction, offering interactive tools and resources to support learners' progress. Through the integration of these diverse methodologies, educators can provide comprehensive and effective instruction in English pronunciation, fostering greater proficiency and confidence among learners (N. M. Arafa, Elbarougy, A. Ewees, & Behery, 2018).

In the realm of education, a plethora of technological tools and applications are employed to enhance learning experiences. These include a diverse range of apparatuses such as computers, digital cameras, scanners, LCD panels or projectors, and distance education or video conferencing systems. Additionally, software tools like word processing, databases, spreadsheets, drawing or graphics programs,

website development platforms, electronic references, and discussion groups or list servers contribute significantly to educational practices. Furthermore, instructional software, encompassing tutorials, drills, and practice modules, stands out as a prominent tool used to aid language learners. These technological resources not only facilitate interactive and engaging learning but also cater to diverse learning styles, ultimately enriching the educational experience for students (Muir-Herzig, 2004).

It can be concluded that, online tools leveraging artificial intelligence (AI) have revolutionized language learning, particularly in the realm of pronunciation. Through automatic speech recognition (ASR) technology, users can engage with on-screen phrases, speaking aloud while the software precisely analyzes their pronunciation against correct English pronunciation standards. This real-time feedback mechanism allows learners to receive instant guidance, with mispronounced words highlighted and specific instructions provided to improve articulation, stress, and intonation. ASR-equipped tools thus empower learners to autonomously refine their pronunciation skill, offering a dynamic and interactive approach to language learning.

d. Assessment of pronunciation

Teachers evaluate their student's pronunciation through speech evaluations. When teachers evaluate speech, they look at a number of factors. The first three are vocabulary, grammar, and pronunciation

(Zaki Abdul Aziz, 2023). Students typically converse with one anothe, and the instructor offers discussion starters like:

1) Expressions of Greeting and Parting

Greeting is an expression used to acknowledge someone's presence and initiate a conversation. It is a way to show recognition and establish communication. Parting expressions, on the other hand, are used when departing from someone. These expressions mark the end of an interaction and signify the separation. Examples of parting expressions include phrases like "goodbye" and "see you later."

2) Expressions of Happiness

Expressions of happiness convey a person's positive feelings about a particular situation or event that brings joy. Examples include receiving a gift on a birthday, discovering a new toy, or acquiring a desired item from the market. Such expressions reflect the emotional response to experiences that elicit happiness.

3) Congratulating and Completing

Congratulating involves expressing congratulations to someone who has achieved or experienced something positive. It is a way to acknowledge their success or happiness. Completing, in this context, refers to concluding or wrapping up a conversation.

This might involve summarizing points discussed or providing final remarks to close the dialogue.

4) Expressions of Giving and Asking for Opinions

Giving and asking for opinions are essential techniques for engaging in meaningful conversations and exchanging viewpoints. These expressions facilitate the sharing and receiving of thoughts and perspectives. In daily activities, it is common to use these expressions to seek input from others or to offer one's own opinion, thereby fostering interactive and collaborative communication.

2. Students' Perception in Using ASR for Pronunciation Skill

Students' perceptions play an important role in the success of learning pronunciation skills. Perception refers to the process of understanding and interpreting learning experiences, including the use of technology in language learning. In this study, students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation are examined because they can influence the effectiveness of the learning process. These perceptions are viewed through three aspects: cognitive, affective, and conative, which reflect students' knowledge, feelings, and behavioral tendencies toward ASR (Mahmudi & Anugerahwati, 2021). Therefore, investigating students' perceptions can provide valuable insights into the role of ASR in improving the

pronunciation skills of eleventh-grade students at SMK Pembangunan Bukittinggi.

According to Jihye, et al., (2018), there are three aspects which reflect students' toward students' perception in using asr for pronunciation skill.

1) Cognitive Aspect

The cognitive aspect refers to students' knowledge, understanding, beliefs, and awareness regarding the use of Automatic Speech Recognition (ASR) in learning pronunciation. This aspect focuses on how students perceive the usefulness of ASR, the information and feedback provided by the application, and its role in helping them improve pronunciation accuracy. Students' cognitive perceptions are reflected in their understanding of how ASR functions as a learning tool and supports the development of pronunciation skills.

2) Affective Aspect

The affective aspect relates to students' feelings, emotions, and attitudes toward the use of ASR in pronunciation learning. It includes students' level of interest, enjoyment, confidence, motivation, and satisfaction when using ASR applications. Positive affective responses indicate that students feel comfortable and encouraged to practice pronunciation, while negative responses may arise from difficulties in using the technology or dissatisfaction with the feedback provided.

3) Conative Aspect

The conative aspect refers to students' intentions, willingness, and actual behaviors related to the use of ASR for pronunciation practice. This aspect reflects students' tendency to continue using ASR, their willingness to engage in pronunciation activities, and their commitment to improving their pronunciation skills through the application. Conative responses are generally influenced by students' cognitive beliefs and affective attitudes toward ASR.

Students' perceptions of using Automatic Speech Recognition (ASR) in learning pronunciation are important because they influence students' engagement and learning outcomes. ASR provides immediate feedback on pronunciation, allowing learners to identify and correct their errors independently. Many students perceive ASR as a useful and user-friendly tool that increases motivation, confidence, and opportunities for pronunciation practice. Through continuous interaction with ASR, learners can become more aware of their pronunciation accuracy and develop their speaking skills more effectively. Although some students may experience difficulties due to recognition errors or system limitations, previous studies have generally reported positive perceptions of ASR as a supportive technology for pronunciation learning. Therefore, understanding students' perceptions of ASR can provide valuable insights into its effectiveness in improving pronunciation skills among EFL learners (Sun, 2023).

In conclusion, students' perceptions of using Automatic Speech Recognition (ASR) in learning pronunciation can be understood through the cognitive, affective, and conative aspects proposed by Bu and Go (2018). The cognitive aspect reflects students' knowledge and beliefs about the usefulness of ASR in improving pronunciation skills, the affective aspect relates to their feelings, motivation, and attitudes toward using ASR, while the conative aspect concerns their willingness and tendency to continue using ASR for pronunciation practice. These three aspects provide a comprehensive framework for understanding how students perceive the use of ASR in pronunciation learning. Therefore, examining students' perceptions based on these dimensions can provide valuable insights into the role of ASR, particularly Google Gemini, in supporting the pronunciation skill development of eleventh-grade students at SMK Pembangunan Bukittinggi.

3. Automatic Speech Recognition (ASR)

a. Definition of Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) stands as an autonomous mechanism, operating in the realm of machines, dedicated to deciphering and transcribing spoken language. In its essence, an ASR system engages in a seamless interaction with oral speech: it captures acoustic signals emanating from a speaker via a microphone, subjects these signals to intricate patterns, models, or algorithms, and ultimately generates an output, predominantly in textual form. This technological

marvel epitomizes the fusion of linguistics and computational prowess, revolutionizing communication and information processing across diverse domains (Levis & Suvorov, 2012).

Automatic speech recognition represents a pinnacle in the field of speech processing, enabling machines to comprehend and interpret human speech, thus facilitating a seamless exchange of information between man and machine. Referred to interchangeably as speech recognition, it functions by converting spoken language into textual form through sophisticated computer algorithms. This transformation not only fosters natural communication but also underpins a wide array of applications spanning various domains. Graphical representations of speech frequencies over time elucidate its essence, while its integration with speech synthesis, processing, speaker identification, and verification technologies forms the backbone of voice interfaces or Human-Machine Interfaces (HMIs). Voice recognition finds extensive utility across diverse sectors, including voice services such as speaking clocks and weather updates, data entry for quality control purposes, aviation training, assistance for individuals with disabilities, and vocal dictation. Furthermore, embedded voice recognition modules, exemplified in mobile phones and automobiles, empower users to interact effortlessly with their devices, whether adjusting car settings, navigating the internet, or controlling onboard systems through voice commands. This confluence of technological advancements not only

enhances user convenience but also augments efficiency and accessibility across various spheres of human activity (Ibrahim & Varol, 2020).

Hence, the continual development and production of Automatic Speech Recognition (ASR) systems remain imperative for human interaction with technology. Speech recognition technology has revolutionized machine comprehension of human languages, offering a pivotal solution for bridging the gap between human beings and information processing devices. With the pervasive influence of information technology permeating various facets of our daily lives, the efficacy of communication between individuals and technological interfaces assumes ever-increasing significance. Historically, human-computer interaction has primarily relied on traditional input methods such as keyboards and screens. However, speech, the most ubiquitous, intuitive, and rapid mode of human communication, has emerged as a paramount channel for interaction. Within speech recognition systems, numerous parameters intricately influence the accuracy and effectiveness of the recognition process, reflecting the interdisciplinary nature of this field and its continuous quest for refinement and optimization (Haton, 2003).

From the statements above, it is concluded that Automatic Speech Recognition (ASR) is a technological breakthrough that connects humans with information processing devices. ASR facilitates

natural communication between humans and machines and supports a wide range of applications in various domains. It is essential for improving accessibility and efficiency in everyday life. ASR also reflects advances in the field of communication and information processing, with its continuous research and development playing a role in improving human interaction with technology.

b. Kinds of ASR Applications

Various Automatic Speech Recognition (ASR) applications have been developed to support pronunciation learning by providing speech analysis and immediate feedback. Popular ASR-based applications such as ELSA Speak, Duolingo, Google Speech Recognition, and speech-to-text features in mobile devices help learners practice pronunciation, stress, intonation, and speaking fluency. These applications allow students to compare their spoken output with standard pronunciation and identify errors for self-correction. Through continuous practice and instant feedback, ASR applications can enhance learners' pronunciation accuracy, confidence, and autonomous learning. Therefore, ASR applications have become valuable tools for supporting pronunciation skill development in English language learning (Sugandhamani, Vairamuthu, & Selvaraj, 2026).

Google Gemini is an artificial intelligence application that can support pronunciation learning through its speech recognition and

conversational features. By interacting with Gemini using voice input, students can practice speaking English and receive responses based on their spoken language. This application helps learners improve pronunciation, fluency, and speaking confidence by providing opportunities for repeated practice in a flexible learning environment. In addition, Gemini can assist students in identifying pronunciation mistakes and encourage them to refine their speaking performance through continuous interaction (Sugandhamani et al., 2026).

ChatGPT is another AI-powered application that can be utilized to support pronunciation learning. Through voice conversation features, students can practice English pronunciation while engaging in interactive communication. ChatGPT allows learners to receive instant responses, practice various speaking topics, and develop greater awareness of pronunciation accuracy. Furthermore, the application promotes autonomous learning by enabling students to practice anytime and anywhere, making pronunciation learning more accessible, engaging, and effective (Sugandhamani et al., 2026).

In conclusion, various ASR applications have been developed to support pronunciation learning by providing speech recognition and immediate feedback. Among these applications, Google Gemini and ChatGPT offer voice interaction features that enable students to practice pronunciation, receive instant responses, and improve their speaking performance. Both applications facilitate repeated practice,

self-correction, and independent learning, allowing students to develop greater pronunciation accuracy, fluency, and confidence in speaking English. By providing accessible and interactive learning opportunities, Google Gemini and ChatGPT can support the development of pronunciation skills among EFL learners. Therefore, these ASR-based applications are considered suitable tools for investigating the use of Automatic Speech Recognition (ASR) toward learning pronunciation skills among eleventh-grade students at SMK Pembangunan Bukittinggi.

c. Characteristics of Automatic Speech Recognition (ASR)

Automatic speech recognition systems are meticulously crafted to address specific challenges. A myriad of parameters intricately shape the design of these systems, necessitating a comprehensive understanding of each aspect to delineate their operational scope effectively. Key considerations encompass modeling units such as words, syllables, and phonemes employed for recognition, the breadth of vocabulary ranging from small to large, and the complexity of tasks addressed, spanning from simple to intricate scenarios employing N-gram language models. Additional factors include task perplexity, speaking modes encompassing isolated, connected, continuous, and spontaneous speech, as well as speaker modes categorized as trained, adaptive, speaker-independent, or speaker-dependent. Furthermore, the speech environment, whether quiet or noisy, and the transducers

utilized, such as high-quality microphones, telephones, or array microphones, profoundly effects system performance. Moreover, the transmission channel presents yet another dimension that warrants meticulous consideration in the design process (S & E, 2016).

