JOLLT Journal of Languages and Language Teaching
https://e-journal.undikma.ac.id/index.php/jollt
Email: jollt@undikma.ac.id
DOI: https://doi.org/10.33394/jollt.v14i2.18880

April 2026. Vol. 14, No.2
p-ISSN: 2338-0810
e-ISSN: 2621-1378
pp. 641-652

DIPHTONG MONOPHTONGIZATION IN JAPANESE LOANWOARDS: AN ACOUSTIC AND SOCIOLINGUISTIC ANALYSIS OF LANGUAGE CONTACT EFFECTS

[1]*Nursyifa Azzahro, [2]Naelul Rohmah, [1]Mauldina Permata Dewi, [1]Alvyna Febrianti
[1]Japanese Literature, Prima Bangsa Institute, Jl. Brigjend Dharsono Bypass No.20, Cideng, Kertawinangun, Cirebon, West Java, Indonesia
[2]English Literature, Prima Bangsa Institute, Jl. Brigjend Dharsono Bypass No.20, Kertawinangun, Cirebon, West Java, Indonesia

*Corresponding Author Email: nursyifa@ipbcirebon.ac.id

Article Info
Article History
Received: January 2026
Revised: December 2025
Accepted: March 2026
Published: April 2026

Keywords
Diphtong analysis;
Japanese phonology;
Language contact;
Loanwords;
Sociolinguistic analysis;

Abstract

This study explores the monophthongization of diphthongs in Japanese loanwords, with a particular focus on the role of foreign language contact. Previous studies have failed to integrate acoustic evidence with sociolinguistic perception in explaining phonological adaptation in Japanese loanwards. This study fills this gap by integrating acoustic phonetic analysis and sociolinguistic perspectives. The dataset consists of 50 Japanese loanwords containing diphthongs (/ai/, /ei/, /oi/, /au/, /ou/), produced by 50 native speakers from two age groups (18–35 and 36–70). A total of 2,500 audio tokens were collected and analyzed using Praat to observe vowel duration and formant transitions. In addition, a perception survey was conducted to assess speakers' attitudes toward monophthongization. The findings indicate a systematic shift from dynamic vowel sequences toward stable long vowel nuclei, suggesting that Japanese speakers are not merely simplifying articulation but restructuring diphthongal sequences into phonologically unified segments. These findings challenge the traditional assumption that Japanese preserves vowel sequences without significant restructuring, suggesting instead that loanward phonology actively reshapes vowel organization under contact pressure. Sociolinguistic results show diverse perceptions: some consider this process natural, while others see it as an impact of media and globalization. The generational differences observed in this stud suggest that monophthongization is not a completed change but a stratified variation, where competing phonetic norms coexist across age groups. This study contributes to the understanding of sound changes in Japanese, supports language education, and provides a curated phonetic dataset for future linguistic and speech technology research.

How to cite: Azzahro, N., Rohmah, N., Dewi, M.P., & Febrianti, A. (2026). Diphtong Monophtongization in Japanese Loanwoards: An Acoustic and Sociolinguistic Analysis of Language Contact Effects. *JOLLT Journal of Languages and Language Teaching, 14*(2), 641-652. Doi: https://doi.org/10.33394/jollt.v14i2.18880

INTRODUCTION

Every language undergoes a process of sound adjustment over time, either due to internal changes or external influences, such as contact with other languages. In Japanese, one phonological phenomenon that is becoming increasingly interesting to study is diphthong monophthongization, which is the process whereby two vowels that occur consecutively in a syllable or phonological unit are simplified into a single long vowel (Azzahro, 2024; Aisyah et al., 2025). This phenomenon is particularly evident in modern loanwords that have entered the language through globalization and cross-linguistic interaction. Although in

standard Japanese phonology, vowel sequences such as /ai/, /ei/, /oi/, /au/, and /oɯ/ are generally considered to be sequential vowels that are retained (Kawakami, 1977), practice in the field shows something different. Many Japanese speakers, both old and young, actually produce simpler vowel forms in their everyday speech (Boersma & Weenink, 2023). For example, the word**タイヤ** (/taiya/, "tire") is often realized as /taːya/, indicating that the vowels /a/ and /i/ are no longer preserved as two separate sounds, but are merged into one long vowel sound. This phenomenon is not only found in this word, but also in many words borrowed from English, German, Portuguese, and other languages that have historically entered the Japanese language (Kubozono, 1999).

