JPBII, Vol. 13 No. 2 September 2025
p-ISSN : 2615-2800 , e-ISSN : 2615-4404 DOI: 10.23887/jpbi.v13i2.5919

THE CORRELATION BETWEEN EXTENSIVE LISTENING AND PRONUNCIATION OF ENGLISH SHORT VOWELS /ɛ/ AND /æ/

I. Novitasari[1], Pariyanto[2]
[12]English Literature, University of 17 August 1945 Surabaya, Surabaya
e-mail: innenovitasari.in@gmail.com, pariyanto@untag-sby.ac.id

This study aims to investigate the correlation between the frequency of English listening activities and the ability to pronounce English short vowels /ɛ/ and /æ/. A total of 45 non-English major university students participated in this correlational research. Quantitative data were collected through voice recordings analyzed using Praat software, while qualitative data on listening preferences and frequency were obtained through questionnaires and interviews. The data were analyzed using SPSS version 25. The findings revealed that students primarily listened to English through music and movies, with weekly durations ranging from 5–7 hours to 21–28 hours. Despite frequent exposure, many students still produced inaccurate pronunciations of /ɛ/ and /æ/ compared to native-like standards. Statistical analysis showed no significant correlation between English listening frequency and the accuracy of short vowel pronunciation.

Keywords: Correlation; English Listening Frequency; Pronunciation; Short Vowels /ɛ/ and /æ/

1. INTRODUCTION

Speaking is one of the four essential language skills since people who acquire a language are recognized as speakers of that language (Ur, 1996). Chaney (as cited in Resha et al., 2015) defines speaking as the process of creating and communicating meaning by using verbal and nonverbal symbols in various circumstances. Speaking helps students enhance their vocabulary, grammar, and overall communicative competence, which eventually supports other language skills such as writing. Through speaking, learners can express ideas, emotions, and intentions; share stories; make requests; and engage in discussions (Rao, 2019). Consequently, proficient speakers often gain better access to educational and professional opportunities (Baker & Westrup, 2003).

One of the most crucial components of speaking is pronunciation, as it directly affects intelligibility and communication success. Pronunciation refers to how words are articulated and how speech sounds are produced (Hornby, 1987, as cited in Manurung et al., 2024). In English, even slight variations in pronunciation—particularly in vowel sounds—can lead to significant changes in meaning. For instance, mispronouncing similar vowel sounds such as /ɛ/ in bed and /æ/ in bad can cause misunderstanding. Many EFL learners struggle with distinguishing and producing these vowels due to limited exposure to authentic English input and insufficient pronunciation practice (Idayani, 2019).

A growing body of research suggests that extensive listening—listening to English frequently through various media such as songs, podcasts, movies, or YouTube videos—can improve learners' pronunciation accuracy, including the perception and production of specific phonemes. Listening activities provide repeated exposure to authentic sounds, enabling learners to internalize correct pronunciation patterns. According to Upa (Upa et al., 2021) and Supeno (Supeno, 2018), listening to English songs or media can enhance learners' familiarity with sounds and words while also increasing motivation. Similarly, Prambudi (as cited in Hidayatullah, 2018) found that watching movies helps students directly observe and imitate accurate pronunciation models. These findings highlight that frequent listening practices can support the accurate production of difficult English vowels such as /ɛ/ and /æ/.

Several previous studies have examined the relationship between listening activities and pronunciation improvement. Apridayanti (Apridayanti, 2021) found a significant

correlation between students' ability to listen to English songs and their pronunciation mastery. Hidayatullah, (2018) demonstrated that using Western films improved students' pronunciation scores across three learning cycles. Syahabuddin et al. (Syahabuddin & Rizqa, 2021) showed that podcasts effectively enhanced listening skills, and Tahmina (Tahmina, 2023) revealed that YouTube usage positively impacted learners' overall English proficiency. However, most of these studies focused on general pronunciation or listening skills rather than specific vowel sounds. Moreover, the participants were often English department students, while limited research has explored how non-English majors develop pronunciation of challenging vowels such as /ɛ/ and /æ/ through extensive listening.

Therefore, this study aims to fill that gap by investigating whether the frequency of listening to English through various media (songs, movies, podcasts, YouTube) correlates with students' ability to pronounce the English short vowels /ɛ/ and /æ/. This study employs voice recordings and the Praat software to analyze vowel quality using acoustic measures (F1 and F2).

