JAPANEDU: Jurnal Pendidikan dan Pengajaran Bahasa Jepang
Vol. 10, No. 01, June 2025, pp. 18-29

JAPANEDU:
Jurnal Pendidikan dan Pengajaran Bahasa Jepang
http://ejournal.upi.edu/index.php/japanedu/index

Japanese Homophone Frequency Analysis
An Acoustic Study

Tanty Aidullia Nasution, Peni Bernita Bangun, Siti Muharami Malayu
Universitas Sumatera Utara, Sumatera, Indonesia
tantynasution2000@gmail.com

ABSTRACT

Japanese homophones, words that share the same phonetic form but differ in meaning, are a prominent feature of the language, largely due to its limited phonemic inventory and rich prosodic system. Distinctions between such homophones are often conveyed through pitch accent and syllable duration, which serve as crucial phonological cues in spoken discourse. This study investigated two homophonic word pairs: *ame* ('rain' vs. 'candy') and *hashi* ('bridge' vs. 'chopsticks'). Recordings were collected from native Japanese speakers, who were asked to produce the target words in a range of natural sentence contexts. Acoustic analyses were conducted using Praat software to examine pitch contours and syllable durations associated with each meaning. Results revealed consistent prosodic patterns that differentiate meanings within each homophone pair. In the case of *ame*, the pronunciation of 'rain' featured a longer vowel duration on the first syllable [a] and a generally higher pitch contour, while 'candy' exhibited greater length on the second syllable [me]. Similarly, for *hashi*, 'chopsticks' was characterized by a longer [ha] syllable and elevated pitch at both the beginning and end of the word, whereas 'bridge' showed a longer [shi] and a higher overall maximum pitch. These findings suggest that prosodic elements—particularly pitch accent and vowel duration—play an essential role in disambiguating homophones in Japanese. The results further highlight the distinct phonological structure of the Japanese language, where subtle variations in prosody contribute significantly to lexical differentiation in spoken communication.

KEYWORDS

Acoustic Phonetics; Homonyms; Japanese; Praat analysis; Prosodic.

ARTICLE INFO

First received: 07 May 2025 *Revised: 24 June 2025* *Final proof accepted: 27 June 2025*
Available online: 30 June 2025

INTRODUCTION

Language evolves in tandem with the society in which it is used. Situations, culture, and ideology are all part of the social context in which language develops and emerges (Halliday 2002; Martin, 1992: 496; Eggins, 2004). Values, attitudes, thought patterns, ideas, culture, and ideology are all represented in language as a result of linguistic evolution and social transformation (Widodo & Purnama, 2020; Wati & Hidayat, 2022). Language demonstrates that language is the community's identity and reflects the features of society.

Phonetics and phonology are the two branches of linguistics that concentrate on studying language sounds, according to Okumura in Tjandra (2004). The study of a given language or languages' sound system is known as phonology in linguistics. Phonology is one of the fundamental areas of linguistics, claims Odden (2005). According to Odden (2005), phonology is made up of different linguistic rules. Based on the frequency and intensity of the sound produced, each sound source can be identified by its qualities (Peterson & Henderson, 2020; White & Taylor, 2021; Davidson & Lee, 2023). Conversely, acoustic phonetics, as defined by Marsono (2013), is the study of language sounds from the standpoint of a physical phenomenon, encompassing elements like vibration frequency, amplitude, intensity, and timbre. The three primary features of acoustic phonetics are intensity, duration, and frequency. An English acoustician named Fry (1955, 1958) measured English speakers' speech sounds using acoustic characteristics like length, intensity, and fundamental frequency. To ascertain lexical level, primary and secondary stress, and phonetic analysis in eight Chickasaw speakers, Gordon (2004), another acoustic researcher, also investigated the use of stress in speech using acoustic metrics such as duration, fundamental frequency, intensity, and voice quality.

Numerous languages, including English, Dutch, French, German, Indonesian, and languages of the Archipelago, have been the subject of phonetic and phonological research (Ningsih, 2007). Every language has a unique set of phonemes, phonetic structures, phonological and phonetic norms, and many even have different orthographic systems, according to the research (Hayes-Harb & Barrios, 2021; Keating, 1984; Mennen, Scobbie, De Leeuw, Schaeffler, & Schaeffler, 2010). There are three primary subfields within phonetic research: 1) Acoustic phonetics, which studies the sound waves generated during articulation and their propagation through the air; 2) Auditory phonetics, which studies how the auditory system perceives language sound waves and the auditory organs, such as the ears and other body parts involved in the hearing system; and 3) Articulatory phonetics, which studies the system of organs that produce language sounds and the processes involved in sound production (Lidwina, 2012; Mücke, Hermes, & Tilsen, 2020; Roettger, 2019; Vigário, Frota, & João, 2011).

The phenomenon of homonyms in the Japanese language is among the fascinating topics examined in this study. One of the languages having a sophisticated and distinctive phonological system is Japanese (Ito & Mester, 2013). Pitch accent, a pattern of frequency variations in pitch that is crucial in differentiating the meanings of specific words, is one of its primary characteristics. Pitch accent in Japanese often happens at the word level, in contrast to tonal languages like Mandarin, where each syllable has a fixed tone (Goss & Tamaoka, 2019; Idemaru, Wei, & Gubbins, 2019). The primary distinction between phonetically similar words is frequently this accent pattern, which also gives the word a prosodic identity.

One of the outcomes of the Japanese phonological system, which has a restricted number of phonemes but a rich prosodic component, is the phenomenon of homonyms, which are words that have the same pronunciation but distinct meanings. For instance, depending on where the pitch accent is placed, the word *hashi* can mean 'bridge' (橋) or 'chopsticks' (箸), whilst *ame* can mean 'rain' (雨) or 'candy' (飴). In situations like this, the pitch frequency pattern or pitch contour that appears on a particular syllable is what makes a difference in meaning.

In addition to being crucial for comprehending the phonological structure of the Japanese language, the study of homonym frequency differences has wide-ranging effects on pragmatics and language processing. Context plays a major role in helping native speakers differentiate between the meanings of homonymous terms in casual discussions. However, these distinctions are frequently hard for non-native speakers to understand when it comes to cross-cultural communication or language acquisition. Therefore, research on acoustic aspects such as frequency becomes important to understand how meaning can be encoded prosodically in language.