In the era of globalization, the massive influx of information from outside Japan has also played a role in accelerating phonological adaptation. Social media, online videos, international music, dramas, movies, and online games provide intensive exposure to foreign pronunciations. Interestingly, however, rather than retaining these foreign diphthongs, most speakers continue to adapt them to the simpler Japanese phonological system (Boersma & Weenink, 2023; Johnson, 2012). This raises an important question: to what extent does contact with foreign languages actually influence the way Japanese speakers produce diphthongs? Is the resulting monophthongization purely a phonological adaptation, or is it also the result of social negotiation, language ideology, and speakers' perceptions of "correct pronunciation"?

Previous studies have discussed diphthong reduction in Japanese from the perspective of generative phonological theory (Azzahro, 2024; Aisyah et al, 2025) or through articulatory description. But these studies have not yet combined phonetic-acoustic analysis with a sociolinguistic approach. In fact, to understand sound changes comprehensively, two types of data are needed: first, objective and measurable acoustic data; second, social perception data that reflects how speakers understand, accept, or reject certain sound forms. However, existing studies fail to explain how phonetic realization and speaker perception interact in shaping this change. Therefore, this study combines both approaches to provide a more comprehensive picture of the phenomenon of diphthong monophthongization in Japanese loanwords. Acoustic analysis was conducted using Praat to see how sound changes occur at the formant, duration, intensity, and pitch contour levels. Meanwhile, the sociolinguistic approach was conducted through surveys and interviews to understand how speakers interpret this phenomenon in their daily lives.

This study has three main objectives, namely 1) to identify how foreign language contact affects the monophthongization of diphthongs in Japanese, 2) to analyze phonological patterns in loanwords that undergo monophthongization, and 3) to evaluate speakers' perceptions and the social implications of this process. This study positions diphthong monophthongization in Japanese loanwords as a case of contact-driven phonological adaptation, supported by sociophonetic evidence. Rather than treating monophthongization as a purely articulatory phenomenon, this research argues that it reflects the interaction between phonological constraints and social exposure to foreign language input. Theoretically, this research is expected to fill a gap in the study of sound changes in the context of foreign language contact. Practically, the results of this research can be useful in Japanese language teaching, especially for non-native learners, and can support the development of language technology such as speech recognition and text-to-speech systems.

RESEARCH METHOD

Research Design

This study uses a convergent parallel mixed-methods design, which is a research design in which quantitative and qualitative data are collected separately but analyzed simultaneously, and the results are then integrated to obtain a more comprehensive understanding. This design is one of the main approaches in mixed research because it allows

researchers to view phenomena from two different but complementary angles (Creswell & Plano Clark, 2018).

The application of this design is important in phonological research, particularly in the analysis of diphthong monophthongization, because sound changes are not only phonetic processes but also social phenomena. Acoustic data is needed to explain what happens to sounds, for example, how formant transitions indicate a shift from diphthongs to monophthongs (Johnson, 2012; Ladefoged & Johnson, 2015). Meanwhile, sociolinguistic data such as speaker perception and language attitudes are needed to understand why the changes occur and how speakers interpret different sound forms (Holmes & Wilson, 2017; Wardhaugh & Fuller, 2021). Using a convergent parallel design, both types of data can be analyzed independently but then combined. This approach is in line with Labov's (2001) argument that sound changes must be understood through the integration of phonological structure and social dynamics. Therefore, this design provides a more complete and in-depth picture of the monophthongization of diphthongs in Japanese, as it not only reveals the form of the sound change but also the social factors that influence this pattern.

Research Participants

The study involved 50 native Japanese speakers, divided into two age groups: 1) the 18–35 age group (young speakers), and 2) the 36–70 age group (adult and elderly speakers). This age range was chosen based on the consideration that sound changes often begin with the younger group, so comparing the two groups could provide an indication of whether monophthongization is a change that is in a transitional phase or a change that has already been established. Each participant was asked to pronounce 50 loanwords containing the diphthongs /ai/, /ei/, /oi/, /au/, and /oɯ/. A total of 2,500 audio tokens were recorded and then analyzed acoustically. For the qualitative component, this study involved subjects who were native Japanese speakers willing to be interviewed about their perceptions of diphthong monophthongization in loanwords. The subjects were selected based on their level of exposure to foreign languages, use of digital media, and experience interacting with variations of modern Japanese.