Based on the background above, this research addresses the following questions: (1) How frequently do the participants listen to English, and what sources do they primarily use? ; (2) What are the participants' F1 and F2 scores when pronouncing the English short vowels /ɛ/ and /æ/? ; and (3) Is there a significant correlation between the frequency of listening to English and the ability to pronounce the English short vowels /ɛ/ and /æ/?

2. RESEARCH METHOD

Research Design

This research uses quantitative and qualitative designs. Quantitative research with a correlational design was used to find out how close the participants' pronunciation of English short vowels /ɛ/ and /æ/ were to the standard pronunciation of native speakers' English short vowels /ɛ/ and /æ/, as well as to find a correlation between the participants' frequency of listening to English and the medium used (music, movies, YouTube, Podcasts) and the participants' pronunciation of English short vowels /ɛ/ and /æ/. Meanwhile, participants' perceptions of listening to English through certain medium (music, movies, YouTube, Podcasts) and their reasons for using the medium were examined by qualitative research with a survey design.

Data Types and Sources

There are two types of data in this study: quantitative and qualitative. The total amount of time participants spent listening to English in a day or week using media such as music, movies, Youtube, podcasts is considered quantitative data. Other quantitative data are Praat's analysis of participants' voice recordings pronouncing the English short vowel /ɛ/ (Let, Set, Yet, Get, and Wet) and the English short vowel /æ/ (Tap, Map, Nap, Gap, and Lap) in the form of Formant 1 (F1) and Formant 2 (F2). Meanwhile, qualitative data are questionnaire responses collected in the form of participants' perceptions of the usage of specific media for listening to English and how effective the media is in assisting participants in correctly pronouncing the short English vowels /ɛ/ and /æ/.

Data Collection Tools and Method

Voice recording and Praat software were used to determine the correlation between the frequency of listening to English and the ability to pronounce English short vowels /ɛ/ and /æ/ correctly according to native speakers' pronunciation standards, as well as the correlation between the medium used to listen to English and the ability to pronounce English short vowels /ɛ/ and /æ/ correctly according to native speakers' pronunciation standards.

p-ISSN : 2615-2800 , e-ISSN : 2615-4404 DOI: 10.23887/jpbi.v13i2.5919

Table 1. Items of English Short Vowels /ɛ/ and /æ/

English short vowel /ɛ/	English short vowel /æ/
LET	TAP
SET	MAP
YET	NAP
GET	GAP
WET	LAP

Besides that, questionnaires and interviews were also conducted to determine how frequently the participants listened to English, what mediums they used, participants' perceptions of the usage of specific media for listening to English and how effective the mediums are in assisting participants to pronounce the English short vowels /ɛ/ and /æ/ correctly according to native speakers' pronunciation standards.

Population and Sample

The researcher of this study was an English teacher in an online speaking class at a non-profit organization. Therefore, the population of this study was 45 non-English major participants who took part in the researcher's online speaking class. The participants consisted of 9 men and 36 women with an age range of 18 – 22 years old.

Data Collection Technique

In collecting the data, the researcher collected the data systematically by using the following steps:

1) The researcher coordinated with participants and provided several instructions regarding filling out the questionnaire
2) Participants filled out the questionnaire according to the instructions given by the researcher
3) Participants uploaded their voice recordings of each word from the English short vowel /ɛ/ (Let, Set, Yet, Get, Wet) and English short vowel /æ/ (Tap, Map, Nap, Gap, Lap) at the end of the questionnaire session
4) The researcher conducted short interviews with participants to obtain additional information from the questionnaires that participants had filled out
5) The researcher downloaded and analyzed participants' voice recordings using Praat to obtain Formant 1 (F1) and Formant 2 (F2) scores of each word from the pronunciation of the English short vowels /ɛ/ (Let, Set, Yet, Get, Wet) and /æ/ (Tap, Map, Nap, Gap, Lap)

Data Analysis Technique

In analyzing the data, the researcher employed both qualitative and quantitative approaches through the following steps:

I. Data Review and Completion

The researcher began by carefully reviewing the questionnaire responses. To ensure data completeness and accuracy, follow-up interviews were conducted with selected participants to clarify or supplement their responses.

II. Data Classification

The questionnaire and interview data were classified according to the research variables, namely:

(1) the frequency of listening to English,
(2) the medium used for listening, and
(3) pronunciation performance of English short vowels /ɛ/ and /æ/.