Speech recognition systems exhibit distinct characteristics across three primary dimensions: speaker dependence, speech continuity, and vocabulary size. Speaker dependence delineates systems as either requiring individual training for each speaker (speaker-dependent), drawing from a diverse training database for universal speaker recognition (speaker-independent), or evolving from speaker-independent to personalized recognition through gradual adaptation (adaptive). Concerning speech continuity, systems vary from isolated word recognition, identifying words uttered in isolation, to connected word recognition, capable of discerning words spoken consecutively without pauses, and further to continuous speech recognition, proficient in understanding complete sentences seamlessly. Additionally, word spotting systems excel in extracting individual words and phrases from continuous speech streams. Lastly, the vocabulary size to which ASR systems are trained serves as a defining factor, categorized as either small or large vocabulary systems. These dimensions collectively shape the capabilities and functionalities of speech recognition technology, catering to diverse user needs and contexts (Levis & Suvorov, 2012).

Understanding the variables inherent in speech recognition systems is crucial for selecting the most suitable algorithms. Among these variables, the types of speech are paramount, typically categorized into four types: isolated words, connected words, continuous speech, and spontaneous speech. Isolated words necessitate pauses between utterances, whereas connected words allow minimal pauses between them. Continuous speech systems enable users to speak naturally while the computer selects the content, posing significant challenges. Spontaneous speech, on the other hand, is natural-sounding and unrehearsed. Additionally, the size of the vocabulary employed profoundly effects system complexity and processing requirements, thereby influencing accuracy. Vocabulary sizes vary, ranging from small vocabularies containing tens of words to very large vocabularies comprising tens of thousands of words. Moreover, speaker dependence is a critical factor, with systems categorized as speaker-dependent, requiring user training, speaker-independent, suitable for any speaker, or speaker-adaptable, capable of adapting to new speaker characteristics. These variables collectively shape the design and functionality of speech recognition systems, catering to diverse user needs and contexts (Ibrahim & Varol, 2020).

In conclusion, the development and implementation of automatic speech recognition (ASR) systems entail meticulous consideration of various parameters and dimensions. These systems

are intricately designed to tackle specific challenges, ranging from modeling units and vocabulary size to speech continuity and speaker dependence. ASR technology encompasses a wide array of functionalities, catering to diverse user needs and contexts. Understanding the variables inherent in ASR systems is crucial for selecting appropriate algorithms and optimizing system performance. By addressing key considerations such as speech types, vocabulary size, and speaker dependence, ASR systems can effectively bridge the gap between human communication and technological interfaces, thereby enhancing efficiency and accessibility across various domains.

d. Developing of Automatic Speech Recognition (ASR)

The integration of automatic speech recognition into computer-based tools for speech and language development in children holds immense potential. While such tools cannot replace the invaluable human interaction involved in teaching children to read, they can significantly enhance individualized assistance and optimize the utilization of time spent with teachers or parents. Considering the considerable advantages offered by these tools, coupled with the economic significance of literacy, it is unsurprising that the speech technology research community is actively addressing this issue. The incorporation of automatic speech recognition technology into educational settings can revolutionize the way children acquire

language skill, providing personalized assistance and optimizing learning opportunities (Russell et al., 1996).

The development process for a general automatic speech recognizer typically involves several sequential steps. Initially, speech is acquired using a unidirectional, noiseless microphone, ensuring clear input. Subsequently, the acquired speech signal undergoes signal parameterization through a suitable feature extraction technique, facilitating further analysis. Acoustic analysis follows, wherein the signal is processed to convert training waveforms into a series of coefficient vectors, enabling a deeper understanding of its acoustic properties. Next, a prototype of Hidden Markov Models (HMM) is defined for each element within the task vocabulary, serving as the foundation for subsequent training. Through model training, each HMM is initialized and trained using acquired data, refining its ability to recognize patterns within the speech data. Task definition comes next, where the grammar for the speech recognizer is established, outlining the rules and constraints for recognition. Finally, the unknown input speech signal is recognized using the trained models, completing the recognition process. It's important to acknowledge that writer may introduce alterations to these steps over time, driven by advancements in existing techniques, the introduction of new methodologies, and variations in the acoustic structures of different languages (Ghai & Singh, 2012).

The standard framework for developing an automatic speech recognition (ASR) system typically involves generating training and development data to build the system, along with evaluation data for final performance analysis. However, this framework presents challenges for systems operating across broad domains or where data collection is limited. In this program, participating institutions are tasked with developing ASR technology capable of accurate machine translation, cross-language information retrieval, and summarization, even in the absence of target domain data for training or testing purposes. This ambitious goal underscores the need for innovative approaches and robust algorithms to overcome the inherent challenges of working with diverse and unstructured datasets (Communication, Ragni, & Gales, 2018).

From the statements above, the integration of automatic speech recognition (ASR) technology into educational tools for children's speech and language development offers significant potential by enhancing individualized assistance and optimizing learning opportunities, albeit unable to replace the essential human interaction involved in teaching children to read. The development process for a general ASR system typically follows a sequential methodology, from speech acquisition to model training and recognition, with ongoing refinements influenced by advancements in techniques and variations in language structures. However, the standard framework for ASR

system development poses challenges, particularly in broad domains or with limited data availability. Nonetheless, initiatives aiming to develop ASR technology capable of accurate machine translation and cross-language information retrieval underscore the necessity for innovative approaches to address diverse and unstructured datasets effectively.

e. Advantages of using Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) technology offers a myriad of benefits in educational settings (Alapján-, 2016):

1) Personalized Learning

ASR enables tailored learning experiences by adapting content to individual students' needs and learning styles, providing personalized suggestions and aids.

2) Flexibility

Students can learn at their own pace and schedule, accommodating various learning styles and ensuring effective learning outcomes.

3) Engagement

Interactive elements like games and exams keep students engaged, enhancing motivation and participation compared to traditional methods.

4) Accessibility

ASR technology transcends physical barriers, bringing education to remote areas and enabling collaboration among diverse student populations.

5) Instructional Quality

AI-powered tutoring services and high-quality instruction are facilitated by ASR, ensuring consistent and effective teaching practices.

6) Time and Effort Savings

Instructors can efficiently create written records of lectures and generate quizzes, allowing more time for individualized instruction and other teaching tasks.

7) Cost-effectiveness

ASR reduces resource requirements and travel time for both teachers and students, making education more accessible and affordable.

8) Enhanced Assessment

Automated grading systems and analysis of spoken responses improve assessment accuracy and enable comprehensive progress tracking.

f. Disadvantages of using Automatic Speech Recognition (ASR)

Although Automatic Speech Recognition (ASR) offers many benefits for pronunciation learning, it also has several limitations. One

of the main disadvantages is the inaccuracy of speech recognition, especially when learners have strong accents, unclear pronunciation, or speak in noisy environments. In some cases, ASR systems may fail to recognize correctly pronounced words or may incorrectly transcribe mispronounced words as correct, resulting in misleading feedback for learners. This limitation can reduce the effectiveness of pronunciation practice and may lead to learner frustration (McCrocklin, 2016). Furthermore, the accuracy of ASR feedback varies across different applications, making some systems more reliable than others for pronunciation training.

Another disadvantage of ASR is that the feedback provided is often indirect and limited. Many ASR applications only display transcription results or pronunciation scores without clearly explaining the specific pronunciation errors and how to correct them. As a result, learners may find it difficult to understand the source of their mistakes and improve their pronunciation effectively. Research has shown that explicit corrective feedback tends to be more beneficial than indirect ASR feedback for pronunciation development (Ngo, Chen, & Lai, 2024).

In addition, technical issues and system limitations may affect students' learning experiences. ASR performance can decrease due to background noise, differences in speech rate, accent variation, microphone quality, or internet connectivity problems. Some learners

may also experience difficulties in understanding the feedback generated by the system or become discouraged when the application repeatedly misinterprets their speech. Therefore, although ASR is a valuable tool for pronunciation learning, its limitations should be considered when integrating it into language learning activities (Liakin, Cardoso, & Liakina, 2015).

In conclusion, although Automatic Speech Recognition (ASR) provides valuable support for pronunciation learning, it also has several limitations that may affect students' learning experiences. Issues such as speech recognition inaccuracies, limited feedback, and technical constraints can reduce the effectiveness of pronunciation practice and create challenges for learners. Despite these disadvantages, ASR remains a useful tool when used appropriately to complement language instruction. Therefore, understanding both the strengths and limitations of ASR is important for evaluating its use in learning pronunciation skills among eleventh-grade students at SMK Pembangunan Bukittinggi.

B. Review of Related Studies

Several previous studies have investigated the use of Automatic Speech Recognition (ASR) technology in language learning, particularly in improving learners' pronunciation skills.

The first study was conducted by Elimat & AbuSeileek, (2014), which examined the effect of ASR technology on the pronunciation performance of EFL students. The study involved 64 third-grade students who were divided into experimental and control groups. The experimental group used ASR through the Tell Me More Performance English program, while the control group received conventional instruction. The findings revealed that students who learned with ASR achieved significantly better pronunciation scores than those who learned through traditional methods. The study also found that individual learning activities using ASR were more effective than pair and group work activities.

The second study was carried out by Bashori et al., (2024), who investigated the effectiveness of two ASR-based language learning systems, namely I Love Indonesia (ILI) and NovoLearning (NOVO), in improving English pronunciation among Indonesian high school students. A total of 117 students participated in the five-week experiment. The results showed that both systems contributed to significant improvements in learners' pronunciation at both the word and sentence levels. However, NOVO, which provided detailed phonetic feedback, produced greater improvement than ILI. The researchers concluded that ASR-based applications have strong potential to support pronunciation learning.

The third study was conducted by Liu et al., (2019), which explored the use of ASR technology for improving English pronunciation among Chinese learners of English. The findings indicated that ASR technology

helped learners improve their pronunciation accuracy and provided opportunities for self-correction through immediate feedback. The study suggested that ASR can serve as an effective instructional tool for both learners and teachers in pronunciation practice.