The purpose of this study is to use Praat software to examine the pitch frequency patterns of Japanese homonyms. Software called Praat was created especially for phonetic analysis. Praat, which was created by Paul Boersma and David Weenink of the University of Amsterdam's Institute of Phonetic Sciences, has several sophisticated capabilities for analyzing various characteristics of sound, including frequency, intensity, and duration. One computer program that can show the frequency of sound waves is called Praat. When a speech recording is entered into this program, Karsono (2013) claims that the sound wave can be used to assess the voice's features. This Praat program will analyze and view Japanese homonyms according to their pitch or frequency.

METHODS

This research methodology experimentally combines qualitative and quantitative data. The purpose of this study is to use Praat software to analyze the pitch contour of Japanese homonyms (De Jong & Wempe, 2009). Analyzing acoustic data

e- ISSN 2528-5548 | p-ISSN 27764478

using an instrumental method enables quantitative frequency measurement. In this instance, statistical analysis of the acoustic measurement data is possible.

Native Japanese speakers' voice recordings were used to gather research data. To record intonation changes that impact pitch frequency patterns, informants were instructed to speak homonyms like *hashi* ('bridge' and 'chopsticks') and *ame* ('rain' and 'candy') in a variety of sentence situations. To guarantee convenience and adaptability in data collecting, this audio recording was made using common recording equipment, like a cell phone.

The next step is to import the recording data into the Praat program so that it may be analyzed. The Praat program illustrates the frequency of sound waves, enabling in-depth acoustic study. When a speech recording is entered into this program, the sound wave of the voice can be used to assess its characteristics (Karsono, 2013). During the analysis phase, variations in the frequency patterns of each homonym syllable are found using the Pitch Contour feature. The frequency differences that can differentiate the meanings of homophones are then analyzed quantitatively using this data.

RESULT AND DISCUSSION

The Praat application will then be used to examine the data collected for this study for frequency. These are the data analysis's findings.

Ame can Signify Either 'Candy' (飴) or 'Rain' (雨)

Depending on the pitch, the Japanese word "*ame*" can signify two different things. The measuring findings for each word are as follows.

Ame *(雨) Means 'Rain'*

The word "*ame*," which signifies 'rain', is the subject of the first analysis. This analysis used the Praat application that the results of which are shown in Figures 1, 2, 3, and 4.

Figure 1 is the result of an analysis of the acoustic properties of the Japanese word *ame* 'rain' as spoken by a native speaker. This figure shows the duration syllable [a] in *ame*, which has the meaning 'rain'. The pronunciation of the syllable [a] takes 0.167574 seconds in Figure 1. The top component, Sound Waves (top), displays a graphic depiction of a sound wave. The pattern of air vibrations that contribute to the sound is visible in the pink region [a]. The loudness of the sound is indicated by the wave's amplitude.

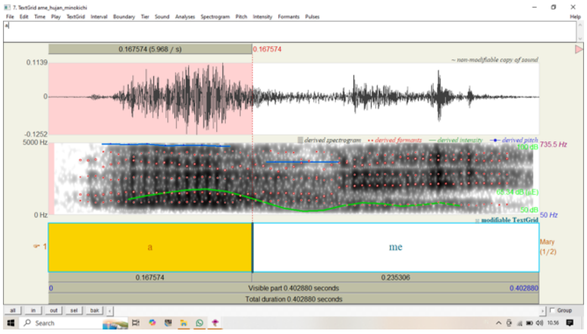

Figure 1: The length of time needed to pronounce the syllable [a] in the word "*ame*".

Spectrogram (center): The spectrogram is situated at the bottom of the sound waveform. This spectrogram shows the sound frequency as a function of time. Time is displayed on the horizontal axis in seconds, and frequency is displayed on the vertical axis in Hertz. The strength of energy at a specific frequency is indicated by the spectrogram's color's blackness. The darker horizontal stripes for the syllable [a] are known as formants. Vowel identification relies on formants, which are the vocal tract's resonance frequencies. To differentiate between various vowel kinds, the first formant (F1) and the second formant (F2) are essential. The quality of the vowel [a] in the word's context is reflected in the location and separation between F1 and F2 in this spectrogram of the syllable [a]. The blue line, which indicates the voice's pitch or high and low tones, is seen in addition to the formants. The loudness or intensity of the voice is indicated by the green line. We can see how the pitch and strength of this syllable [a] vary throughout its duration. The TextGrid, which enables us to label sound segments in the audio, sits at the absolute bottom of this analytic process. The sound [a] is labeled in time intervals ranging from 0 to around 0.167 seconds, as may be seen here. The phonetic identification of that sound segment is indicated by the label "a". The label "me" follows the subsequent syllable.

In Figure 2, the pronunciation of the syllable [me] takes 0.235306 seconds. And now for the distribution of frequency energy in the spectrogram below. Observe the current shift in the formant patterns. The position and motion of formants that are typical of the change from the vowel [a] to the consonant [m], and then to the vowel [e]. Energy at

e- ISSN 2528-5548 | p-ISSN 27764478

lower frequencies can be used to identify the nasal consonant [m]. The vowel [e] will thereafter be represented by a shift in formants. To understand the different articulations of the two vowels, compare the formant positions in the [me] and [a] syllables. Pitch fluctuates throughout the [me] syllable, as indicated by the blue line. This analysis determines whether the fundamental frequency of this syllable differs from that of the preceding one. The shift in loudness inside the [me] syllable will also be indicated by the green line that indicates intensity. The [me] syllable occupies the time interval from 0.167574 seconds to roughly 0.235306 seconds, as can be seen in the TextGrid at the bottom. This sound segment can be easily identified by the label "me". The phonetic properties of the [me] syllable in the pronunciation of the word (雨) are fully demonstrated by this acoustic study. We can identify [a] and [me] from one another scientifically because of the variations in their acoustic characteristics.

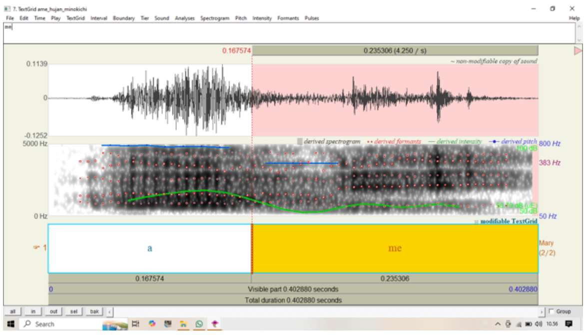

Figure 2: The length of time needed to pronounce the syllable [me] in the word "*ame*".