III. Acoustic Data Processing

Participants' voice recordings were collected and analyzed using Praat software to obtain the acoustic measurements of Formant 1 (F1) and Formant 2 (F2) for each target vowel (/ɛ/ in let, set, yet, get, wet and /æ/ in tap, map, nap, gap, lap). These measurements were then compiled in tabular form.

IV. Data Categorization

To facilitate statistical analysis, the variables were categorized into ordinal scales:

1. Listening frequency was categorized from very high to very low based on weekly hours of listening.
2. Listening media were classified by type (e.g., music, movies, YouTube, others).
3. Pronunciation performance (F1 and F2 values) was categorized from very good to poor according to standard formant value ranges.
4. Detailed range values were used internally for analysis but are not presented here for brevity.

V. Statistical Analysis

The categorized data were analyzed using SPSS 25.0 to examine:

1. The correlation between English listening frequency and pronunciation accuracy of the vowels /ɛ/ and /æ/.
2. The correlation between the listening medium and pronunciation accuracy of the vowels /ɛ/ and /æ/.

VI. Interpretation and Presentation

The results of the correlation analyses were presented in both tabular and graphical forms, followed by a comprehensive interpretation to address the research questions and draw conclusions.

Hypothesis

The hypothesis of this study are proposed in terms of null hypotheses (H_0) and alternative hypotheses (H_a). They are as follows:

1. H_a: There is a significant positive correlation between the frequency of listening to English through various media and the ability to accurately pronounce the English short vowels /ɛ/ and /æ/.
2. H_0: There is no significant correlation between the frequency of listening to English and the ability to pronounce the English short vowels /ɛ/ and /æ/.

3. FINDINGS AND DISCUSSION

Participants' frequency of listening to English and the medium used to listen to English

The table below shows the data that the researcher has collected from the questionnaire. The data contains the frequency with which the participants listen to English as well as the medium used to listen to English. The data can be seen below:

Table 2. Participants' frequency of listening to English and the medium used

No	Participant's Name	The medium used to listen to English	Duration of listening to English (Hours/Week)
1	P1	Music	5
2	P2	Music	5
3	P3	Music	5
4	P4	Music	5
5	P5	Music	5
6	P6	Music	5
7	P7	Music	7

No	Participant's Name	The medium used to listen to English	Duration of listening to English (Hours/Week)
8	P8	Movies	7
9	P9	Movies	7
10	P10	Music	7
11	P11	Music	7
12	P12	Music	7
13	P13	Movies	7
14	P14	Movies	7
15	P15	YouTube	7
16	P16	YouTube	7
17	P17	Music	10
18	P18	Music	10
19	P19	Movies	10
20	P20	Movies	10
21	P21	Movies	10
22	P22	YouTube	10
23	P23	YouTube	10
24	P24	Music & Movies	10
25	P25	Music & Movies	12
26	P26	Music	12
27	P27	Music	12
28	P28	YouTube	12
29	P29	Movies & YouTube	12
30	P30	Music	14
31	P31	Music	14
32	P32	Music	14
33	P33	Music	14
34	P34	Music	14
35	P35	Music	14
36	P36	Music	14
37	P37	Music	14
38	P38	Music	14
39	P39	Movies	14
40	P40	YouTube	14
41	P41	Music	21
42	P42	Music	21
43	P43	Music	21
44	P44	Music	28
45	P45	Music	28
None of the participants chose Podcasts as a medium for listening to English			

From the data, music was the most frequently chosen medium, selected by 28 participants, followed by movies (7 participants), YouTube (7 participants), and combinations such as music & movies (2 participants) and movies & YouTube (1 participant). In terms of listening frequency, 13 participants reported listening for 5–7 hours per week (very low interest), 13 participants for 10–12 hours per week (low interest), 11 participants for 14 hours per week (high interest), and 5 participants for 21–28 hours per week (very high interest).

Although the majority of participants reported listening to English regularly, especially through music, the expected correlation between listening frequency and accurate pronunciation of English short vowels /ɛ/ and /æ/ was not strongly established. This finding suggests that exposure alone may not guarantee accurate phonetic acquisition, particularly when the exposure is incidental rather than explicitly focused on pronunciation.

Music, as the most popular medium, offers a rhythmic and repetitive input that may help with memorizing words but not necessarily with accurate vowel perception and production.