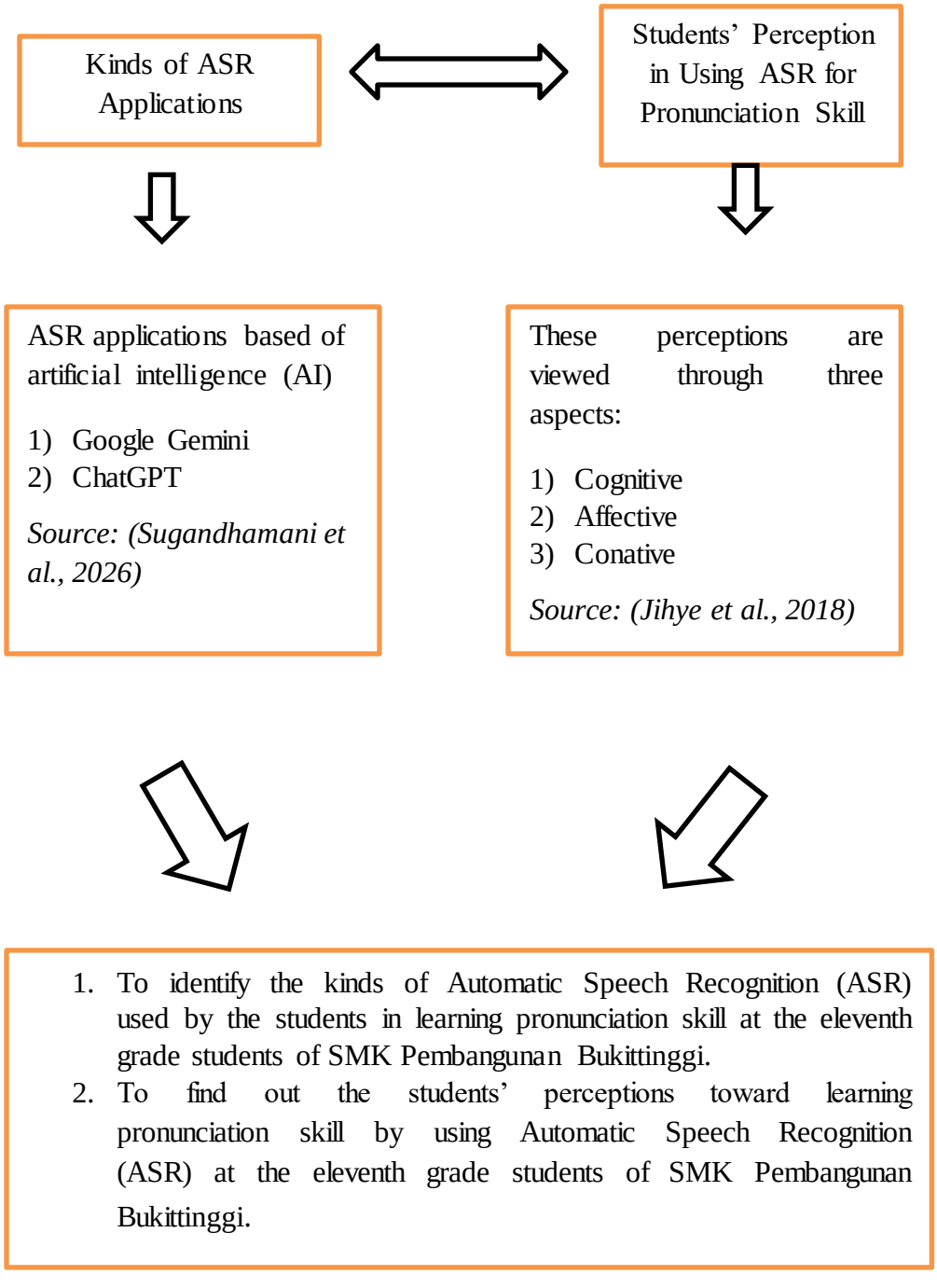
The three studies have the same focus, namely on the application of Automatic Speech Recognition (ASR) technology in improving language pronunciation skill. In all of these studies, participants or students were involved in the language learning process, ranging from elementary school students to high school students, as well as students learning English as a second or foreign language. The results showed a significant improvement in pronunciation skill in the group using ASR technology, both results and at the word and sentence level. From the results, it was concluded that ASR technology can be an effective tool in improving pronunciation skill in language learning.

There are significant differences between the three studies. First, the research subjects are different. The first study involved grade 3 students at Al-Abrar Basic Mixed School, the second study involved high school students in Indonesia, while the third study involved English language learners in China. Second, the variety of learning methods: The first study compared the use of ASR technology with regular learning methods, while the second study compared two ASR-based language learning systems, and the third study focused on the application of ASR technology for pronunciation improvement. Third, different cultural and language contexts: The first and second studies

were conducted in Indonesia, while the third study was conducted in China. Finally, different research objectives: The first study aimed to explore the effect of using ASR in learning English pronunciation, the second study to evaluate two ASR-based language learning systems, while the third study was to examine the effectiveness of ASR in improving English pronunciation among Chinese language learners.

Based on the previous studies, it can be concluded that ASR technology has a positive impact on learners' pronunciation development. Most studies reported improvements in pronunciation accuracy, increased opportunities for independent practice, and effective feedback for learners. However, the present study differs from those previous studies in terms of its focus, participants, and research setting. While the previous studies mainly examined the effectiveness of ASR applications in improving pronunciation performance, this study focuses on the use of ASR, specifically Google Gemini, toward learning pronunciation skills among eleventh-grade students at SMK Pembangunan Bukittinggi. In addition, this study aims to explore students' experiences and perceptions regarding the use of ASR in pronunciation learning within the context of an Indonesian vocational high school.

C. Conceptual Framework



Figures 2.1 Conceptual Framework

CHAPTER III

RESEARCH METHOD

A. Research Design

This section presented the research design of the study. A research design was important to conduct research systematically and to obtain accurate and relevant information while minimizing time and effort. This research used a quantitative research approach with a survey research design. Quantitative research focuses on collecting numerical data and analyzing them statistically to describe a particular phenomenon (Creswell & Creswell, 2018).

This research used a quantitative survey research design because the writer intended to identify the kinds of Automatic Speech Recognition (ASR) applications used by students and to describe students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. A quantitative survey research design was considered appropriate because it enabled the writer to collect numerical data from respondents through questionnaires and analyze the data using descriptive statistics to provide an objective description of the research findings (Fraenkel et al., 2012).

According to quantitative research principles, a quantitative survey research design aims to collect numerical data from a sample of participants without manipulating variables or providing experimental treatment. The writer collected data directly from the respondents through questionnaires to describe their perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The collected data were analyzed using

descriptive statistical analysis to present the findings in the form of frequencies, percentages, and interpretation categories (Creswell & Creswell, 2018).

This research used a quantitative survey research design because it focused on identifying the kinds of Automatic Speech Recognition (ASR) applications used by students and analyzing students' perceptions toward learning pronunciation skills by using ASR at the eleventh-grade students of SMK Pembangunan Bukittinggi. Through this design, the writer was able to obtain quantitative data regarding students' perceptions and describe the findings systematically based on the research objectives.

B. Research Setting

This research was conducted at SMK Pembangunan Bukittinggi, located in Bukittinggi, West Sumatra, Indonesia. The school was selected as the research site because it provides English language instruction and supports the integration of technology in the learning process. The participants of this study were eleventh-grade students who were learning English as a foreign language (EFL).

The study focused on the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, as tools to support students' pronunciation learning. The research was conducted from July 1st to July 13th, 2026, during the 2025/2026 academic year. The research schedule was carried out based on the recommendation of the English teacher at SMK

Pembangunan Bukittinggi, who determined that the selected period was appropriate for conducting the study without disrupting the teaching and learning process. The writer chose this setting because pronunciation is one of the essential components of English learning, and many students still experience difficulties in producing accurate English sounds. Therefore, the implementation of ASR technology was expected to provide opportunities for students to practice pronunciation independently and receive immediate feedback on their speech production.

During the research period, students were given opportunities to use Google Gemini and ChatGPT in pronunciation learning activities. The writer distributed questionnaires to collect data on students' perceptions regarding the use of Automatic Speech Recognition (ASR) in learning pronunciation. The findings of this study were expected to provide insights into how ASR technology can support pronunciation skill development among eleventh-grade students at SMK Pembangunan Bukittinggi.

C. Population And Sample

The population of this research consisted of all eleventh-grade students of SMK Pembangunan Bukittinggi in the 2025/2026 academic year. The population included students from six study programs, namely Teknik Alat Berat (TAB), Teknik Pemesinan (TPM), Teknik dan Bisnis Sepeda Motor (TBSM), Teknik Komputer dan Jaringan (TKJ), Teknik Sepeda Motor (TSM), and other eleventh-grade classes at SMK Pembangunan Bukittinggi. These students were selected as the population because they had participated in

English learning activities and had experience using Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

Table 3.1 Population

No	Class	Number of Students
1.	XI TAB	24
2.	XI TPM	21
3.	XI TBSM	20
4.	XI TKJ	19
5.	XI TSM	19
TOTAL		103

The sample of this research was the eleventh-grade students of the XI Teknik Alat Berat (TAB) class. The sample was selected using the random sampling technique, which provides every member of the population with an equal opportunity to be selected as the research sample (Creswell & Creswell, 2018). Through this technique, the researcher selected one class randomly to represent the population. The XI TAB class was chosen as the research sample because it represented the characteristics of the population and had experience using Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

. Based on the explanation above, it can be concluded that the population of this research consisted of all 103 eleventh-grade students of SMK Pembangunan Bukittinggi from the TAB, TPM, TBSM, TKJ, and TSM study programs. Meanwhile, the sample consisted of 24 students from class XI TAB, who were selected using the random sampling technique. The sample was considered representative of the population because the selected students had participated in English learning activities and had experience using Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

D. Instrument and Technique of Data Collection

A questionnaire was employed as the primary instrument for collecting data in this research. According to Siyoto and Sodik (2015), a questionnaire is a research instrument consisting of a series of written statements or questions designed to obtain information from respondents in a systematic manner. In this study, the questionnaire was constructed by the writer based on the research objectives and variables, namely the kinds of Automatic Speech Recognition (ASR) applications used by students and students' perceptions toward the use of ASR in learning pronunciation skills. The questionnaire items were developed based on the cognitive, affective, and conative aspects proposed by Bu and Go (2018).

The questionnaire was administered to all eleventh-grade students of SMK Pembangunan Bukittinggi. It consisted of statements intended to gather information about students' experiences, opinions, attitudes, and responses

regarding the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in pronunciation learning. The responses obtained from the questionnaire provided quantitative data that reflected students' perceptions and identified the ASR applications used during pronunciation learning activities.

To measure the respondents' perceptions, the writer used a five-point Likert Scale. This scale was selected because it allows respondents to express the extent of their agreement with each statement related to the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The response categories consisted of Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). Each response category was assigned a numerical score to facilitate quantitative data analysis.

Table 3.2 Score Criteria of Questionnaire

Statement Answer	Score
Strongly Agree (SA)	5
Agree (A)	4
Nutral (N)	3
Disagree (D)	2
Strongly Disagree (SD)	1

E. Validity Dan Reliability

The validity and reliability tests were conducted to ensure that the questionnaire was appropriate and consistent for measuring the research variables. Validity refers to the extent to which an instrument measures what it is intended to measure, while reliability refers to the consistency of the instrument in producing stable and dependable results (Kothari, 2004).

In this research, the content validity of the questionnaire was assessed using Aiken's V Index. Each questionnaire item was evaluated by three expert validators using Aiken's V formula to determine the level of agreement regarding the relevance and appropriateness of each item. Aiken's V is widely used to measure the content validity of research instruments based on expert judgment. The formula is presented as follows:

$$\mathbf{V} = \frac{\sum \mathbf{S}}{\mathbf{n}(\mathbf{c}-\mathbf{1})}$$

Descriptions:

V= rater agreement index

∑S = the score assigned each rater minus the lowest score in the category

n = number of raters

c = the number of categories a rater can select

The results of the validity test for each item were analyzed using Microsoft Excel and can be seen in the following table.

Table 3.3 Index Icon Scores Each Item

Aspects	Validator			S1	S2	S3	∑S	n(c-1)	V	Description
	I	II	II							
Item _1	0	4	4	-1	3	3	5	9	0.555555556	Sedang
Item _2	3	4	4	2	3	3	8	9	0.888888889	Tinggi
Item _3	4	4	3	3	3	2	8	9	0.888888889	Tinggi
Item _4	4	4	4	3	3	3	9	9	1	Tinggi
Item _5	4	4	4	3	3	3	9	9	1	Tinggi
Item _6	4	4	4	3	3	3	9	9	1	Tinggi
Item _7	4	4	4	3	3	3	9	9	1	Tinggi

From the description of the table above, the following results are obtained:

Table 3.4 Results of Index Icon Scores

Aspects	Validator			S1	S2	S3	∑S	V	Description
	I	II	II						
Item _1-7	23	28	27	16	21	20	57	0.904761905	valid

According to the table above, the Aiken's V Index scores were obtained by calculating the agreement among the expert validators to determine the validity of each questionnaire item. The calculation was based on the scores given by the validators to assess the relevance and appropriateness of the instrument. The results of the Aiken's V Index indicated the extent to which each questionnaire item represented the research variables and measured what it was intended to measure. Therefore, the Aiken's V Index was used to determine the content validity of the research instrument.

Table 3.5 Score Criteria of Validity Test

- If V score < 0.6, then the data is not valid
- If V score ≥ 0.6, then the data is valid

The validators in this research were three expert lecturers from UIN Sjech M. Djamil Djambek Bukittinggi. The validity test was conducted to determine whether the questionnaire accurately measured the variables under investigation. According to Kothari (2004), validity was one of the most important criteria in research because it indicated the extent to which an instrument measured what it was intended to measure.