The research subjects were interviewed using a semi-structured approach to explore 1) their views on diphthongal vs. monophthongal forms, 2) their social assessment of pronunciation variations, 3) the factors that influence pronunciation preferences, and 4) their perceptions of the influence of foreign languages and media in shaping sound changes. This approach provides a deeper understanding of social and perceptual aspects that cannot be captured through acoustic analysis alone.

Instruments

This study uses three main types of instruments to collect and analyze data, namely Praat phonetic software, perception questionnaires, and semi-structured interview guides. These three instruments were chosen because they provide a complementary picture of the phenomenon of diphthong monophthongization, both in terms of sound production and in terms of speakers' perception and social assessment.

Phonetic Software: Praat

Acoustic phonetic analysis can be conducted using *Praat* software, which is widely employed in linguistic research to measure speech parameters such as frequency, duration, and intensity through spectrographic visualization and formant analysis (Mawarni et al., 2024). Praat has become an international standard in phonetic research due to its ability to display and measure acoustic aspects in detail, such as 1) Formants (F1, F2): Used to see differences in vowel quality. Changes in F1 and F2 can indicate whether a diphthong retains two vowel targets or has been simplified into a monophthong (Johnson, 2012; Ladefoged & Johnson, 2015), 2) Vowel duration: Important for determining whether the combination of

two vowels produces a long vowel with stable duration, 3) Fundamental frequency (F0): Used to observe intonation patterns and changes in vowel stability, 4) Intensity (dB): Reflects sound energy, useful for assessing whether there is a decrease in energy that typically occurs in the process of sound reduction, and 5) Spectrogram: Visually displays the acoustic structure so that the transition movement of diphthongs can be observed more clearly (Boersma & Weenink, 2023). Praat has been utilized in previous studies on Japanese phonetics to investigate acoustic properties of speech sounds (Virdaus, 2023). Using Praat allows researchers to perform objective and standardized analyses that can be replicated by other researchers.

Perception Questionnaire

The second instrument is a perception questionnaire, which is used to collect qualitative data on how Japanese speakers assess the monophthongization of diphthongs in loanwords. This questionnaire was designed to explore 1) speakers' preferences between diphthongal and monophthongal pronunciations, 2) the reasons behind those choices, 3) the extent to which they consider foreign languages or media to have influenced certain pronunciation patterns, and 4) whether they view monophthongization as part of the natural development of the Japanese language. This approach is in line with sociolinguistic principles that emphasize the importance of language perception and attitudes in understanding sound changes (Holmes & Wilson, 2017; Wardhaugh & Fuller, 2021). Speaker perception is often an indicator of whether a phonetic change will persist, disappear, or develop into a standard variation.

Semi-Structured Interview Guide

The third instrument is a semi-structured interview guide, which is used to explore qualitative data in greater depth. Through interviews, researchers can explore personal experiences and social factors that influence how speakers produce and evaluate phonological variation. The interview guide includes questions related to 1) speakers' experiences in learning or hearing foreign languages, 2) their level of exposure to international media (e.g., movies, anime, music, or digital content), 3) their general attitudes toward loanwords and sound changes in Japanese, and 4) their perceptions of "correct pronunciation" in a modern context. Semi-structured interviews were chosen because they are flexible, allowing for the exploration of new information that might not emerge in a questionnaire. This technique is commonly used in research on language identity and phonological variables because it provides richer insights into the social dimensions of a phenomenon (Labov, 2001; Tagliamonte, 2006).

Data Analysis

Data analysis in this study was conducted by combining acoustic-phonetic and sociolinguistic approaches to understand how the monophthongization of diphthongs in Japanese loanwords occurs and is perceived by speakers. The entire analysis process was carried out in stages so that each dimension of the data (both sound and perception) could be examined thoroughly.

Acoustic-Phonetic Analysis

The audio data obtained from 50 speakers was processed and analyzed using Praat software. Each utterance was extracted per vowel segment to obtain the following values: 1) First formant (F1) as an indicator of vowel height, 2) Second formant (F2) as an indicator of vowel fronting or backing, 3) Third formant (F3) as a support for resonance characteristics, 4) Vowel duration, 5) Pitch (fundamental frequency), and 5) Intensity. Formant values were taken at representative points (20%, 50%, and 80% of vowel duration) to see whether there was a transition, which is characteristic of diphthongs, or stabilization, which indicates monophthongization. Spectrograms and waveforms were used to observe the energy patterns

and temporal structure of vowels, so that any changes could be analyzed both visually and numerically. The results of this analysis are then compared between age groups to see if there are significant differences in production patterns between young speakers (18–35 years old) and adult/elderly speakers (36–70 years old). This approach allows for the identification of whether monophthongization is generational or has spread evenly throughout the speech community.