In Figure 3, the word's sound wave (雨) is still visible in the upper portion. But this time, blue vertical lines are visible. The glottal pulses or vocal cord vibrations that underpin sound generation are indicated by these lines. The basic frequency or pitch of the sound is directly correlated with the density of these pulses. The pitch contour is particularly visible in the image's center. The voice's pitch shift over time is depicted by the blue line with control points. The recording starts at around 0.101440 seconds, which is the first point indicated in red. The pitch value, which is approximately 119.9 Hz, is now visible on the left side. This shows how often the vocal cords vibrate at the start of the word (雨). The fundamental aspect of the speaker's pitch at that moment is this frequency.

This first red dot, which is the lowest frequency point in pronunciation, is a crucial indicator for determining a spoken word's first pitch. As you can see, the voice's fundamental frequency at the start of pronunciation is approximately 119.9 Hz. The blue line's subsequent movement indicates the pitch shift that occurs after this moment. In the study of a language's prosody, which encompasses intonation, stress, and tone patterns that convey linguistic and emotional meaning, pitch analysis such as this is crucial. We can comprehend how pitch is utilized to convey information in Japanese by comparing the pitches of various word or phrase components.

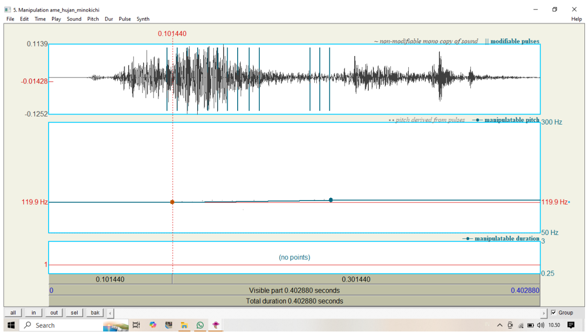

Figure 3: First and Lowest Word Frequencies for "*ame*".

Furthermore, in Figure 4, according to the measurement data, there is no discernible variation in the frequency of this word's pronunciation of this syllable. The final and greatest frequency of this word's pronunciation is 124.6 Hz. This point is at the time coordinates of approximately 0.231440 seconds. This depicts the speaker's voice pitch at that precise instant. The utterance of this word exhibits a tiny increase in pitch over time, if we compare it to the first red dot in the preceding figure (about 119.9 Hz). The pitch of the voice fluctuates continually throughout the word (雨), as indicated by the blue lines that connect the pitch control points. A sample measurement of pitch at a specific moment is represented by the red dot. Despite being minor in this instance, these pitch shifts can transmit significant linguistic information, particularly about intonation and syllable stress. Although Japanese is not a tonal language like some other Asian languages, words can have distinct meanings depending on their pitch, a phenomenon called pitch accent. These pitch shifts are also correlated with the glottal pulses shown at the top of the picture. As the fundamental frequency (pitch) rises, so does the pulse density.

e- ISSN 2528-5548 | p-ISSN 27764478

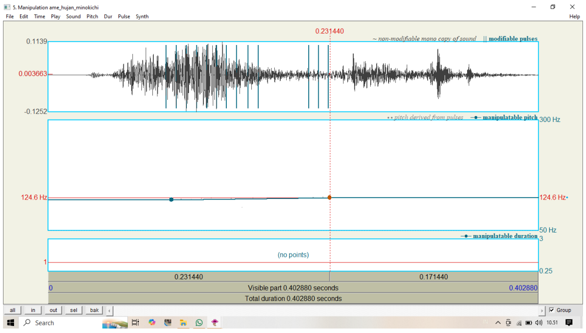

Figure 4: Final and highest frequency of the word "*ame*".

Ame *(飴) Means 'Candy'*

The measuring result for the candy-related term *ame* (飴). Additionally, this word has two syllables: [a] and [me]. Praat analysis for *Ame* (飴) means 'candy', shown in Figures 5, 6, 7, and 8.

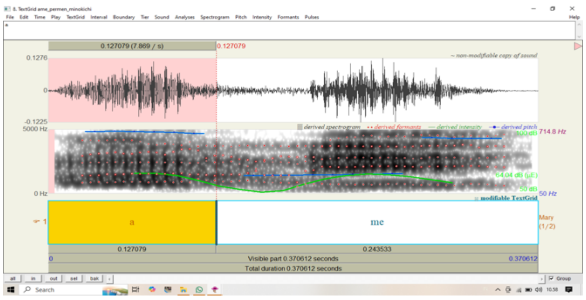

Figure 5: Duration of Syllable Pronunciation [a].

According to the measured data, the pronunciation of the syllable [a] in the word "*ame*," which means 'candy', takes 0.127079 seconds in Figure 5. A sound wave may be seen at the very top. The sound [a] is represented by the pink section. In comparison to the segment right after it, the amplitude of the wave in this section shows how loud the sound [a] is. The creation of the vowel sound is physically represented by this wave vibration pattern [a]. The strength of energy at a specific frequency is indicated by the spectrogram's color's blackness. We may see horizontal bands known as formants for the syllable [a]. In the context of pronunciation (飴), the location and separation between F1 and F2 in this spectrogram are typical for the vowel [a]. Take note of the formants' comparatively low and widely spaced positions, which are characteristic of open vowels like [a]. TextGrid, the bottom part, gives several time segments phonetic labels. In this case, the label "a" indicates that the duration starts at the beginning and ends at roughly 0.127079 seconds. The label "me" follows the subsequent syllable.

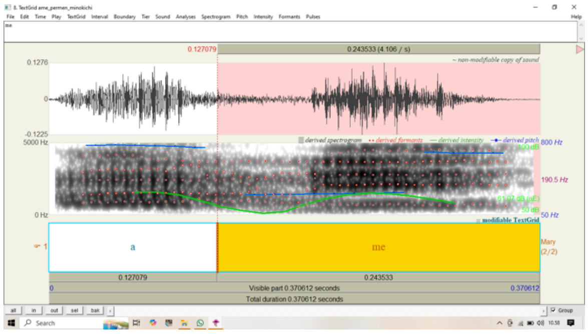

Figure 6: Duration of Syllable Pronunciation [me].