In this research, the writer used content validity to evaluate the relevance and appropriateness of each questionnaire item in measuring the kinds of Automatic Speech Recognition (ASR) applications used by students and students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The content validity was assessed through expert judgment using Aiken's V Index.

F. Trustworthiness of Data

The trustworthiness of the data in this research was ensured through the use of a structured questionnaire distributed to all respondents. The questionnaire was developed based on the research objectives and indicators, namely the kinds of Automatic Speech Recognition (ASR) applications used by students and students' perceptions toward the use of ASR in learning pronunciation skills. The questionnaire items were constructed based on the cognitive, affective, and conative aspects proposed by Bu and Go (2018).

Before the questionnaire was administered, it was evaluated by three expert lecturers from UIN Sjech M. Djamil Djambek Bukittinggi to examine the relevance, clarity, and appropriateness of each item. The questionnaire was then revised according to the suggestions provided by the validators. Furthermore, the instrument was tested for validity using Aiken's V Index and for reliability using Cronbach's Alpha to ensure that it consistently measured the intended variables.

After the validation and reliability testing processes were completed, the questionnaire was distributed to all eleventh-grade students of SMK Pembangunan Bukittinggi. The responses obtained from the questionnaire were analyzed quantitatively using descriptive statistical analysis. These procedures ensured that the data collected were accurate, consistent, and trustworthy in describing students' perceptions toward the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

G. Technique of Data analysis

The data obtained from the questionnaire were analyzed to answer the research questions and achieve the objectives of the study. The analysis aimed to identify the kinds of Automatic Speech Recognition (ASR) applications used by the students and to describe students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills at the eleventh-grade students of SMK Pembangunan Bukittinggi. The questionnaire responses were analyzed using descriptive statistical analysis through the following procedures.

1. The writer collected all completed questionnaires from the eleventh-grade students of SMK Pembangunan Bukittinggi and checked each questionnaire to ensure that every item had been answered completely and accurately.
2. The writer classified the respondents' answers based on each questionnaire item and assigned scores according to the five-point Likert Scale. The

scoring process was conducted to measure students' perceptions toward the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills. The scoring criteria are presented in the following table.

3. The writer calculated the percentage of the responses for each questionnaire item by using the formula proposed by Sudijono (2014). The percentage calculation was used to describe the distribution of students' responses regarding the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The formula was as follows:

$$\left(Percentage = \frac{f}{n} x\ 100\%\right)$$

Descriptions:

P = percentage

f = total of respondents' answer

n = total totals of maximal score

4. The writer calculated the mean score and percentage for each questionnaire item by using the formula proposed by Gay and Mills (2016):

$$\mathcal{X} = \frac{\sum \chi}{n}$$

$\mathcal{X}$ = mean score

$\sum \mathcal{X}$ = the sum of the score

n = total of subjects

5. The writer organized the results into a rating scale table to provide numerical interpretations of the respondents' perceptions. The rating scale was used to classify and interpret the percentage results obtained from the questionnaire. The interpretation criteria were presented in the following table:

Table 3.6 Interpretation of Using ASR Toward Learning Pronunciation Skill

Percentage of Rating Scale	Rating Quality
81 – 100%	Very positive
61 – 80%	Positive
41 – 60%	Sufficiently positive
21 – 40%	Negative
0 -20%	Very negative

Source : Adopted from (Roza, 2021)

6. Calculated into the figures of data interpretation.

Finally, the writer interpreted the questionnaire results based on the mean scores and percentages to describe students' perceptions toward the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills. The findings were then used to answer the research questions and achieve the objectives of the study.

CHAPTER IV

FINDING AND DISCUSSION

A. Finding

This chapter presented the findings of the research on students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills at the eleventh-grade students of SMK Pembangunan Bukittinggi. It also described the data analysis conducted to answer the research question: "What kind of Automatic Speech Recognition (ASR) is used by the students in learning pronunciation skill at the eleventh grade students of SMK Pembangunan Bukittinggi? and What are the students' perceptions of learning pronunciation skill by using Automatic Speech Recognition (ASR) at the eleventh grade students of SMK Pembangunan Bukittinggi?"

1. Kinds of ASR Applications

The data for this research were collected from all eleventh-grade students of SMK Pembangunan Bukittinggi. The population consisted of all students enrolled in the eleventh grade during the academic year in which the research was conducted. Since the number of the population was manageable, the writer employed total sampling, in which all eleventh-grade students were selected as the research respondents. This sampling technique enabled the writer to obtain comprehensive data regarding the kinds of Automatic Speech Recognition (ASR) applications used by the students and their perceptions toward the use of Automatic Speech

Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

The scores of the questionnaire on Kinds of ASR Applications towards English learning were collected by the rubric score below:

Table 4.1 Description of the Data of Kinds of ASR Applications towards English learning

No	Indicators	Statements	Answers				
			STS	TS	N	S	SS
1.	Menggunakan Google Gemini	Statement_1	1	0	0	9	13
		Statement_2	0	1	2	9	11
2.	Menggunakan ChatGPT	Statement_3	0	0	2	9	12
		Statement_4	0	0	7	10	6

Based on the data presented in the table above, the results reflected students' responses to the questionnaire designed to investigate their perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The questionnaire measured students' perceptions based on three aspects proposed by Bu and Go (2018), namely cognitive, affective, and conative, as well as identifying the kinds of ASR applications used by the students, particularly Google Gemini and ChatGPT.

The scores for each questionnaire item were calculated based on the frequency of students' responses in each category. Subsequently, the scores were converted into percentages and interpreted according to the predetermined criteria to identify students' perceptions regarding the use

of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills. The data analysis was conducted using the Likert Scale, and the calculation procedures are presented as follows:

Table 4.2 Analysis of the Data of Kinds of ASR Applications towards English learning

Indicators	Statements	Data Analysis				
		Answers	**Respondents**	**Percentage**	**Average**	**Category**
1) Menggunakan Google Gemini	Statement_1	SS	13	57	**89%**	**Very positive**
		S	9	39		
		N	0	0		
		TS	0	0		
		STS	1	4		
	Statement_2	SS	11	48	**86%**	**Very positive**
		S	9	39		
		N	2	9		
		TS	1	4		
		STS	0	0		
2) Menggunakan ChatGPT	Statement_3	SS	12	52	**89%**	**Very positive**
		S	9	39		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_4	SS	6	26	**79%**	**Positive**
		S	10	43		
		N	7	30		
		TS	0	0		
		STS	0	0		

The data in Table 4.2 show the kinds of Automatic Speech Recognition (ASR) applications used by the eleventh-grade students of SMK Pembangunan Bukittinggi in learning pronunciation skills. Based on

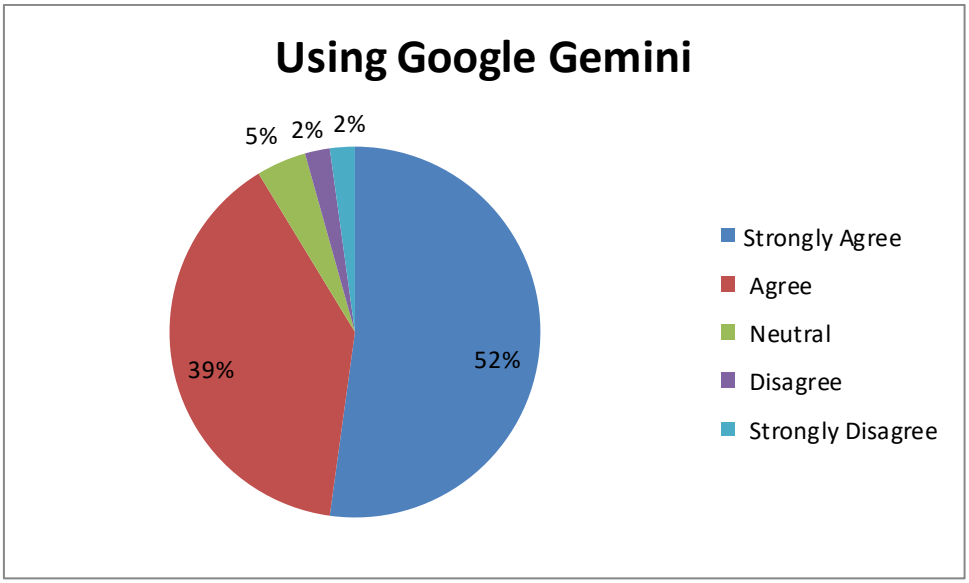
Statement 1, 57% of the respondents strongly agreed and 39% agreed that they used Google Gemini in learning pronunciation, while 4% strongly disagreed. The average score for this statement was 89%, which was categorized as Very Positive. In Statement 2, 48% of the respondents strongly agreed, 39% agreed, 9% were neutral, and 4% disagreed. The average score was 86%, which also fell into the Very Positive category.

Regarding the use of ChatGPT, Statement 3 showed that 52% of the respondents strongly agreed, 39% agreed, and 9% were neutral that they used ChatGPT in learning pronunciation skills. The average score was 89%, which was categorized as Very Positive. Meanwhile, Statement 4 indicated that 26% of the respondents strongly agreed, 43% agreed, and 30% were neutral. The average score for this statement was 79%, which was categorized as Positive.

In conclusion, the findings indicate that both Google Gemini and ChatGPT were used by the students as Automatic Speech Recognition (ASR) applications in learning pronunciation skills. However, Google Gemini received more consistently positive responses than ChatGPT, as reflected in its higher average scores (89% and 86%) compared to ChatGPT (89% and 79%). These findings suggest that students had very positive perceptions toward the use of Google Gemini and positive to very positive perceptions toward the use of ChatGPT for pronunciation learning.

To present the research findings systematically, the writer described the results of each questionnaire item individually by used the pie chart. The explanation was based on the questionnaire scores obtained from the respondents and was supported by the figures presented below:

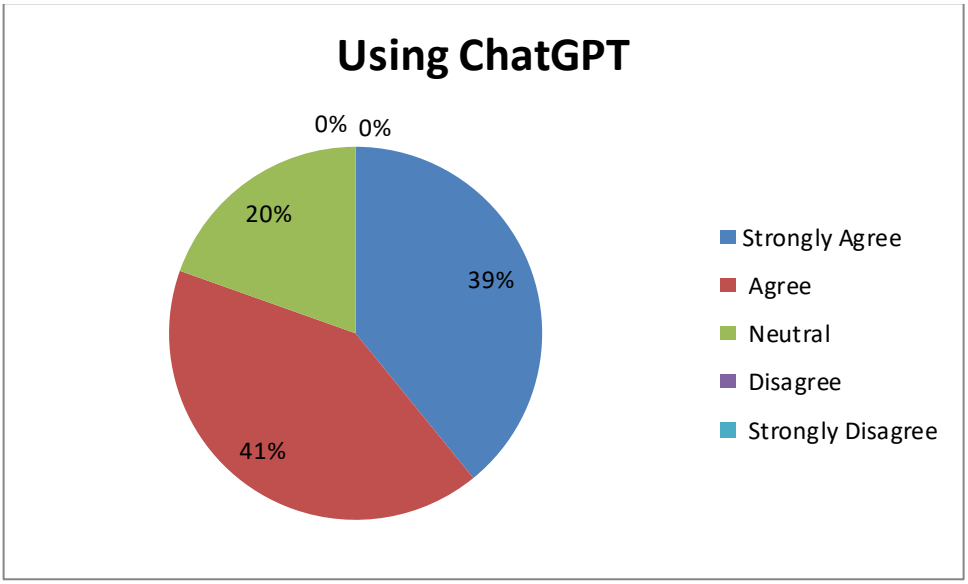
Figure 4.1: Students' Perceptions Using Google Gemini



The result of Figure 4.1 showed that the majority of respondents had positive perceptions toward using Google Gemini. A total of 91% of respondents (52% strongly agreed and 39% agreed) stated that they had positive experiences using Google Gemini. Meanwhile, 5% of the respondents gave a neutral response. Only a small proportion of respondents expressed negative perceptions, with 4% in total (2% disagreed and 2% strongly disagreed). These findings indicate that most students viewed Google Gemini positively and considered it beneficial for supporting their English learning activities.