Sociolinguistic Analysis

In addition to acoustic analysis, this study also included a sociolinguistic component to understand how speakers interpret the phenomenon of monophthongization in everyday life. Data was collected through a perception survey covering 1) speakers' awareness of sound changes, 2) attitudes toward the use of diphthongs vs. monophthongs in loanwords, 3) the influence of exposure to foreign languages (media, education, travel, etc.), and 4) preferences for pronunciation forms considered "correct" or "natural." Respondents' answers were then categorized into three perception groups: 1) Considering monophthongization natural, 2) Associating it with the influence of globalization, and 3) Having no clear opinion. This data was analyzed to see whether these perceptions were related to the acoustic production patterns found in the previous stage. By linking the two types of data, the study was able to identify the relationship between how speakers produce sounds and how they assess these changes.

Integration of Phonetic and Sociolinguistic Analysis

The final stage of the analysis was to combine the acoustic and social findings to see the dynamics of sound change more comprehensively. The relationship between production and perception was examined through: Comparing monophthongization tendencies in formant data, 1) Matching production patterns with speaker attitudes, and 2) Variations based on age, language experience, and media exposure. With this integrative approach, the study not only describes how diphthongs undergo phonetic change, but also explains why these changes are accepted, maintained, or even noticed by the speaker community.

RESEARCH FINDINGS AND DISCUSSION

Research Findings

Acoustic-Phonetic Analysis

The results of acoustic analysis of the five main diphthongs in Japanese, namely /ai/, /ei/, /oi/, /au/, and /oɯ/, show that monophthongization patterns occur consistently across all speech data studied. However, the degree of monophthongization varies between diphthongs, with some diphthongs showing complete simplification, while others still retain slight traces of transition.

Table 1
Acoustic-Phonetic Findings Based on Diphthongs

Diftong	F1 (Hz)	F2 (Hz)	F3 (Hz)	Duration (ms)	Pitch (Hz)	Intensity(dB)
/ai/	~650	~1200	~2600	170-190	200-215	70-73
/ei/	~520	~1800	~2700	160-180	195-210	68-72
/oi/	~540	~1300	~2650	170-185	205-220	69-72
/au/	~600	~1100	~2550	175-190	190-205	70-74
/oɯ/	~500	~900	~2500	155-170	185-200	68-71

Monophthongization is consistently observed across all diphthong types, indicating a systematic shift toward simplified vowel structures. This is evidenced by stable formant values and the absence of significant transitional movement, as shown in Table 1. These patterns suggest that speakers are not merely reducing articulation but are restructuring diphthongal sequences into phonologically unified segments. The diphthongs /ai/, /ei/, and /oɯ/ are the groups with the most stable and uniform monophthongization, while /oi/ shows

intergenerational variation, and /au/ is at a medium level of simplification. These findings show that sound changes in Japanese are uneven, but tend to move towards a single long vowel system that is more consistent with the phonological structure of Japanese.

Table 2
Summary of Acoustic-Phonetic Findings Based on Parameters

Parameter	Function	Finding
F1	Vowel height	Stable (no movement toward a second vowel)
F2	Vowel frontness-backness	No transition (loss of diphtonal features)
F3	Resonance	Stable (Single vowel quality)
Duration	Vowel length	Single phase (monophthong)
Pitch	Fundamental frequency	Linear (no dual vowel nuclei)
Intensity	Acoustic energy	Single peak (single vowel)

The acoustic results consistently demonstrate that diphthongs in the dataset are realized as monophthongal units rather than dynamic vowel sequences. This pattern is evidenced by the stability of key acoustic parameters, including F1, F2, and F3, which show minimal variation throughout the vowel duration, as presented in Table 2. The absence of significant F1 and F2 transitions indicates that speakers do not actively move toward a secondary vowel target, while the stability of F3 further supports the presence of a single, unified resonance structure. Taken together, these findings suggest that diphthongal sequences are not merely reduced in articulation, but are produced as phonologically stable long vowels. Temporal and prosodic parameters also support these findings. The vowel duration shows only a single phase without the two distinct durations characteristic of a true diphthong, while the pitch shows a smooth, linear contour without the two peaks that typically occur when two vowels are arranged in a syllable.