In Figure 6, the pronunciation of the syllable [me] takes 0.243533 seconds. Compared to the 'rain' variant, [a] has a shorter duration, while [me] appears to occupy a larger portion of the total utterance, suggesting emphasis on the second syllable. This syllable analysis shows that there is greater space before and after the syllable [me] as it moves from the syllable [a] to the syllable [me]. Below the waveform is the spectrogram, showing frequency on the vertical axis and time on the horizontal axis. The darkness of the spectrogram reflects the intensity of formants (resonant frequencies). The green pitch contour line indicates the variation in fundamental frequency (F0), which begins high (around 714.8 Hz) and gradually descends, a typical pitch pattern for words like *ame* 'candy' with an unaccented prosodic pattern. The blue line indicates overall intensity, which peaks around the vowel region. The yellow and blue boxes at the bottom segment the word into its syllables: [a] and [me], labeled in a TextGrid. This acoustic pattern supports findings that "candy" (*ame*) has a higher duration in the [me] syllable and displays a falling pitch contour, helping distinguish it from its homophone "rain".

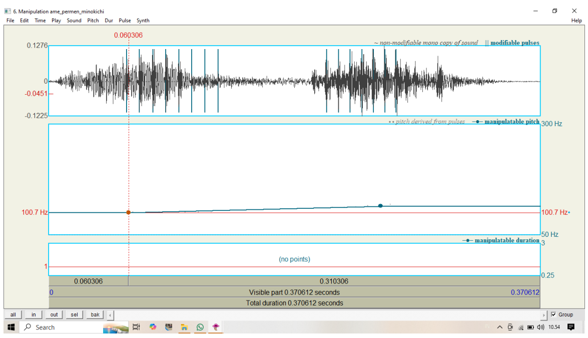

Figure 7: Lowest and Initial Frequencies.

e- ISSN 2528-5548 | p-ISSN 27764478

In Figure 7, it showed the initial frequency is shown at 100.7 Hz, which is also the lowest frequency. At a time of roughly 0.060306 seconds, the red dot marks the start of the pitch contour. A frequency value of roughly 100.7 Hz is shown on the left side of the graph, parallel to the red dot. A sound waveform, which visually depicts the variations in air pressure when a word is pronounced, is displayed in the upper section. The glottal pulses, which are caused by the vocal folds vibrating, are shown by the blue vertical lines. The fundamental frequency or pitch of the voice is intimately correlated with the density of these pulses. The fundamental frequency of the voice at the start of the word (飴) pronunciation is 100.7 Hz. This is an indication of the speaker's voice pitch at that moment. The speed at which the vocal cords vibrate produces this frequency. The beginning of the voice pitch of this word is indicated by this red dot. The blue line that follows the red dot illustrates how the voice's pitch varies over time, and it indicates that the pitch slightly rises beyond this initial point. The fact that this term has an increase from the beginning to the end is explained indirectly.

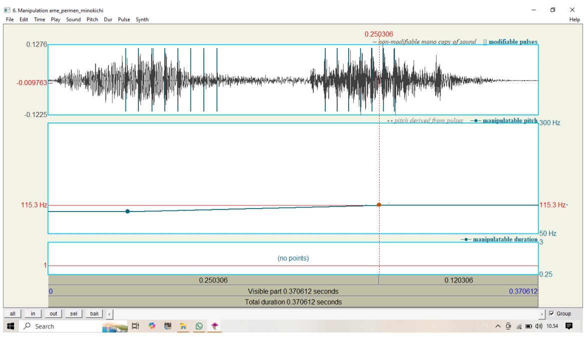

Figure 8: Highest and Final Frequency.

Figure 8 shows that frequency tests revealed that the end frequency, which is also the highest frequency, is at 115.3 Hz. The sound shown here is the Japanese word *ame,* meaning 'candy' (飴), with a total duration of approximately 0.370612 seconds. At the top, the waveform displays the amplitude of the speech signal. We can observe that there is a rise in pitch during the word's pronunciation, followed by a tiny drop back to 115.3 Hz at the finish, if we compare it to the red dot in the previous image, which is about 100.7 Hz at the beginning of the word. The smooth, gently rising pitch is typical of unaccented words in Japanese like *ame* 'candy', helping distinguish it from its homophone *ame* 'rain', which has a different pitch pattern. The bottom panel is the duration tier, currently empty, indicating that no time-stretching or compression has been applied. If filled, it would allow researchers to adjust the timing of specific segments. In general, the red frequency points in this image give precise information about the word's final pitch (飴), which is around 115.3 Hz. Together with the pitch variations throughout time, this is the basic frequency of the speaker's vocal cords vibrating at the end of the word, which creates the word's intonation pattern.

Hashi can Signify Either 'chopsticks' (箸) or 'bridge' (橋)

Depending on the accent, the term "*hashi*" can also indicate two different things. *Hashi* can be defined as a 'bridge' (橋) and 'chopsticks' (箸). The frequency difference between these two terms can then be calculated.

Hashi (橋) Which Means Bridge

There are two syllables in the word "*hashi*," which means bridge: [ha] and [shi]. The length of time it takes to pronounce each syllable will then be used to gauge that. Analysis of this word is shown in Figures 9, 10, 11, and 12.

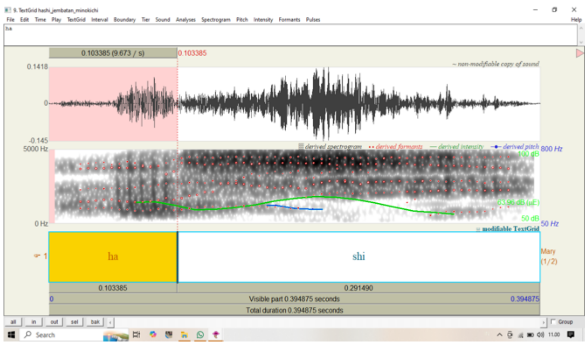

Figure 9: Pronunciation duration of the syllable [ha].

According to the measurement results, the word *hashi* (橋), which means bridge, may be said with a duration of 0.103385 s for the syllable [ha] in figure 9. The red vertical line at 0.103385 seconds marks the boundary between the first syllable [ha] and the second syllable [shi]. The [ha] syllable occupies about 0.103 seconds, while [shi] spans a longer duration, suggesting a prosodic emphasis on the second syllable. The loudness of the sound [ha] in this portion is indicated by the wave's amplitude in comparison to the subsequent segment. The physical manifestation of the sound's generation is

e- ISSN 2528-5548 | p-ISSN 27764478

this wave vibration pattern [ha]. The presence of a fricative consonant component [h] is indicated by the aperiodic period at the beginning, as we can see. The spectrum is comparatively flat and diffuse at high frequencies at the beginning of the syllable [ha], which is typical of the voiceless consonant [h]. The vowel [a] then follows the usual formant pattern. Darker horizontal bars will be visible for the first formant (F1) and the second formant (F2). The quality of the vowel [a] following the consonant [h] is indicated by the position and separation between F1 and F2.