Based on the data, it can be concluded that students' perceptions of using Google Gemini are generally positive. The overwhelming percentage of students who selected strongly agree and agree demonstrates that Google Gemini is well accepted as a learning tool. Nevertheless, the presence of a few neutral and negative responses suggests that some students may still experience challenges or have different preferences when using Google Gemini, indicating the need for further support to ensure its effective use for all learners.

Figure 4.2: Students' Perceptions in Using ChatGPT



The data presented in Figure 4.2 show that the majority of respondents had positive perceptions of using ChatGPT. A total of 80% of the respondents (39% strongly agreed and 41% agreed) expressed positive opinions about using ChatGPT for learning. Meanwhile, 20% of the respondents selected the neutral option. Notably, none of the respondents chose either the disagree or strongly disagree categories, indicating that

there were no negative perceptions toward the use of ChatGPT among the participants.

Based on these findings, it can be concluded that students generally have positive perceptions of using ChatGPT as a learning tool. The high percentage of positive responses suggests that ChatGPT is considered useful and supportive in students' learning activities. Furthermore, the absence of negative responses indicates a high level of acceptance of ChatGPT. However, the presence of neutral responses suggests that some students may still be uncertain about its benefits or have limited experience using the platform, highlighting the need for continued guidance to maximize its educational potential.

2. Students' Perception in Using ASR for Pronunciation Skill

The data in this section were obtained from 24 eleventh-grade students of class XI TAB at SMK Pembangunan Bukittinggi, who were selected using the random sampling technique. The questionnaire was used to collect data on students' perceptions toward the use of Automatic Speech Recognition (ASR), particularly cognitive aspects, affective aspects, and conative aspects, in learning pronunciation skills. The results were analyzed using descriptive statistical analysis and are presented below:

Table 4.3 Description of the Data of Students' Perception in Using ASR for Pronunciation Skill

No	Indicators	Statements	Answers				
			STS	TS	N	S	SS
1.	Aspek Kognitif	Statement_5	0	1	0	8	14
		Statement_6	0	0	2	10	11
		Statement_7	0	0	2	7	14
		Statement_8	0	0	5	11	7
2.	Aspek Afektif	Statement_9	0	1	0	7	15
		Statement_10	1	0	1	10	11
		Statement_11	0	0	4	6	13
		Statement_12	0	0	2	12	9
3.	Aspek Konatif	Statement_13	0	1	3	8	11
		Statement_14	0	0	2	14	7
		Statement_15	0	0	0	4	19
		Statement_16	1	0	3	9	10

Regarding the cognitive aspect, most respondents agreed that Automatic Speech Recognition (ASR) helped them improve their understanding of English pronunciation by providing immediate feedback and increasing their awareness of pronunciation accuracy. For the affective aspect, most students expressed positive feelings toward the use of ASR, made pronunciation learning more interesting, enjoyable, and motivating. In terms of the conative aspect, the majority of respondents showed a willingness to continue using ASR applications for pronunciation practice and demonstrated positive behavioral intentions toward integrating the technology into their English learning activities.

The scores for each questionnaire item were calculated based on the frequency of students' responses in each category. Subsequently, the scores were converted into percentages and interpreted according to the

predetermined criteria to identify students' perceptions regarding the use of Automatic Speech Recognition (ASR), particularly cognitive aspects, affective aspects, and conative aspects, in learning pronunciation skills. The data analysis was conducted using the Likert Scale, and the calculation procedures are presented as follows:

Table 4.4 Data Analysis of Students' Perception in Using ASR for Pronunciation Skill

Indicators	Statements	Data Analysis				
		Answers	**Respondents**	**Percentage**	**Average**	**Category**
1) Aspek Kognitif	Statement_5	SS	14	61	**90%**	**Very positive**
		S	8	35		
		N	0	0		
		TS	1	4		
		STS	0	0		
	Statement_6	SS	11	48	**88%**	**Very positive**
		S	10	43		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_7	SS	14	61	**90%**	**Very positive**
		S	7	30		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_8	SS	7	30	**82%**	**Very positive**
		S	11	48		
		N	5	22		
		TS	0	0		
		STS	0	0		
2) Aspek Afektif	Statement_9	SS	15	65	**91%**	**Very positive**
		S	7	30		
		N	0	0		
		TS	1	4		
		STS	0	0		

	Statement_10	SS	11	48	**86%**	**Very positive**
		S	10	43		
		N	1	4		
		TS	0	0		
		STS	1	4		
	Statement_11	SS	13	57	**88%**	**Very positive**
		S	6	26		
		N	4	17		
		TS	0	0		
		STS	0	0		
	Statement_12	SS	9	39	**86%**	**Very positive**
		S	12	52		
		N	2	9		
		TS	0	0		
		STS	0	0		
3) Aspek Konatif	Statement_13	SS	11	48	**85%**	**Very positive**
		S	8	35		
		N	3	13		
		TS	1	4		
		STS	0	0		
	Statement_14	SS	7	30	**84%**	**Very positive**
		S	14	61		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_15	SS	19	83	**97%**	**Very positive**
		S	4	17		
		N	0	0		
		TS	0	0		
		STS	0	0		
	Statement_16	SS	10	43	**83%**	**Very positive**
		S	9	39		
		N	3	13		
		TS	0	0		
		STS	1	4		

The data in Table 4.4 presented the analysis of students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills based on the cognitive, affective, and conative aspects. Overall, the results indicate that students had very positive perceptions, with an average percentage score of 87%, which falls into the Very Positive category.

From the cognitive aspect, Statements 5 to 8 obtained average scores of 90%, 88%, 90%, and 82%, respectively. All statements were categorized as Very Positive. These findings indicate that most students believed that the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, helped them improve their understanding of English pronunciation, identify pronunciation errors, and enhance their pronunciation skills through immediate feedback.

In the affective aspect, Statements 9 to 12 obtained average scores of 91%, 86%, 88%, and 86%, respectively, all of which were categorized as Very Positive. These results suggest that students had positive feelings toward the use of ASR in pronunciation learning. Most respondents agreed that learning with Google Gemini and ChatGPT was enjoyable, motivating, and increased their confidence in practicing English pronunciation.

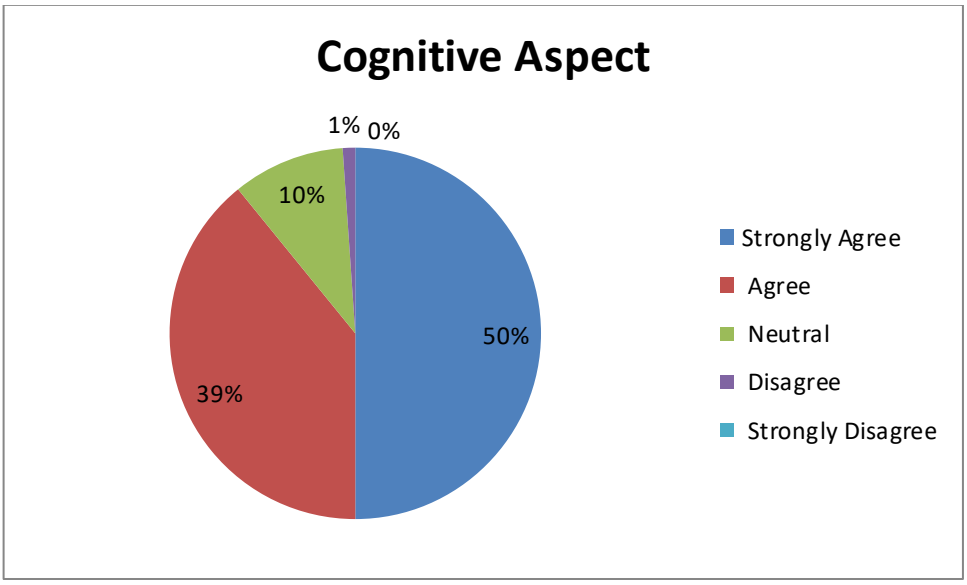
Regarding the conative aspect, Statements 13 to 16 obtained average scores of 85%, 84%, 97%, and 83%, respectively. All statements

were also categorized as Very Positive. These findings indicate that students showed positive behavioral intentions toward the use of ASR. Most students expressed their willingness to continue using Google Gemini and ChatGPT for practicing pronunciation, both inside and outside the classroom.

In conclusion, the findings demonstrate that students had very positive perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The overall average score of 87% indicates that Google Gemini and ChatGPT were positively perceived as effective learning tools that support students' pronunciation awareness, motivation, confidence, and willingness to practice English pronunciation independently.

To ensure a systematic presentation of the research findings, the writer analyzed and described the results of each questionnaire item individually by used the pie chart. The explanation was based on the questionnaire data obtained from the respondents and supported by the figures presented below:

Figure 4.3: Students' Perceptions in Cognitive Aspect

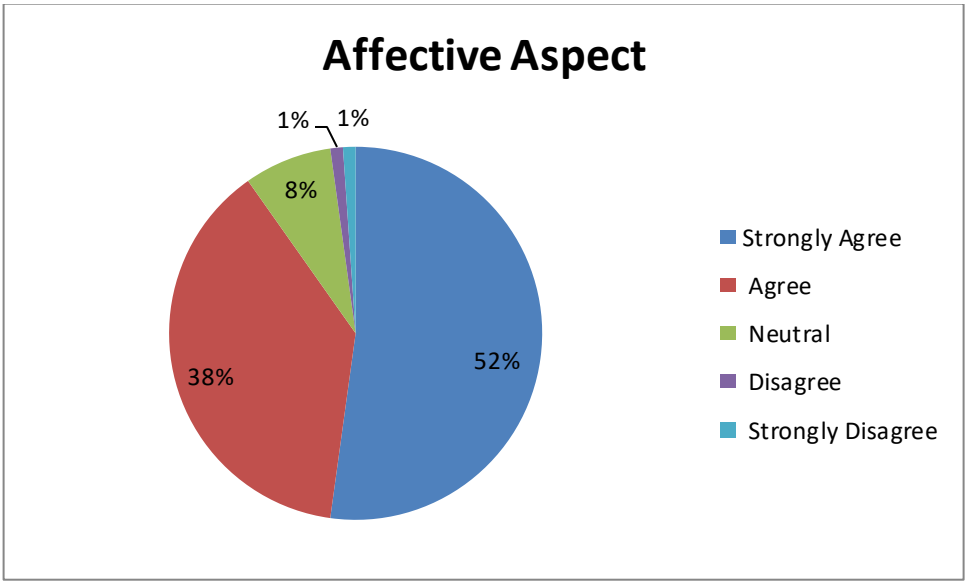


The data presented in Figure 4.3 indicate that students showed highly positive perceptions in the cognitive aspect of using AI tools for learning English. A total of 89% of the respondents (50% strongly agreed and 39% agreed) believed that using AI tools positively contributed to their cognitive learning experiences. Meanwhile, 10% of the respondents gave neutral responses, while only 1% disagreed. No respondents selected the strongly disagree category, indicating that negative perceptions were almost nonexistent.

Based on these findings, it can be concluded that students generally perceive AI tools as beneficial in supporting the cognitive aspect of English learning. The high percentage of positive responses suggests that AI tools help students understand learning materials, improve comprehension, and facilitate the learning process. Although a small number of students expressed neutral or negative perceptions, the results

findings demonstrate that AI tools are widely accepted as effective learning support in enhancing students' cognitive development.