As pointed out by Field (Field, 2005), vowel salience in connected speech is often reduced, especially in songs where rhythm and melody distort natural vowel quality. Therefore, participants who mainly rely on music may misperceive or oversimplify vowel distinctions such as /ɛ/ versus /æ/.

Movies and YouTube provide more contextual and visual cues, which could aid semantic understanding but may still not highlight subtle acoustic differences between English vowels. Since the perception of these vowels is influenced by phonetic contrasts not present in the participants' L1 (likely Indonesian), it is plausible that L1 transfer limits their ability to discern and reproduce these vowels accurately, even with frequent exposure.

According to Flege's Speech Learning Model (Flege, 1995), second language learners tend to assimilate unfamiliar L2 sounds into the closest L1 categories. In Indonesian, for instance, the distinction between /ɛ/ and /æ/ does not exist, which may lead learners to perceive both as a single category (often /e/). Thus, even when participants are exposed to English frequently, their phonetic perception remains filtered by L1 categories, leading to inaccurate production despite frequent listening.

The quality and focus of listening activities also play a crucial role. Listening to English through entertainment media such as music and movies often serves recreational purposes rather than form-focused learning. Without explicit attention to pronunciation features—such as vowel length, height, and frontness—learners may not develop phonological awareness necessary for accurate articulation. As Derwing & Munro (Derwing & Munro, 2015) argue, input enhancement and focused training are required to notice and internalize subtle phonetic differences, which casual exposure alone fails to achieve.

Interestingly, no participant chose podcasts, which could have provided rich, natural spoken input. Participants' perception that podcasts are less engaging and difficult to follow may reflect limited listening proficiency and low tolerance for ambiguity—factors that hinder phonetic learning. The lack of visual support in podcasts may also make it harder to focus on vowel quality.

Overall, while the data indicate that participants are frequently exposed to English through various media, especially music, this exposure does not directly translate into accurate vowel pronunciation. Factors such as L1 phonological interference, low perceptual salience of target vowels, and lack of focused listening practice contribute to the weak correlation between listening frequency and pronunciation accuracy. Future interventions should integrate explicit pronunciation training, contrastive listening exercises, and phonetic awareness activities to enhance learners' ability to perceive and produce English vowels accurately.

Participants' pronunciation of the English short vowels /ɛ/ and /æ/

This section presents the findings from the analysis of participants' pronunciation of the English short vowels /ɛ/ and /æ/. Data were collected through voice recordings and analyzed using Praat software to obtain the F1 and F2 formant frequencies. These values were compared with native speaker reference standards, and each participant's weekly listening frequency (in hours) was considered to examine whether increased exposure correlates with greater pronunciation accuracy.

Table 3. The average of English short vowel /ɛ/ and listening frequency

No	Participant's Name	Frequency	Native Speaker F1	Native Speaker F2	Participant F1	Participant F2
1	P1	5	550	1770	755,2	1965,4
2	P2	5	550	1770	776,6	2282,6
3	P3	5	550	1770	705,8	1901,8
4	P4	5	550	1770	814,6	1968,8
5	P5	5	550	1770	776,6	2048,4
6	P6	5	550	1770	709,8	2090
7	P7	5	550	1770	753,6	2062,6

No	Participant's Name	Frequency	Native Speaker F1	Native Speaker F2	Participant F1	Participant F2
8	P8	5	550	1770	747	1404,2
9	P9	5	550	1770	661,4	1799,8
10	P10	5	550	1770	714,6	2075
11	P11	7	550	1770	672	2049,6
12	P12	7	550	1770	738,6	1991,8
13	P13	7	550	1770	713	1863,8
14	P14	7	550	1770	617,4	1507,4
15	P15	7	550	1770	652	2140,6
16	P16	7	550	1770	732,6	2048
17	P17	7	550	1770	600,6	2129,6
18	P18	10	550	1770	649,2	2046,4
19	P19	10	550	1770	661,6	2056,2
20	P20	10	550	1770	679,6	1935,4
21	P21	10	550	1770	586	2128,8
22	P22	10	550	1770	650,4	2171
23	P23	10	550	1770	590,8	2306,6
24	P24	12	550	1770	469,8	1466,6
25	P25	12	550	1770	507	1635,2
26	P26	12	550	1770	665,2	1886,4
27	P27	12	550	1770	426,6	1924,6
28	P28	12	550	1770	481,8	1699,6
29	P29	12	550	1770	559	1831,4
30	P30	14	550	1770	855,4	1939,2
31	P31	14	550	1770	723	1692,4
32	P32	14	550	1770	661,6	1842,2
33	P33	14	550	1770	769	2199,8
34	P34	14	550	1770	401	711
35	P35	14	550	1770	667,2	2223,6
36	P36	14	550	1770	689,2	1977,4
37	P37	14	550	1770	769	1711,2
38	P38	14	550	1770	804,4	2249
39	P39	14	550	1770	577,8	1809,2
40	P40	14	550	1770	561	1802,6
41	P41	21	550	1770	745,4	2209,2
42	P42	21	550	1770	583,4	2032,8
43	P43	21	550	1770	673,6	1908
44	P44	28	550	1770	573,8	1852,6
45	P45	28	550	1770	650,6	1830,8