The next one is the syllable that has 0.291490 s for the syllable [shi] in Figure 10. The total duration of the utterance is approximately 0.394875 seconds. A spectrogram is shown below. We shall observe the spectral properties of the vowel [i] after the consonant [ʃ] (as in the English word "she") for the syllable [shi]. A voiceless post-alveolar glide, the consonant [ʃ] is distinguished by a comparatively high, dispersed spectral energy in the mid-to-high frequencies. The flat high-frequency spectrum we observe for the consonant [h] in the syllable [ha] will differ from this pattern. We shall observe a shift to the formant pattern of the vowel [i] following the consonant spectrum [ʃ]. As is typical of the vowel [i], the first formant (F1) will be very low, while the second formant (F2) will be quite high and far from F1. The waveform at the top illustrates the amplitude of the sound over time. Below the waveform is the spectrogram, which shows the frequency content of the sound. The darker regions represent areas of higher acoustic energy. Overlaid are red dots indicating formant frequencies, green line showing pitch contour (fundamental frequency), and blue line representing intensity. In this token of *hashi* ('bridge'), the pitch rises gradually through [ha] and peaks around the beginning of [shi], then falls slightly, consistent with the unaccented pattern typical for this meaning. The intensity (loudness) also peaks around the syllable boundary.

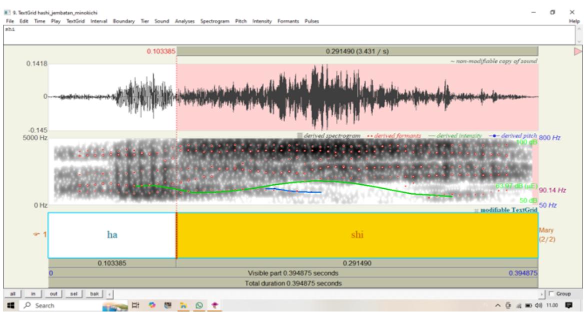

Figure 10: Pronunciation duration of the syllable [shi].

Figure 11 illustrates the initial frequency of this word, which is also the lowest frequency in the frequency measurement, at 106.1 Hz with marked pitch points at 0.062438 seconds. A frequency value of roughly 106.1 Hz is shown on the left side of the graph, parallel to the red dot. The lowest fundamental frequency detected at the start of this word's pronunciation is 106.1 Hz. The pitch contour's blue line, which follows the red dot, indicates a slow rise in frequency. This suggests that as the word advances, the speaker's pitch rises. These pitch changes are an important element in prosody, which includes intonation and tone patterns that can carry linguistic or emotional meaning. Overall, the word utterance with the lowest observed fundamental frequency—106.1 Hz—begins with the red dot in this figure. This is a crucial place to start if you want to comprehend how the speaker's pitch shifts as the word is uttered; in this instance, the trend is upward after the outset.

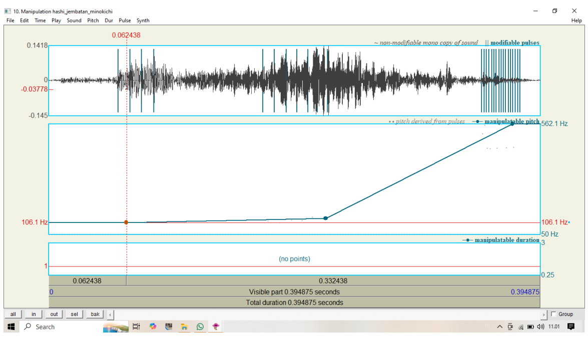

Figure 11: Initial and Lowest Frequency.

Figure 12 shows that the final and maximum frequency is at 559.6 Hz. The total duration of the utterance is approximately 0.394875 seconds, with marked pitch points at 0.332438 seconds. The waveform, which shows the speech signal's amplitude over time, is at the top. Pulses, which are moments of vocal fold vibration necessary for pitch tracking and synthesis, are indicated by vertical blue lines. A particular time point is highlighted for study or editing by the red vertical cursor. The pitch tier, known as "manipulatable pitch," is displayed in the main panel. Two editable pitch targets are connected by a blue line. The starting pitch, which reflects a sharply ascending intonation contour, begins at 106.1 Hz and rises steadily to about 562.1 Hz. The term "bridge" (*hashi*) refers to an unaccented prosodic rhythm, which is typified by this rising pitch. The duration tier, which is currently empty and shows that no time-stretching has been used, is below this. Researchers could mimic variations in speech tempo if it were altered.

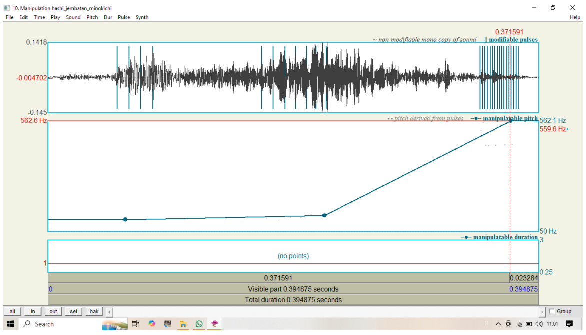

Figure 12: Final and Highest Frequency.

Hashi *(橋) Means 'Chopsticks' (箸)*

The word *hashi*, which means 'chopsticks' (箸), also has two syllables, the syllables [ha] and [shi], which can then be measured for the difference in pronunciation duration between the two syllables. The analysis of *hashi* (箸), meaning 'chopsticks', shows in Figures 13, 14, 15, 16, 17, and 18.

Through the measurements that have been conducted seen in Figure 13, the duration of the syllable pronunciation in the word *hashi* (箸), which means 'chopsticks', was found. Where it is that the pronunciation of the syllable [ha] has a duration of 0.132605 s. The syllable [ha] spans about 0.132605 seconds. The loudness of the sound [ha] about the subsequent segment is indicated by the amplitude of the waveform in this part. This waveform's vibration pattern physically depicts how the sound is produced [ha]. The presence of a sliding consonant component is indicated by the aperiodic interval at the beginning [h]. A spectrogram can be seen beneath the sound waveform. This spectrogram shows the sound's frequency content over time.