Figure 4.4: Students' Perceptions in Affective Aspect

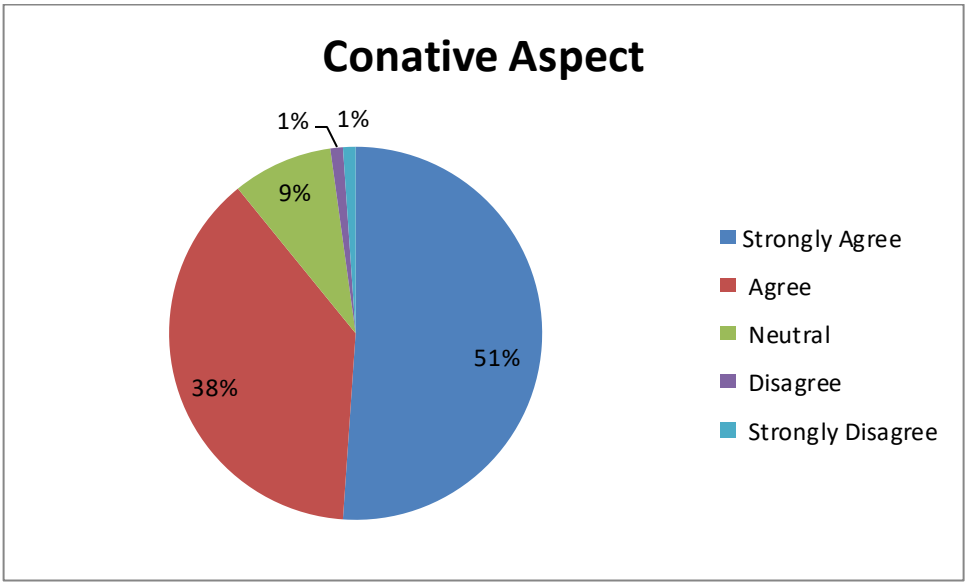


The results presented in Figure 4.4 indicate that students had highly positive perceptions in the affective aspect of using AI tools for English learning. A total of 90% of the respondents (52% strongly agreed and 38% agreed) expressed positive feelings toward using AI tools in their learning process. Meanwhile, 8% of the respondents selected the neutral option, while only 2% (1% disagreed and 1% strongly disagreed) expressed negative perceptions. These findings suggest that AI tools create a positive emotional learning experience for most students.

Based on these findings, it can be concluded that students generally have positive affective perceptions of using AI tools in learning English. The high percentage of positive responses indicates that AI tools help

students feel more confident, interested, and comfortable during the learning process. Although a small proportion of students remained neutral or expressed negative perceptions, the results demonstrate that AI tools positively influence students' attitudes and emotional engagement in English learning.

Figure 4.5: Students' Perceptions in Conative Aspect



The results presented in Figure 4.5 indicate that students had positive perceptions in the conative aspect of using AI tools for English learning. A total of 89% of the respondents (51% strongly agreed and 38% agreed) expressed positive behavioral intentions toward using AI tools in their learning activities. Meanwhile, 9% of the respondents selected the neutral option, while only 2% (1% disagreed and 1% strongly disagreed) expressed negative perceptions. These findings suggest that most students are willing to continue using AI tools to support their English learning.

Based on these findings, it can be concluded that students generally have positive conative perceptions of using AI tools. The high percentage of positive responses indicates that students are motivated to continue using AI tools, recommend them to others, and integrate them into their learning activities. Although a small proportion of respondents remained neutral or expressed negative perceptions, the results demonstrate that AI tools have a positive influence on students' behavioral intentions and their willingness to adopt technology in English learning.

Table 4.5 presented the analysis of students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The questionnaire consisted of sixteen statements representing five indicators: the use of Google Gemini, the use of ChatGPT, the cognitive aspect, the affective aspect, and the conative aspect. The responses from 24 eleventh-grade students were analyzed using percentage analysis and classified based on the predetermined interpretation criteria. The results of each statement are presented as follows:

Table 4.5 Data Analysis of Kinds of ASR Applications and Students' Perception in Using ASR for Pronunciation Skill

Indicators	Statements	Data Analysis				
		Answers	Respondents	Percentage	Average	Category
1) Menggunakan Google Gemini	Statement_1	SS	13	57	**89%**	**Very positive**
		S	9	39		
		N	0	0		
		TS	0	0		
		STS	1	4		
	Statement_2	SS	11	48	**86%**	**Very**

		S	9	39		**positive**
		N	2	9		
		TS	1	4		
		STS	0	0		
2) Menggunakan ChatGPT	Statement_3	SS	12	52	**89%**	**Very positive**
		S	9	39		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_4	SS	6	26	**79%**	**Positive**
		S	10	43		
		N	7	30		
		TS	0	0		
		STS	0	0		
3) Aspek Kognitif	Statement_5	SS	14	61	**90%**	**Very positive**
		S	8	35		
		N	0	0		
		TS	1	4		
		STS	0	0		
	Statement_6	SS	11	48	**88%**	**Very positive**
		S	10	43		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_7	SS	14	61	**90%**	**Very positive**
		S	7	30		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_8	SS	7	30	**82%**	**Very positive**
		S	11	48		
		N	5	22		
		TS	0	0		
		STS	0	0		
4) Aspek Afektif	Statement_9	SS	15	65	**91%**	**Very positive**
		S	7	30		
		N	0	0		
		TS	1	4		
		STS	0	0		

	Statement_10	SS	11	48	**86%**	**Very positive**
		S	10	43		
		N	1	4		
		TS	0	0		
		STS	1	4		
	Statement_11	SS	13	57	**88%**	**Very positive**
		S	6	26		
		N	4	17		
		TS	0	0		
		STS	0	0		
	Statement_12	SS	9	39	**86%**	**Very positive**
		S	12	52		
		N	2	9		
		TS	0	0		
		STS	0	0		
5) Aspek Konatif	Statement_13	SS	11	48	**85%**	**Very positive**
		S	8	35		
		N	3	13		
		TS	1	4		
		STS	0	0		
	Statement_14	SS	7	30	**84%**	**Very positive**
		S	14	61		
		N	2	9		
		TS	0	0		
		STS	0	0		
	Statement_15	SS	19	83	**97%**	**Very positive**
		S	4	17		
		N	0	0		
		TS	0	0		
		STS	0	0		
	Statement_16	SS	10	43	**83%**	**Very positive**
		S	9	39		
		N	3	13		
		TS	0	0		
		STS	1	4		
SUM					**1393%**	
Average					**87%**	**Very positive**

Based on the table above, all of the percentages of the data were gotten by using the following formula:

$$\left(Percentage = \frac{f}{n} x\ 100\%\right)$$

Descriptions:

P = percentage

f = total of respondents' answer

n = total totals of maximal score

For example, the statements number 1:

Answers	**Respondents**	**Count**	**Score**
Strongly agree	13	13 x 5	65
Agree	9	9x 4	36
Neutral	0	0x 3	0
Disagree	0	0x 2	0
Strongly disagree	1	1x 1	1
Score			102

Then:

$$\left(Percentage = \frac{f}{n} x\ 100\%\right)$$

$$\left(\frac{102}{115} x\ 100\%\right) = (89\%)$$

After getting the percentage of items, this research used formula to find the average as follows:

$$\bar{x} = \frac{\sum x}{n}$$

x = mean score

∑x = the sum of the score

n = total of subjects

$$x = \left(\frac{1393\%}{16}\right) = 87\%$$

Furthermore, the results questionnaire results were calculated and interpreted using the predetermined data interpretation table to determine students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills at the eleventh-grade students of SMK Pembangunan Bukittinggi.

Table 4.6 Interpretation of the Questionnaire Result

Rating Scale	Percentage	Rating Quality
81 – 100 %		Very positive
61 – 80 %	87%	Positive
41 – 60 %		Sufficiently positive
21 – 40 %		Negative
0 – 20 %		Very negative

Sources: Adopted from (Roza, 2021)

Based on Table 4.6, the overall average percentage score of students' perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills was 87%. According to the interpretation scale adopted from Roza (2021), this percentage falls within the range of 81–100%,

which is categorized as "Very Positive." These findings indicate that the eleventh-grade students of SMK Pembangunan Bukittinggi had very positive perceptions toward the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills.

B. DISCUSSION

The findings of this study revealed that the eleventh-grade students of SMK Pembangunan Bukittinggi generally had positive perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. The questionnaire measured students' perceptions based on the cognitive, affective, and conative aspects, as well as the kinds of ASR applications used by the students, particularly Google Gemini and ChatGPT.

Regarding the kinds of ASR applications based on Artificial Intelligence (AI), the results showed that 91% of the students had positive perceptions toward using Google Gemini for pronunciation practice (52% strongly agree and 39% agree), while 80% had positive perceptions toward using ChatGPT (39% strongly agree and 41% agree). According to Sugandhamani et al. (2026), Google Gemini supports pronunciation learning through speech recognition and conversational features that enable students to practice speaking, receive immediate responses, identify pronunciation errors, and improve their fluency and speaking confidence. Similarly, ChatGPT facilitates pronunciation learning through interactive voice conversations, allowing students to receive instant feedback, practice various speaking topics, and develop greater pronunciation awareness while promoting autonomous

learning. These findings indicate that both Google Gemini and ChatGPT were perceived as useful AI-powered ASR applications because they provided opportunities for repeated pronunciation practice, immediate feedback, and independent learning (Sugandhamani et al., 2026).

Regarding the cognitive aspect, the results indicated that 89% of the students agreed that Automatic Speech Recognition (ASR) helped them improve their understanding of English pronunciation (50% strongly agree and 39% agree). They perceived that the immediate feedback provided by Google Gemini and ChatGPT enabled them to recognize and correct pronunciation errors more effectively. Meanwhile, 1% of the respondents showed negative responses, suggesting that some students still experienced difficulties in understanding or utilizing the feedback provided by the applications.

In terms of the affective aspect, 90% of the students expressed positive feelings toward the use of ASR in pronunciation learning (52% strongly agree and 38% agree). They reported that learning pronunciation through Google Gemini and ChatGPT was enjoyable, motivating, and increased their confidence in speaking English. However, 2% of the respondents expressed less positive attitudes (1% disagree and 1% strongly disagree), indicating that some students encountered challenges or felt less comfortable when using the applications.

For the conative aspect, the findings showed that 89% of the students demonstrated positive behavioral intentions toward using ASR (51% strongly

agree and 38% agree). Most respondents indicated their willingness to continue using Google Gemini and ChatGPT to improve their pronunciation skills, both inside and outside the classroom. On the other hand, 2% of the students showed lower intentions to continue using ASR (1% disagree and 1% strongly disagree), suggesting that several factors, such as technical limitations or personal preferences, might have influenced their responses.

It can be concluded, the findings indicate that students had positive perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills across the cognitive, affective, and conative aspects. According to Jihye et al. (2018), the cognitive aspect reflects students' knowledge and understanding of ASR as a pronunciation learning tool, the affective aspect relates to their feelings, motivation, confidence, and attitudes toward using ASR, while the conative aspect refers to their willingness and behavioral intention to continue using ASR for pronunciation practice. In this study, 89% of the students showed positive cognitive perceptions, 90% demonstrated positive affective perceptions, and 89% expressed positive conative perceptions toward the use of Google Gemini and ChatGPT. These findings suggest that ASR applications not only enhanced students' understanding of pronunciation but also increased their motivation, confidence, and willingness to continue practicing English pronunciation independently, supporting the theoretical framework proposed by Jihye et al. (2018).

It can be concluded, the findings indicated that students had positive perceptions toward the use of Automatic Speech Recognition (ASR), particularly Google Gemini and ChatGPT, in learning pronunciation skills. Based on the questionnaire results, the results percentage score was 89%, which fell into the "Positive" category according to the predetermined interpretation scale. These findings suggested that the use of ASR supported students in improving their pronunciation skills, increased their motivation to practice English, and encouraged autonomous learning among the eleventh-grade students of SMK Pembangunan Bukittinggi.

CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

Based on the findings of this research, it was concluded that the eleventh-grade students of SMK Pembangunan Bukittinggi generally had positive perceptions toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills. This conclusion was supported by the results obtained from the questionnaire, which measured students' perceptions through the cognitive, affective, and conative aspects, as well as identified the kinds of ASR applications used, particularly Google Gemini and ChatGPT.

From the cognitive aspect, 89% of the students agreed (50% strongly agree and 39% agree) that the use of ASR helped them improve their understanding of English pronunciation by providing immediate feedback and increasing their awareness of pronunciation accuracy. Regarding the affective aspect, 90% of the students (52% strongly agree and 38% agree) expressed positive feelings toward the use of ASR because it made pronunciation practice more enjoyable, interesting, and motivating. Meanwhile, 89% of the students (51% strongly agree and 38% agree) demonstrated positive conative responses, indicating their willingness to continue using ASR applications for practicing English pronunciation in the future.