Table 3 displays the formant frequencies for the vowel /ɛ/, with native speaker reference values of F1 = 550 Hz and F2 = 1770 Hz. The descriptive results indicate that only 3 participants (6.7%) achieved very good pronunciation, producing F1 values close to the standard range (559–573.8 Hz). A further 8 participants (17.8%) were categorized as good, while the majority produced F1 values considerably deviating from the target, categorized as average (16 participants) or poor (18 participants).

A similar distribution is found in the F2 values, with only 3 participants producing near-native realizations (1799.8–1809.2 Hz). Most participants' F2 values were either overly fronted (F2 > 2000 Hz) or backed (F2 < 1600 Hz), reflecting inconsistent articulatory control.

Crucially, the data show no strong correlation between listening frequency and formant accuracy. Several participants with higher listening exposure (≥14 hours per week) produced formant values farther from the native standard, whereas some with lower exposure (5–7 hours per week) achieved more accurate pronunciation.

This finding suggests that quantity of exposure alone is insufficient for accurate vowel acquisition. According to Flege's Speech Learning Model (Flege, 1995) and Best and Tyler's Perceptual Assimilation Model (Best & Tyler, 2007), L1 phonological categories strongly influence L2 perception and production. Indonesian speakers, for example, may perceive /ɛ/ as similar to their native /e/, which typically has a lower F1 and higher F2 than the English /ɛ/. Consequently, learners often produce a more fronted vowel, consistent with the elevated F2 values observed.

Furthermore, Escudero (Escudero, 2006) emphasizes the role of perceptual salience—learners must be able to distinguish subtle acoustic differences between L2 vowels. Without explicit phonetic awareness, learners may continue to map English /ɛ/ onto their L1 /e/ category, leading to persistent inaccuracies despite frequent listening.

Table 4. The average of English short vowel /æ/ and listening frequency

No	Participant's Name	Frequency	Native Speaker F1	Native Speaker F2	Participant F1	Participant F2
1	P1	5	690	1660	651	2127,8
2	P2	5	690	1660	776,6	2282,6
3	P3	5	690	1660	750,2	2143,2
4	P4	5	690	1660	901,6	1754,6
5	P5	5	690	1660	810,2	2055
6	P6	5	690	1660	714,2	2238,4
7	P7	5	690	1660	624,4	1449,4
8	P8	5	690	1660	875,4	1644
9	P9	5	690	1660	726,6	1879,6
10	P10	5	690	1660	644,4	1705,4
11	P11	7	690	1660	702,2	2337
12	P12	7	690	1660	728,2	1915,2
13	P13	7	690	1660	896,4	2040,8
14	P14	7	690	1660	722,2	1792,8
15	P15	7	690	1660	699,8	2114,8
16	P16	7	690	1660	822,2	1996,8
17	P17	7	690	1660	580,4	1960,2
18	P18	10	690	1660	766,4	1897,4
19	P19	10	690	1660	633,8	1450
20	P20	10	690	1660	896	1808,8
21	P21	10	690	1660	690,8	2070
22	P22	10	690	1660	737,4	1735,8
23	P23	10	690	1660	692,6	2102,8
24	P24	12	690	1660	456,8	2366,8
25	P25	12	690	1660	553,6	1502,4
26	P26	12	690	1660	742,2	2142,4
27	P27	12	690	1660	581,4	1764,2
28	P28	12	690	1660	546,6	1836,8
29	P29	12	690	1660	621	1697
30	P30	14	690	1660	943,2	1892,4
31	P31	14	690	1660	659,8	1399,2
32	P32	14	690	1660	754,6	1850,6
33	P33	14	690	1660	848,2	2007,2
34	P34	14	690	1660	375,4	812,4
35	P35	14	690	1660	880	1989,8
36	P36	14	690	1660	682,8	2141,2
37	P37	14	690	1660	780,6	1841,8
38	P38	14	690	1660	849,6	1818,2
39	P39	14	690	1660	708,2	1748