Time is on the horizontal axis, and frequency is on the vertical axis. The intensity of energy at a specific frequency is indicated by the color's blackness. The voiceless consonant [h] is characterized by a somewhat flat and diffuse spectrum at the beginning of the high frequencies, as in the word [ha]. The formant pattern that is typical for the vowel [a] will then be shown. Darker horizontal bands will indicate the presence of the first formant (F1) and the second formant (F2). The quality of the vowel [a] following the consonant [h] is indicated by the location and separation between the F1 and F2.

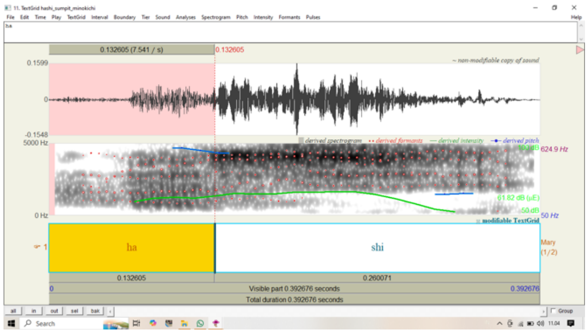

Figure 13: Duration of Pronunciation of the Syllable [ha].

In Figure 14, the pronunciation of the syllable [shi] has a duration of 0.132605 s. The total duration of the utterance is approximately 0.392676 seconds, and it is segmented into two syllables: [ha] and [shi], as indicated in the TextGrid layer at the bottom. The syllable [shi] extends for a significantly longer duration, roughly 0.260071 seconds, suggesting greater emphasis on the second syllable. The red vertical line indicates the syllable boundary, while the waveform at the top shows the voice signal's amplitude over time. Below this, the spectrogram displays the utterance's frequency content, with darker regions indicating stronger acoustic energy. Red dots on the spectrogram indicate formant frequencies, or the vocal tract's resonant frequencies; a green line shows the pitch contour; a blue line indicates loudness; and the pitch value peaks at about 624.9 Hz during the [shi] syllable. The pitch contour displays a pattern that is consistent with an accented word's initial pitch decrease. In keeping with the prosodic characteristics of *hashi*, which means 'chopstick', the pitch begins high at [ha] and decreases considerably on [shi].

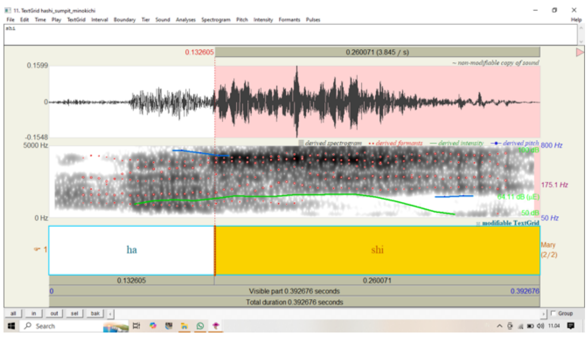

Figure 14: Duration of Pronunciation of the Syllable [shi].

Additionally, in Figure 15, the initial frequency was found at 116.5 Hz and is also the lowest frequency in the word *hashi* (箸), which means chopsticks. Similar to the preceding word's frequency analysis, this analysis of the sound waveform, a graphic depiction of the variations in air pressure as the word is pronounced, is displayed in the upper portion. The glottal pulses, which are caused by the vocal cords vibrating, are depicted by the blue vertical lines. The pitch contour is visible in the image's center. The shift in the sound's fundamental frequency over time is depicted by the blue line with control points. At a time of roughly 0.091338 seconds, the red dot marks the start of the pitch contour. The frequency value of roughly 116.5 Hz is shown on the left side of the graph, parallel to the red dot. The red dot indicates the voice's fundamental frequency at the start of the word pronunciation, which is 116.5 Hz. This frequency, which is created by the speed at which the vocal cords vibrate, indicates the speaker's voice's high and low pitch at that moment. The beginning marker of the voice pitch in this word is this red dot. The red dot indicates the voice's fundamental frequency at the start of the word pronunciation, which is 116.5 Hz. This frequency, which is created by the speed at which the vocal cords vibrate, indicates the speaker's voice's high and low pitch at that moment. The beginning marker of the voice pitch in this word is this red dot.

Figure 16 shows the final frequency result at 573.8 Hz. The red dot is situated at the end of the pitch contour, at a time of roughly 0.251338 seconds, as can be observed in this analytical proof. The graph shows frequency values of roughly 573.8 Hz on the left and right sides, parallel to the red dot. The red dot at the end of the word pronunciation indicates the voice's fundamental frequency, which is 573.8 Hz. This shows how high and low the speaker's voice is at that precise moment.

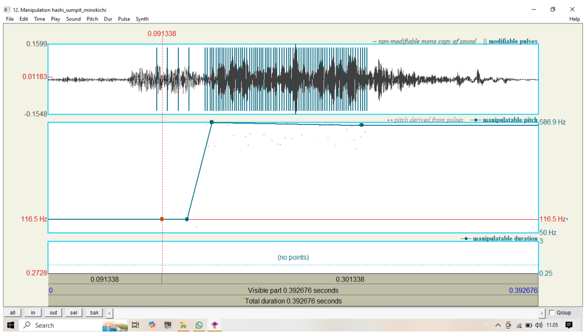

Figure 15: Initial Frequency.

As comparing it to the red dot in the previous figure, which is around 116.5 Hz at the beginning of the word, the pitch increases significantly as this word is pronounced, peaking around the red dot. The entire pitch shift during word utterance is depicted by the blue line joining the pitch control locations. The fundamental frequency after the word, which is quite high in this instance, is indicated by the red dot. Such abrupt shifts in pitch could be a sign of emotional content in the speech, stress, or even unique intonation. All things considered, the red frequency points in this image give precise details on the pitch at the word's utterance end, which is approximately 573.8 Hz. Compared to the word's commencement, this fundamental frequency is extremely high, suggesting a substantial shift in intonation during the utterance.

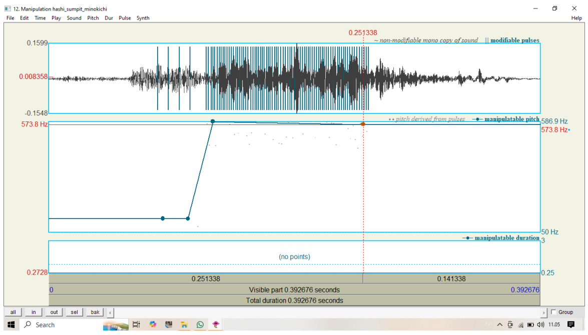

Figure 16: Final Frequency.