The findings also revealed that 91% of the students (52% strongly agree and 39% agree) had positive perceptions toward the use of Google Gemini, while 80% (39% strongly agree and 41% agree) had positive

perceptions toward the use of ChatGPT in pronunciation learning. These applications provided opportunities for repeated pronunciation practice, immediate responses, and independent learning, which supported the improvement of students' pronunciation skills.

It can be concluded, the questionnaire results showed that students had a positive perception toward the use of Automatic Speech Recognition (ASR) in learning pronunciation skills, with a results percentage score of 88%, which was categorized as "Positive" based on the predetermined interpretation scale. These findings indicated that the use of Google Gemini and ChatGPT contributed positively to pronunciation learning by increasing students' motivation, confidence, and opportunities to practice English independently.

B. Suggestion

Based on the findings of this research, the writer would like to offer the following suggestions:

1. For English Teachers

English teachers were encouraged to integrate Automatic Speech Recognition (ASR) applications, particularly Google Gemini and ChatGPT, into pronunciation learning activities. These applications could provide students with more opportunities to practice pronunciation independently, receive immediate feedback, and improve their speaking confidence. Teachers were also expected to provide appropriate guidance

to help students use the applications effectively during the learning process.

2. For Students

Students were encouraged to utilize Automatic Speech Recognition (ASR) applications, especially Google Gemini and ChatGPT, as supplementary learning tools for practicing English pronunciation. Through regular practice and active engagement with these applications, students could improve their pronunciation accuracy, speaking confidence, and independent learning skills.

3. For Future Researchers

Future researchers were recommended to conduct further studies on the use of Automatic Speech Recognition (ASR) by involving larger populations, different educational levels, or other English language skills. They were also encouraged to investigate the effectiveness of various ASR applications or employ different research designs to provide more comprehensive findings regarding the implementation of ASR in English language learning.

REFERENCES

Alapján-, V. (2016). 済無No Title No Title No Title, (1995), 1–23.

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). I Can Speak: improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, 1–19. https://doi.org/10.1080/17501229.2024.2315101

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). Teaching pronunciation: A course book and reference guide., *7*(7), 556.

Communication, S., Ragni, A., & Gales, M. J. F. (2018). White Rose Research Online URL for this paper : Automatic Speech Recognition System Development in the " Wild ".

Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal*, *10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

Ghai, W., & Singh, N. (2012). Literature Review on Automatic Speech Recognition. *International Journal of Computer Applications*, *41*(8), 42–50. https://doi.org/10.5120/5565-7646

Gilakjani, A. P. (2018). Abbas Pourhosein Gilakjani 1*. *International Journal of Research in English Education*, (June), 95–108.

Haton, J. P. (2003). Automatic speech recognition: A Review. *ICEIS 2003 - Proceedings of the 5th International Conference on Enterprise Information Systems*, *1*(9), IS5–IS10. https://doi.org/10.5120/9722-4190

Ibrahim, H., & Varol, A. (2020). A Study on Automatic Speech Recognition Systems. *8th International Symposium on Digital Forensics and Security, ISDFS 2020*, *10*(3), 77–85. https://doi.org/10.1109/ISDFS49300.2020.9116286

Jihye, P., Leslie, S., & Sharron, J. L. (2018). Cognitive, affective and conative responses to visual simulation: The effects of rotation in online product presentation. *Journal of Consumer Behaviour*, *50*(October), 35–50. https://doi.org/10.1002/cb

Leong, L., & Ahmadi, S. M. (2017). c r v i h o e f.

Levis, J., & Suvorov, R. (2012). Encyclopedia of Applied Linguistics: Automatic Speech Recognition. *The Encyclopedia of Applied Linguistics*, *1976*. https://doi.org/10.1002/9781405198431.wbeal0066

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French/y/. *CALICO Journal*, *32*(1), 1–25.

https://doi.org/10.1558/cj.v32i1.25962

Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., … Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, *25*(2), 126–140. https://doi.org/10.1504/IJIL.2019.097674

Mahmudi, N., & Anugerahwati, M. (2021). International Journal of Language Teaching and Education. *International Journal of Language Teaching and Education*, *10*(2015), 1–15. https://doi.org/10.22437/ijolte.v10i1.49746

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Muir-Herzig, R. G. (2004). Technology and its impact in the classroom. *Computers and Education*, *42*(2), 111–131. https://doi.org/10.1016/S0360-1315(03)00067-8

N. M. Arafa, M., Elbarougy, R., A. Ewees, A., & Behery, G. M. (2018). A Dataset for Speech Recognition to Support Arabic Phoneme Pronunciation. *International Journal of Image, Graphics and Signal Processing*, *10*(4), 31–38. https://doi.org/10.5815/ijigsp.2018.04.04

Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Prashant, P. D. (2018). Importance of Pronunciation in English Language Communication. *Academic Research in Educational Sciences*, *7*(2), 15–20.

Purwanto, A. (2019). Teaching Pronunciation Using Varieties of Pronunciation Teaching Materials and Practices. *Scope : Journal of English Language Teaching*, *3*(2), 81. https://doi.org/10.30998/scope.v3i2.4129

Rai, N., & Thapa, B. (2019). A study on purposive sampling method in research. *Kathmandu:Kathmandu School of Law*, 1–12.

Rosalia, T., Dayu, A. T., & Aprizani, Y. (2021). A study on teaching English pronunciation in Vocational School Banjarbaru.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., & Barker, P. (1996). Applications of automatic speech recognition to speech and language development in young children. *International Conference on Spoken Language Processing, ICSLP, Proceedings*, *1*, 176–179. https://doi.org/10.21437/icslp.1996-45

S, K., & E, C. (2016). A Review on Automatic Speech Recognition Architecture and Approaches. *International Journal of Signal Processing, Image*

Processing and Pattern Recognition, *9*(4), 393–404. https://doi.org/10.14257/ijsip.2016.9.4.34

Sugandhamani, N., Vairamuthu, R. E., & Selvaraj, L. K. (2026). Artificial Intelligence in Connectivism Learning : A Pathway to Communication Skill Development, *5*(2), 214–227.

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, *14*(August). https://doi.org/10.3389/fpsyg.2023.1210187

Tergujeff, E. (2014). English Pronunciation Teaching: Four Case Studies from Finland. *Journal of Language Teaching and Research*, *3*(4). https://doi.org/10.4304/jltr.3.4.599-607

W, C. J. (2008). *Educational Research Planning, Conducting and Evaluating Quantitative and Qualitative Research*. United States of America.

Yates, L., & Zielinski, B. (2009). *Give it a go: Teaching pronunciation to adults. Australia: AMEP Research.*

Zaki Abdul Aziz. (2023). The Correlation Between Students' Pronunciation And Their English Listening Skills.

APPENDICES

Appendix 1: Research Guideline

Variable	Sub Variable	Indicators	Description
The Use of Automatic Speech Recognition (Asr) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi	1) Kinds of ASR Applications	a. Google Gemini	Google Gemini is an artificial intelligence application that can support pronunciation learning through its speech recognition and conversational features. By interacting with Gemini using voice input, students can practice speaking English and receive responses based on their spoken language.
		b. ChatGPT	ChatGPT is another AI-powered application that can be utilized to support pronunciation learning. Through voice conversation features, students can practice English pronunciation while engaging in interactive communication. ChatGPT allows learners to receive instant responses, practice various speaking topics, and develop greater awareness of pronunciation accuracy.
	2) Students' Perception in Using ASR for Pronunciation Skill	a. Cognitive	The cognitive aspect refers to students' knowledge, understanding, beliefs, and awareness regarding the use of Automatic Speech Recognition (ASR) in learning pronunciation. This aspect focuses on how students perceive the usefulness of ASR, the information and feedback provided by the application, and its role in helping them improve pronunciation accuracy.

		b. Affective	The affective aspect relates to students' feelings, emotions, and attitudes toward the use of ASR in pronunciation learning. It includes students' level of interest, enjoyment, confidence, motivation, and satisfaction when using ASR applications. Positive affective responses indicate that students feel comfortable and encouraged to practice pronunciation, while negative responses may arise from difficulties in using the technology or dissatisfaction with the feedback provided.
		c. Conative	The conative aspect refers to students' intentions, willingness, and actual behaviors related to the use of ASR for pronunciation practice. This aspect reflects students' tendency to continue using ASR, their willingness to engage in pronunciation activities, and their commitment to improving their pronunciation skills through the application. Conative responses are generally influenced by students' cognitive beliefs and affective attitudes toward ASR.

Source: (Sugandhamani et al., 2026) & (Jihye et al., 2018)

Appendix 2: Students' Questionnaire

Angket Penerapan Teknologi Pengenalan Suara Otomatis/ASR (Penggunaan Gemini Dan ChatGPT) Dalam Pembelajaran Pronunciation Siswa

Identitas responden

Nama :

Kelas :

Nama sekolah :

Instruksi:

Mohon baca petunjuk berikut ini dengan seksama sebelum Anda mulai.

1. Mohon baca setiap pernyataan dengan seksama dan tunjukkan tingkat persetujuan Anda dengan memberi tanda centang (✓) pada kotak yang sesuai.
2. Jawaban Anda bersifat anonim dan akan dijaga kerahasiaannya. Tidak ada jawaban yang benar atau salah, jadi mohon jawablah sejujur mungkin berdasarkan pengalaman dan perasaan Anda.
3. Angket ini dibagi menjadi beberapa bagian. Setiap bagian berisi pernyataan yang berkaitan tentang penerapan teknologi pengenalan suara otomatis/ASR (penggunaan Gemini dan chatGPT) dalam pembelajaran pronunciation siswa.
4. Mohon pastikan bahwa Anda menjawab semua pernyataan di setiap bagian. Angket yang tidak lengkap dapat mempengaruhi validitas temuan penelitian.
5. Jika Anda menemukan pernyataan yang kurang jelas atau memiliki pertanyaan mengenai angket, jangan ragu untuk meminta klarifikasi kepada peneliti.
6. Setelah Anda mengisi angket, mohon kirimkan ke tempat pengumpulan yang telah ditentukan atau sesuai instruksi peneliti.

No	Variable	Indikator	Pernyataan	Jawaban				
				STS	TS	N	S	SS
1.	Jenis-jenis Aplikasi ASR	a. Menggunakan Google Gemini	1) Saya senang menggunakan Google Gemini untuk berlatih pengucapan bahasa Inggris.					
			2) Google Gemini membantu saya berlatih pengucapan melalui interaksi suara.					
		b. Menggunakan ChatGPT	3) Saya senang menggunakan ChatGPT untuk berlatih pengucapan bahasa Inggris.					
			4) ChatGPT memberikan kesempatan untuk berlatih pengucapan melalui percakapan interaktif.					
2.	Persepsi Siswa dalam Menggunakan ASR untuk Keterampilan Pengucapan	a. Aspek Kognitif	5) Saya memahami bagaimana aplikasi ASR membantu meningkatkan pengucapan bahasa Inggris.					
			6) Umpan balik yang diberikan oleh ASR membantu saya mengenali kesalahan pengucapan saya.					
			7) Menggunakan ASR meningkatkan pemahaman saya tentang pengucapan bahasa Inggris yang benar.					
			8) ASR membuat latihan pengucapan menjadi lebih mudah daripada metode konvensional.					
		b. Aspek Afektif	9) Saya senang belajar pengucapan menggunakan aplikasi ASR.					
			10) Saya merasa termotivasi untuk berlatih					

			pengucapan menggunakan Google Gemini atau ChatGPT.					
			11) Saya merasa lebih percaya diri saat berbicara dalam bahasa Inggris setelah menggunakan ASR.					
			12) Menggunakan ASR membuat pembelajaran pengucapan menjadi lebih menarik.					
		c. Aspek Konatif	13) Saya berniat untuk terus menggunakan ASR guna meningkatkan pengucapan saya.					
			14) Saya sering berlatih pengucapan bahasa Inggris menggunakan Google Gemini atau ChatGPT di luar kelas.					
			15) Saya bersedia merekomendasikan aplikasi ASR kepada teman-teman sekelas saya.					
			16) Saya akan terus menggunakan ASR karena hal itu membantu meningkatkan pengucapan saya.					