No	Participant's Name	Frequency	Native Speaker F1	Native Speaker F2	Participant F1	Participant F2
40	P40	14	690	1660	457,6	2069,6
41	P41	21	690	1660	724,2	2230,6
42	P42	21	690	1660	764,6	1870,2
43	P43	21	690	1660	643,6	1588
44	P44	28	690	1660	595,8	1783,6
45	P45	28	690	1660	715,6	1803,8

Table 4 shows participants' production of /æ/, with native reference values of F1 = 690 Hz and F2 = 1660 Hz. The descriptive results indicate that 12 participants (26.7%) achieved very good pronunciation, while 16 (35.6%) performed well. However, 11 participants (24.4%) demonstrated poor accuracy, producing F1 values much higher or lower than the standard (375.4–457.6 Hz and 848.2–943.2 Hz). Similarly, the F2 data reveal that only 3 participants approximated the native standard (1644–1705.4 Hz), while many others produced either excessively fronted (F2 > 2000 Hz) or backed (F2 < 1500 Hz) variants. Again, listening frequency did not strongly predict pronunciation accuracy. Some participants with higher exposure produced vowels that deviated significantly from the target, suggesting that exposure without perceptual adaptation fails to improve production.

These results align with findings by Bohn and Flege (Bohn & Flege, 1997), who assert that input quantity must be accompanied by perceptual quality. Learners must first notice and categorize L2 vowel contrasts distinct from their L1 before accurate articulation can occur. In the case of /æ/, which has no direct equivalent in Indonesian, learners may substitute it with /a/ or /ɛ/, leading to misplacement along the F1–F2 acoustic space. The pervasive deviations indicate negative L1 transfer and insufficient perceptual differentiation. As Iverson and Evans (Iverson & Evans, 2007) argue, successful L2 vowel learning requires the formation of new phonetic categories, which is often hindered by L1 dominance and lack of explicit instruction.

The findings reveal that while some participants demonstrate near-native vowel production, most struggle to accurately produce /ɛ/ and /æ/ due to L1 interference, perceptual assimilation, and limited phonetic awareness. The absence of a strong relationship between listening exposure and pronunciation accuracy underscores the limitation of implicit learning in phonetic acquisition.

From a theoretical perspective, these findings support the view that L2 phonetic learning is constrained by perceptual and articulatory factors rather than exposure alone. Learners' perceptual systems often filter L2 sounds through L1 categories, preventing them from noticing subtle acoustic cues critical for accurate production (Best & Tyler, 2007). Therefore, pedagogical approaches should integrate explicit phonetic training, including formant visualization tools (e.g., Praat), minimal pair discrimination, and targeted articulatory feedback. As Derwing and Munro (Derwing & Munro, 2015) highlight, explicit instruction can enhance learners' metaphonological awareness, enabling them to adjust their vowel production more effectively.

Correlation between the frequency of participants listening to English and their pronunciation of English short vowels /ɛ/ and /æ/

The analysis using SPSS revealed that there was no significant correlation between participants' frequency of listening to English and their pronunciation accuracy of the English short vowels /ɛ/ and /æ/, as measured by the acoustic parameters F1 and F2. For the vowel /ɛ/, the correlation between listening frequency and F1 yielded a significance value of .052 with a Pearson coefficient of .292, while the correlation with F2 produced a significance value of .085 and a coefficient of .260. For the vowel /æ/, the correlation between listening frequency and F1 resulted in a significance value of .942 with a Pearson coefficient of –.011, and the correlation with F2 showed a significance value of .642 with a coefficient of .071. Since all significance values exceeded the .05 threshold, these results indicate that there is no statistically significant relationship between the frequency of listening to English and the

accuracy of vowel production.

These findings suggest that increased listening frequency alone does not lead to improved accuracy in pronouncing the English short vowels /ɛ/ and /æ/. This result contradicts the initial assumption that frequent exposure to English input would enhance learners' ability to imitate and produce target sounds more accurately. Several factors may account for this unexpected outcome.