Meanwhile, in Figure 17, the highest frequency is at 584.7 Hz. At a time of roughly 0.131245 seconds, the red dot marks the start of the pitch contour. A frequency value of roughly 587.4 Hz is shown on the left side of the graph, parallel to the red dot. The red dot at the beginning of the word indicates the voice's fundamental frequency, which is 587.4 Hz. The pace of the vocal cords produces the pitch of the speaker's voice at that precise moment, which is represented by this frequency. The start of this word's extremely high pitch is indicated by this red dot. The pitch contour's blue line, which follows the red dot, displays a sharp and quick drop in frequency. This suggests that the speaker's pitch peaks at the beginning of the word and then rapidly becomes much lower. Such drastic shifts in pitch frequently convey significant prosodic meanings, such as emphasis, the conclusion of a sentence, or the expression of a particular emotion. Time in seconds is shown on the graph's horizontal axis. Overall, the word starts at the red frequency point in this image, which has a very high fundamental frequency of 587.4 Hz. This is a crucial place to start if you want to comprehend how the speaker's pitch shifts during the word. In this

e- ISSN 2528-5548 | p-ISSN 27764478

instance, the pitch drops off quite sharply following the high onset.

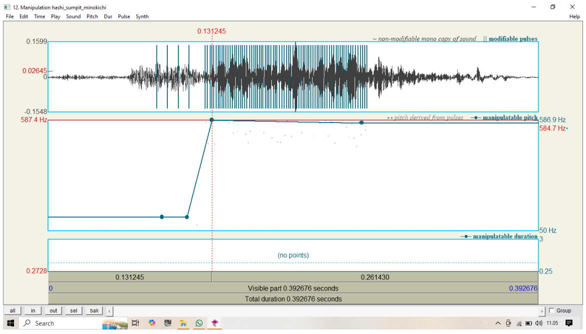

Figure 17: Highest Frequency.

In Figure 18, the lowest frequency is at 116.0 Hz. The total duration of the sound is 0.392676 seconds, and the syllables [ha] and [shi] are separated in the waveform and pitch tiers. The waveform displays the sound signal's amplitude at the top. The primary voiced section is represented by the region between around 0.091338 s and 0.301338 s, where a large number of modification pulses (vertical blue lines) show where the signal can be adjusted for prosodic variations. The word's pitch trajectory is depicted by the middle part, or manipulable pitch tier. The pitch is fixed at about 116.5 Hz at first, but it climbs abruptly to about 568.9 Hz. By the pronunciation of *hashi*, which means 'chopsticks', which is usually indicated by a non-accented or rising pitch, this artificial pitch adjustment implies the development of a rising intonation pattern. The pitch-derived-from-pulses line shows how the software interpolates the pitch contour from vocal fold pulses. There has been no temporal lengthening or shortening done because the duration tier at the bottom is empty (no alterations). Thus, the main adjustment in this case is intonational (pitch-based), changing the prosody to use pitch accent to differentiate lexical meaning.

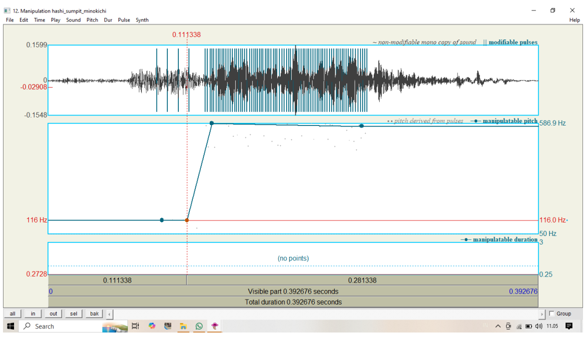

Figure 18: Lowest Frequency.

Pitch and frequency are prioritized in Japanese, therefore, it's critical to discern between words' meanings (Goss & Tamaoka, 2019; Idemaru, Wei, & Gubbins, 2019). It is possible to draw conclusions and compare the two pronunciations of "*ame*" with distinct meanings based on the total amount of data gathered. Table 1 below displays the comparison results.

Table 1: Comparison of the pronunciation duration of the word "*ame*".

	[a]	[me]
Ame (雨)	0,167574 s	0.235306 s
Ame (飴)	0,127079 s	0,243533 s

It is evident from Table 1, which shows the comparison above, that the syllable [a] in the word *ame* (雨), which means rain, is pronounced with more duration than in the term *ame* (飴), which means candy. The term *ame* (飴), which means candy, has a greater duration than the word *ame* (雨), which means rain, when it comes to the pronunciation of the syllable [me]. The comparation of the two words' frequencies as seen in Table 2.

Table 2: Comparison of the frequency.

	Initial	Final	High	Low
Ame (雨)	119,9 Hz	124,6 Hz	124,6 Hz	119,9 Hz
Ame (飴)	100,7 Hz	115,3 Hz	115,3 Hz	100,7 Hz

The frequency of the initial, final, high, and low pronunciations of the word *ame* (雨), which means rain, is higher than that of the word *ame* (飴), which means candy, as Table 2 makes evident.

By comparing the measuring findings of the two words "*hashi*," one might make inferences about this term. According to the comparison from Table 3, the term *hashi* (箸), which means chopsticks, has a greater duration in the pronunciation of the syllable [ha] than the word *hashi* (橋), which means bridge. However, the term *hashi* (橋), which means bridge, has a much longer length in the pronunciation of the syllable [shi] than the phrase *hashi* (箸), which means chopsticks.

Table 3: Comparison of the Duration from "*Hashi*".

	[ha]	[shi]
hashi (橋)	0,103385 s	0,291490 s
hashi (箸)	0,132605 s	0,132605 s

e- ISSN 2528-5548 | p-ISSN 27764478

The comparison table 4 revealed that the word *hashi* (箸), which means chopsticks, has a different frequency in each category, while the word *hashi* (橋), which means bridge, has the same frequency at the beginning and lowest, as well as at the end and highest. Additionally, this comparison revealed that the word *hashi* (箸), which means chopsticks, has a higher value at the initial, final, and lowest frequencies, while the word *hashi* (橋), which means bridge, has a higher frequency value in the highest category.

Table 4: Comparison of the frequency.