Source: (Sugandhamani et al., 2026) & (Jihye et al., 2018)

Appendix 3: Validation Sheets

LEMBAR VALIDASI INSTRUMENT

THE USE OF AUTOMATIC SPEECH RECOGNITION (ASR) TOWARD LEARNING PRONUNCIATION SKILL AT THE ELEVENTH GRADE STUDENTS OF SMK PEMBANGUNAN BUKITTINGGI

Nama : Reviyolla Herman

Judul Penelitian : The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi

Fakultas/Jurusan : Fakultas Tarbiyah dan Ilmu Keguruan / Pendidikan Bahasa Inggris

Petunjuk :

1. Untuk memberikan penilaian terhadap instrumen "The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi" Bapak/ibu cukup memberikan tanda chek list (□) pada kolom yang di sediakan
2. Angka- angka yang terdapat pada kolom yang dimaksud berarti:
   - 0 : Tidak Valid
   - 1 : Kurang Valid
   - 2 : Cukup Valid
   - 3 : Valid
   - 4 :Sangat Valid
3. Huruf- huruf yang terdapat pada kolom yang di maksud berarti:
   - A : Dapat di gunakan tanpa revisi
   - B : Dapat di gunakan dengan sedikit revisi
   - C : Dapat di gunakan dengan revisi sedang
   - D : Dapat di gunakan dengan banyak revisi
   - F : Tidak dapat di gunakan

A. ASPEK PENILAIAN

No	Aspek Penilaian	Penilaian					Keterangan
		0	1	2	3	4	
1	Aspek Petunjuk						
	a Petunjuk ditanyatakan dengan jelas					✓	

	b Indikator yang diamati pengamat mudah dipahami					✓	
	c Masing-masing indicator dapat dibedakan dengan jelas				✓		perlu diperjelas perbedaan beberapa indikator (cth : 13 & 16, 9 & 12)
2	Aspek Isi						
	a Indikator yang diamati sudah mencakup semua aspek					✓	
	b Kebenaran isi instrumen					✓	
3	Aspek Bahasa						
	a Kalimat yang digunakan berdasarkan kaidah bahasa yang benar					✓	
	b Menggunakan kalimat yang mudah dipahami					✓	

A. PENILAIAN SECARA UMUM

No	Uraian	A	B	C	D	E
1	Penilaian secara umum terhadap instrument		✓			

B. KOMENTAR/ SARAN

Secara umum angket ini sudah dapat digunakan dengan sedikit revisi

Bukittinggi, 30 Juni 2026

Validator



Dr. Afdaleni. M.Pd., M.Pd.
NIP. 196404091989032001

1 of 2

LEMBAR VALIDASI INSTRUMENT

THE USE OF AUTOMATIC SPEECH RECOGNITION (ASR) TOWARD LEARNING PRONUNCIATION SKILL AT THE ELEVENTH GRADE STUDENTS OF SMK PEMBANGUNAN BUKITTINGGI

Nama : Reviyolla Herman

Judul Penelitian : The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi

Fakultas/Jurusan : Fakultas Tarbiyah dan Ilmu Keguruan / Pendidikan Bahasa Inggris

Petunjuk :

1. Untuk memberikan penilaian terhadap instrumen "The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi" Bapak/ibu cukup memberikan tanda chek list (⊔) pada kolom yang di sediakan
2. Angka- angka yang terdapat pada kolom yang dimaksud berarti:
   - 0 : Tidak Valid
   - 1 : Kurang Valid
   - 2 : Cukup Valid
   - 3 : Valid
   - 4 :Sangat Valid
3. Huruf- huruf yang terdapat pada kolom yang di maksud berarti:
   - A : Dapat di gunakan tanpa revisi
   - B : Dapat di gunakan dengan sedikit revisi
   - C : Dapat di gunakan dengan revisi sedang
   - D : Dapat di gunakan dengan banyak revisi
   - F : Tidak dapat di gunakan

A. ASPEK PENILAIAN

No	Aspek Penilaian	Penilaian					Keterangan
		0	1	2	3	4	
1	Aspek Petunjuk						
	a Petunjuk ditanyatakan dengan jelas					✓	

Dipindai dengan CamScanner

	b Indikator yang diamati pengamat mudah dipahami					✓	
	c Masing-masing indicator dapat dibedakan dengan jelas					✓	
2	Aspek Isi						
	a Indikator yang diamati sudah mencakup semua aspek					✓	
	b Kebenaran isi instrumen					✓	
3	Aspek Bahasa						
	a Kalimat yang digunakan berdasarkan kaidah bahasa yang benar					✓	
	b Menggunakan kalimat yang mudah dipahami					✓	

A. PENILAIAN SECARA UMUM

No	Uraian	A	B	C	D	E
1	Penilaian secara umum terhadap instrument	✓				

B. KOMENTAR/ SARAN

~~Ada beberapa item yang memiliki unsur kesamaan dalam satu indicator, lebih baik direvisi Kembali.~~

Instrumen ini bisa digunakan.

Bukittinggi, 30 Juni 2026

Validator

Dr. Widya Syafitri, M.Pd

Dipindai dengan CamScanner

1 of 2

LEMBAR VALIDASI INSTRUMENT

THE USE OF AUTOMATIC SPEECH RECOGNITION (ASR) TOWARD LEARNING PRONUNCIATION SKILL AT THE ELEVENTH GRADE STUDENTS OF SMK PEMBANGUNAN BUKITTINGGI

Nama : Reviyolla Herman

Judul Penelitian : The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi

Fakultas/Jurusan : Fakultas Tarbiyah dan Ilmu Keguruan / Pendidikan Bahasa Inggris

Petunjuk :

1. Untuk memberikan penilaian terhadap instrumen "The Use of Automatic Speech Recognition (ASR) toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK Pembangunan Bukittinggi" Bapak/ibu cukup memberikan tanda chek list (☐) pada kolom yang di sediakan
2. Angka- angka yang terdapat pada kolom yang dimaksud berarti:
   - 0 : Tidak Valid
   - 1 : Kurang Valid
   - 2 : Cukup Valid
   - 3 : Valid
   - 4 :Sangat Valid
3. Huruf- huruf yang terdapat pada kolom yang di maksud berarti:
   - A : Dapat di gunakan tanpa revisi
   - B : Dapat di gunakan dengan sedikit revisi
   - C : Dapat di gunakan dengan revisi sedang
   - D : Dapat di gunakan dengan banyak revisi
   - F : Tidak dapat di gunakan

A. ASPEK PENILAIAN

No	Aspek Penilaian	Penilaian					Keterangan
		0	1	2	3	4	
1	Aspek Petunjuk						
	a Petunjuk ditanyatakan dengan jelas	√					Tambahkan petunjuk pd instrumen

Dipindai dengan CamScanner

	b Indikator yang diamati pengamat mudah dipahami				✓		Tambahkan definisi masing2 aspek, tulis dll beda kurung
	c Masing-masing indicator dapat dibedakan dengan jelas					✓	
2	Aspek Isi						
	a Indikator yang diamati sudah mencakup semua aspek					✓	
	b Kebenaran isi instrumen					✓	
3	Aspek Bahasa						
	a Kalimat yang digunakan berdasarkan kaidah bahasa yang benar					✓	
	b Menggunakan kalimat yang mudah dipahami					✓	

$3+20 = \frac{23}{7} = 3,2$

A. PENILAIAN SECARA UMUM

No	Uraian	A (4)	B (3)	C (2)	D (1)	E (0)
1	Penilaian secara umum terhadap instrument		✓ dengan revisi			

B. KOMENTAR/ SARAN

~~Ada beberapa item yang memiliki unsur kesamaan dalam satu indicator, lebih baik~~ direvisi Kembali.

Bukittinggi, 30 Juni 2026

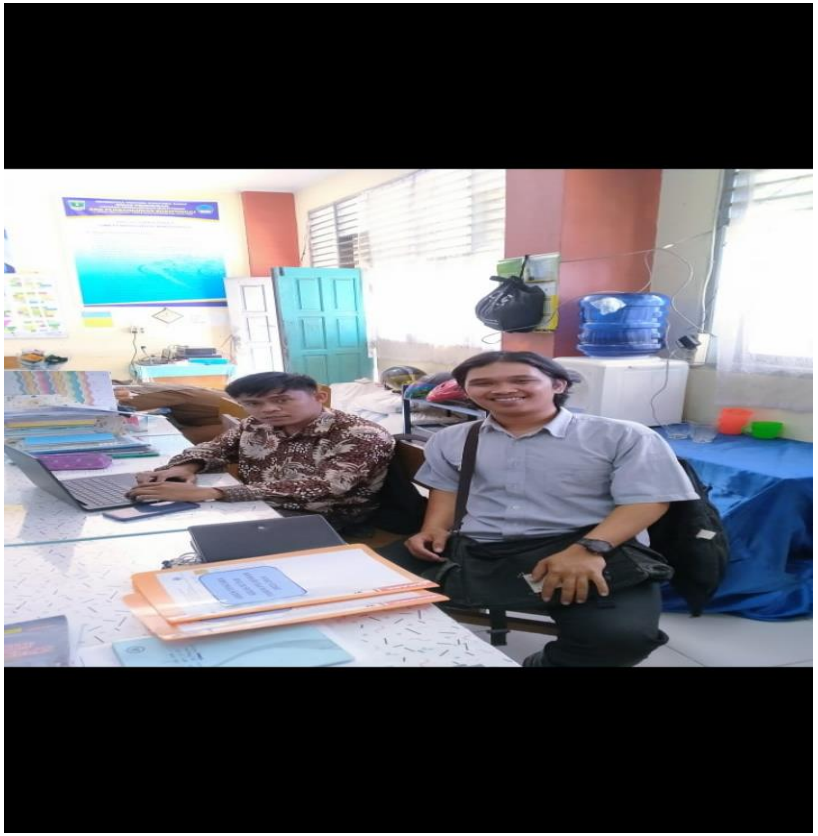
Validator

Rahmiliahma Zulyadi, M.Pd

Dipindai dengan

Appendix 4: Documentations





Appendix 5: Surat Izin Penelitian

KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SJECH M. DJAMIL DJAMBEK BUKITTINGGI
FAKULTAS TARBIYAH DAN ILMU KEGURUAN
Kampus II: Jl. Gurun Aur Kubang Putih - Kabupaten Agam - Sumatera Barat – Telepon/ Fax. (0752) 22875
Website: http://ftik.uinbukittinggi.ac.id | email: ftik@uinbukittinggi.ac.id

Nomor : B-0704/Un.26.1/PP.07/05/2026 Bukittinggi, 29 Juni 2026
Lamp : 1 (Satu) rangkap
Hal : Mohon Izin Penelitian

Kepada YTH : Kepala SMK S Pembangunan Bukittinggi
Bapak/Ibu .....

Assalamu'alaikum Wr. Wb

Dalam rangka mengumpulkan data untuk penyusunan skripsi bagi mahasiswa Fakultas Tarbiyah dan Ilmu Keguruan (FTIK) Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, maka bersama ini kami mohon kepada Bapak/Ibu untuk dapat memberi izin penelitian bagi mahasiswa dibawah ini:

Nama	:	**Revi Yolla Herman**
NIM	:	**2319159**
Semester	:	14
Program Studi	:	Pendidikan Bahasa Inggris
Fakultas	:	Tarbiyah dan Ilmu Keguruan
Judul Skripsi	:	*"The Use of Automatic Speech Recognition (ASR) Toward Learning Pronunciation Skill at the Eleventh Grade Students of SMK pembangunan Bukittinggi"*
Lokasi Penelitian	:	SMK pembangunan Bukittinggi
Waktu	:	Juni 2026 s.d Selesai

Demikian permohonan ini kami sampaikan atas perhatian dan kerjasamanya diucapkan terimakasih.

DEKAN,

Dr. JUNAIDI, M.Pd
NIP 197706162003121003