First, from the perspective of perceptual salience and phonetic similarity, the vowels /ɛ/ and /æ/ pose particular challenges for Indonesian learners due to the absence of an equivalent contrast in the Indonesian vowel system. According to Flege's Speech Learning Model (SLM) (Flege, 1995), when L2 sounds are perceived as similar to an existing L1 category, learners may fail to establish distinct phonetic representations. In the case of /ɛ/ and /æ/, both may be assimilated to the Indonesian /e/, resulting in reduced perceptual discrimination and inaccurate production despite extensive listening exposure. Escudero (Escudero, 2005) also notes that perceptual assimilation can limit learners' ability to form new categories unless they receive focused perceptual training.

Second, the measure of "listening frequency" used in this study likely reflects the quantity rather than the quality of input. Passive listening—such as through songs, movies, or casual conversations—may not facilitate phonetic learning unless learners engage in focused, analytic listening that directs attention to segmental features. Research by Derwing and Munro (Derwing & Munro, 2015) emphasizes that explicit instruction and corrective feedback play a more crucial role in improving pronunciation than mere exposure. Without targeted attention and feedback, learners may continue to perceive and produce sounds inaccurately.

Third, L1 transfer exerts a strong influence on learners' pronunciation patterns. Indonesian vowels are generally centralized and less dispersed in the acoustic space than English vowels, which can lead to systematic deviations in F1 and F2 values when learners attempt to produce front or low vowels like /ɛ/ and /æ/. This articulatory habit persists even with frequent exposure to native speech, as learners may not consciously modify their tongue height and position to match native-like targets.

Finally, pronunciation development is shaped by individual and instructional factors such as phonetic aptitude, motivation, and the presence or absence of pronunciation-focused teaching. Learners who merely consume English media without receiving explicit pronunciation training or engaging in active production practice may not achieve significant gains in segmental accuracy.

Taken together, these findings indicate that listening frequency alone is an insufficient predictor of vowel pronunciation accuracy. While frequent listening may improve general comprehension and prosodic awareness, accurate production of challenging vowels like /ɛ/ and /æ/ requires explicit phonetic instruction, perceptual training, and corrective feedback. This aligns with findings from previous studies showing that focused, form-oriented practice yields greater improvement in L2 pronunciation (Thomson & Derwing, 2014).

Future research should therefore incorporate qualitative measures of listening engagement—such as the types of input, level of attention, and presence of feedback—as well as experimental designs that include targeted training sessions. Such approaches would provide deeper insight into how different aspects of input exposure contribute to the acquisition of L2 vowel contrasts.

4. CONCLUSION AND SUGGESTIONS

The findings of this study reveal that there is no statistically significant correlation between the frequency of participants' listening to English and their accuracy in pronouncing the English short vowels /ɛ/ and /æ/. Although it was initially assumed that greater listening exposure would enhance pronunciation accuracy, the data show otherwise. Participants with higher listening frequencies did not necessarily produce vowels closer to the native standard, as measured by F1 and F2 formant frequencies.

The results highlight that listening quantity alone is insufficient to improve segmental pronunciation accuracy. Instead, learners' performance is strongly shaped by L1 transfer,

perceptual assimilation, and limited phonetic awareness. Indonesian learners tend to assimilate English vowels /ɛ/ and /æ/ to their native /e/ or /a/, which prevents the formation of new phonetic categories and leads to persistent deviations in vowel production.

From a theoretical perspective, these outcomes support models such as Flege's Speech Learning Model (1995) and Best & Tyler's Perceptual Assimilation Model (2007), which emphasize the role of perceptual and articulatory constraints in L2 sound acquisition. Without explicit phonetic training, learners are unlikely to achieve native-like pronunciation merely through passive exposure.

Suggestions: (1) For Teaching Practice: Incorporate explicit pronunciation instruction into English learning programs, particularly for segmental features like vowels, Utilize formant visualization tools (e.g., Praat) and minimal pair discrimination exercises to raise learners' phonetic awareness, and Provide corrective feedback and guided articulatory practice to help learners consciously adjust tongue height and position for accurate vowel production; (2) For Learners: Engage in active and analytic listening rather than passive exposure, focusing on identifying and mimicking subtle vowel differences, and Combine listening with production practice and self-monitoring using phonetic tools or teacher feedback; (3) For Future Research: Include qualitative measures of listening engagement, such as the type of input (e.g., songs vs. pronunciation drills), level of attention, and feedback received, Conduct experimental studies that compare the effects of passive exposure and explicit phonetic training on vowel acquisition, and Explore individual learner differences (e.g., aptitude, motivation, previous training) to understand their roles in L2 pronunciation development.