	Initial	**Final**	**High**	**Low**
hashi (橋)	106,1 Hz	559,6 Hz	559,6 Hz	106,1 Hz
hashi (箸)	116,5 Hz	573,8 Hz	584,7 Hz	116,0 Hz

It can be concluded that the Japanese language has a distinct phonological structure based on measurements made using Praat. (De Jong & Wempe, 2009). Where syllable frequency and pitch are used to pronounce homophones. It can be concluded from measurements made using Praat that the Japanese language has a distinct phonological structure. Where pitch and syllable frequency are used to pronounce homophones (Chen, Chen, & Li, 2019; Furukawa & Nakamura, 2024; Takahashi, Onohara, & Ihara, 2023).

CONCLUSION

The results show that for "a*me*," the syllable [a] has a longer duration in the meaning of "rain," whereas the syllable [me] has a longer duration in the meaning of 'candy.' The overall frequency is also higher for the meaning 'rain.' For *hashi*, the syllable [ha] has a longer duration in the meaning of 'chopsticks,' while the syllable [shi] has a much longer duration in the meaning of 'bridge.' The frequency patterns also differ between the two meanings, where the initial, final, and lowest frequencies of the word *hashi* (箸) meaning chopsticks have higher values, while in the highest category, the word *hashi* (橋) meaning bridge has higher frequency values. Thus, it may be said that the Japanese language's pitch accent and frequency patterns are crucial in defining word meanings.

ACKNOWLEDGMENT

The University of Sumatra Utara and the course lecturers who helped with the journal writing are both acknowledged by the researcher. The resource people who were willing to supply research data are also thanked by the author. Every step of the study process is carried out entirely by the researcher alone, without any other funding.

REFERENCES

Chen, X. S., Chen, S., & Li, B. (2019). The influence of speech prosodic structure and syllable composition on consonant VOT in Japanese. *The Journal of the Acoustical Society of America, 146*(4_Supplement), 3009–3009. https://doi.org/10.1121/1.5137418

Davidson, K., & Lee, Y. (2023). Acoustic analysis of speech sounds in phonetics. *Journal of Phonetic Research, 28*(3), 45-62.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Eggins, S. (2004). *An introduction to Systemic Functional Linguistics* (2nd ed.). London: Continuum.

Fry, D. B. (1955). The physics of speech. *Acoustical Society of America Journal, 28*(6), 1132-1146.

Fry, D. B. (1958). The acoustic aspects of the phoneme. *Journal of the Acoustical Society of America, 30*(9), 1342-1347.

Furukawa, K., & Nakamura, S. (2024). Boundary-driven downstep induced by syntax–prosody mapping. *The Journal of the Acoustical Society of America, 156*(2), 1440–1460. https://doi.org/10.1121/10.0028340

Gordon, M. (2004). A phonological and phonetic study of word-level stress in Chickasaw. *International journal of American linguistics, 70*(1), 1-32.

Goss, S. J., & Tamaoka, K. (2019). Lexical accent perception in highly proficient L2 Japanese learners: The roles of language-specific experience and domain-general resources. *Second Language Research, 35*(3), 351–376. https://doi.org/10.1177/0267658318775143

Halliday, M. A. K. (2002). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Continuum.

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching, 54*(3), 297–326. https://doi.org/10.1017/S0261444820000658.

Idemaru, K., Wei, P., & Gubbins, L. (2019). Acoustic Sources of Accent in Second Language Japanese Speech. *Language and Speech, 62*(2), 333–357. https://doi.org/10.1177/0023830918773118

e- ISSN 2528-5548 | p-ISSN 27764478

Ito, J., & Mester, A. (2013). Prosodic subcategories in Japanese. *Lingua, 124*, 20–40. https://doi.org/10.1016/j.lingua.2012.08.016

Karsono, H. (2013). *Analisis Akustik Fonetik Menggunakan Praat [Phonetic Acoustic Analysis Using Praat]*. Yogyakarta: Pustaka Pelajar.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*(2), 286. https://doi.org/10.2307/413642

Lidwina, F. (2012). Kajian fonetik auditori dalam pengajaran Bahasa [The study of auditory phonetics in language teaching]. *Linguistik Journal, 12*(1), 45-58.

Martin, J. R. (1992). *English text: System and Structure*. Amsterdam: John Benjamins.

Mennen, I., Scobbie, J. M., De Leeuw, E., Schaeffler, S., & Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research, 26*(1), 13–41. https://doi.org/10.1177/0267658309337617

Miller, G., & Brown, J. (2018). *Speech Acoustics and Language*. New Jersey: Wiley.

Mücke, D., Hermes, A., & Tilsen, S. (2020). Incongruencies between phonological theory and phonetic measurement. *Phonology, 37*(1), 133–170. https://doi.org/10.1017/S0952675720000068

Ningsih, H. (2007). Fonetik dan fonologi dalam kajian linguistic [Phonetics and phonology in linguistic studies]. *Jurnal Linguistik, 10*(2), 42-50.

Odden, D. (2005). *Introducing Phonology*. Cambridge: Cambridge University Press.

Peterson, R., & Henderson, J. (2020). Phonetic features and their role in speech perception. *Speech and Hearing Studies, 14*(2), 101-115.

Roettger, T. B. (2019). Researchers' degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10*(1), 1. https://doi.org/10.5334/labphon.147

Takahashi, T., Onohara, A., & Ihara, Y. (2023). Bayesian phylogenetic analysis of pitch-accent systems based on accentual class merger: A new method applied to Japanese dialects. *Journal of Language Evolution, 8*(2), 169–191. https://doi.org/10.1093/jole/lzae004

Tjandra, S. N. (2004). Fenomena erologi Jepang-Indonesia pada mahasiswa bahasa Jepang tahap menengah di Universitas Indonesia [The phenomenon of Japanese-Indonesian erology in Intermediate Japanese language students at the University of Indonesia]. *Makara Human Behavior Studies in Asia, 8*(1), 33-38.

Vigário, M., Frota, S., & João, M. (2011). Phonetics and Phonology: Interactions and Interrelations. *Phonetica, 68*(1–2), 113–115. https://doi.org/10.1159/000329393

Widodo, H., & Purnama, E. (2020). *Sociolinguistics in Language and Culture*. London: SAGE Publications.

Wati, F., & Hidayat, I. (2022). Perubahan sosial dan bahasa: Implikasinya terhadap komunikasi [Social and language change: Implications for communication]. *Jurnal Ilmu Komunikasi, 21*(4), 59-70.

White, P., & Taylor, K. (2021). *Auditory Phonetics and Sound Perception*. London: Routledge.

e- ISSN 2528-5548 | p-ISSN 27764478