REFERENCES

Apridayanti, R. (2021). *The Correlation Between Students' Ability in Listening to English Songs and Their Pronunciation Mastery*. https://repository.radenintan.ac.id/16100/

Baker, J., & Westrup, H. (2003). *Essential speaking skills: A handbook for English language teachers*. Continuum. https://books.google.co.id/books?id=hnE8CwAAQBAJ&printsec=frontcover

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 13–34. John Benjamins. https://doi.org/10.1075/lllt.18.03bes

Bohn, O.-S., & Flege, J. E. (1997). Perception and Production of a New Vowel Category by Adult Second Language Learners. In A. James & J. Leather (Eds.), *Second-Language Speech: Structure and Process*, 53–74. https://doi.org/10.1515/9783110882933.53

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://www.researchgate.net/publication/280198453_Pronunciation_Fundamentals_Evidence-based_Perspectives_for_L2_Teaching_and_Research

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [University of Amsterdam]. https://hdl.handle.net/11245/1.147612

Escudero, P. (2006). The Phonological and Phonetic Development of New Vowel Contrasts in Spanish Learners of English. In B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish–English Interphonology*, 41–55. https://www.researchgate.net/publication/254901634_The_phonological_and_phonetic_development_of_new_vowel_contrasts_in_Spanish_learners_of_English

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423. https://doi.org/10.2307/3588495

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language*

research, 233–277. https://www.researchgate.net/publication/333815781_Second_language_speech_learning_Theory_findings_and_problems

Hidayatullah, M. S. (2018). Improving students' pronunciation through Western movie media. *Al-Lisan: Jurnal Bahasa*, *4*(1), 93–111. https://doi.org/10.30603/al.v4i1.381

Idayani, A. (2019). An analysis of students' errors in pronouncing English words. *English Journal of Indonesia*. https://doi.org/10.32520/eji.v3i1.469

Iverson, P., & Evans, B. G. (2007). Learning English Vowels with Different First-Language Vowel Systems: Perception of Formant Targets, Formant Movement, and Duration. *Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://www.researchgate.net/publication/5661051_Learning_English_vowels_with_different_first-language_vowel_systems_Perception_of_formant_targets_formant_movement_and_duration

Manurung, L., Ismiyati, Y., & Yurni. (2024). Analysis of Student's Pronunciation Ability of Twelfth Grade Student at SMAN 8 Jambi. *Journal of English Language Teaching*. https://jelt.unbari.ac.id/index.php/jelt/article/download/173/170

Rao, P. S. (2019). The importance of speaking skills in English classrooms. *Alford Council of International English \& Literature Journal (ACIELJ)*, *2*(2), 6–18. https://www.acielj.com/Papers/vol2issue2/1.ACIELJ -Srinu sir (1-12) OK.pdf

Resha, P. W., Mawardin, M. S., & Abdul, W. (2015). Developing Speaking Skill of Grade VIII Students Through Short Conversation. *Unpublished Manuscript*. https://media.neliti.com/media/publications/244500-none-65dd45e8.pdf

Supeno, S. (2018). The effect of listening to English songs toward students' pronunciation. *Journal of English Education Studies*, *1*(1), 1–10. https://doi.org/10.31316/jees.v1i1.123

Syahabuddin, K., & Rizqa, K. (2021). Improving students' listening skill using podcasts. *Journal of Digital Education, Communication, and Arts (DECA)*, *4*(1), 51–61. https://doi.org/10.30871/deca.v4i01.2867

Tahmina, T. (2023). Students' perception of the use of YouTube in English language learning. *Journal of Languages and Language Teaching*, *11*(1), 151. https://doi.org/10.33394/jollt.v11i1.6883

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A research synthesis and meta-analysis. *Applied Linguistics*, *35*(3), 1–26. https://doi.org/10.1093/applin/amt018

Upa, R., Ola, T., & Wahyono, E. (2021). Correlation Between Students' Learning Habit in Listening English Song and Their Pronunciation Capacity. *Ethical Lingua*, *8*(1), 1–7. https://ethicallingua.org/25409190/article/view/275

Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge University Press. https://sacunslc.wordpress.com/wp-content/uploads/2015/03/penny-ur-a-course-in-language-teaching-practice-of-theory-cambridge-teacher-training-and-development-1996.pdf