Sociolinguistic Analysis

In addition to acoustic analysis, this study also explores how speakers understand and assess these sound changes. The results of surveys and interviews show that speakers' attitudes toward monophthongization vary considerably. Some speakers view monophthongization as natural. According to them, sound simplification is part of the natural development of the Japanese language, especially since the Japanese phonological system does not recognize diphthongs as independent phonemic units. For this group, monophthongization is not a "mistake," but rather an adaptation that is in harmony with the phonological structure of the Japanese language.

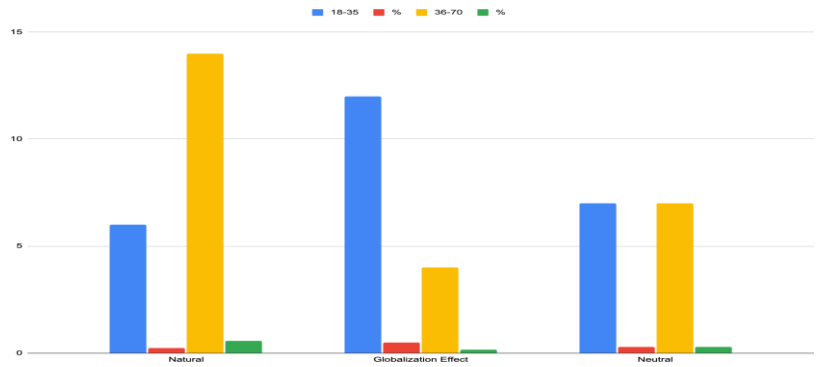

Figure 1. Speaker Perception of Diphtong Monophtongization

However, there are also speakers who consider monophthongization to be the result of external influences, particularly the media and globalization. This group argues that exposure to foreign languages through films, music, games, and social media has encouraged the emergence of new pronunciations, which ultimately influence the way Japanese people pronounce loanwords. For them, this change is not entirely natural, but rather a social

response to the increasing contact with foreign languages in everyday life.

This difference in perception is quite clear between the older and younger generations. Younger speakers are generally more aware that the original form of loanwords should contain diphthongs. Their exposure to the original pronunciation, for example through English-language digital content, makes them more familiar with the source form. Therefore, even though they pronounce the monophthong form, they remain aware that the diphthongal form is closer to the donor language. In addition, speakers who have a high level of exposure to foreign languages or who use foreign languages daily in their work or education show a more open attitude towards pronunciation variations. They are more tolerant of both monophthong and diphthong forms. This group also tends to view sound variation as something flexible and part of the social dynamics of language. Thus, sociolinguistic findings show that phonetic changes cannot be separated from social factors. Perception, language experience, level of exposure to foreign languages, and the generation to which speakers belong play an important role in influencing how sounds are produced and maintained in speech communities.

Speaker Perception and Social Impact on Diphthong Monophthongization

Research results related to speaker perception show that the monophthongization of diphthongs in loanwords is not only influenced by phonetic mechanisms, but also by social experience, exposure to foreign languages, and the language attitudes of each speaker. Based on a perception survey of 50 participants, it was found that speakers were divided into three main attitude groups: those who consider monophthongization to be natural, those who see it as an influence of globalization and the media, and those who do not have a strong opinion.

The largest group consists of speakers who consider monophthongization to be a natural process in the phonological development of the Japanese language. They argue that sound changes such as the simplification of diphthongs into long vowels are natural because Japanese does not inherently recognize full diphthongs. Speakers in this group are generally adults and the elderly. Their attitudes are consistent with the stable vowel production patterns found in acoustic analysis, indicating a correlation between phonetic practice and language attitudes.

The second group, which consists mostly of young speakers, views the phenomenon of monophthongization as a result of the influence of globalization, digital media, and contact with foreign languages, especially English. They are aware that the original form of loanwords in the source language has clear diphthongs, so the change to monophthongs is considered a form of linguistic adaptation. Young speakers in this group demonstrate higher language awareness, mainly because they hear the original pronunciation more often through the internet, games, music, and movies. This makes them more sensitive to phonological differences between Japanese and foreign languages. The last group consists of speakers who do not have a clear opinion. They tend to be unaware of the difference between diphthongs and monophthongs or do not see this phenomenon as important. Interestingly, the acoustic production patterns of this group also tend to be unstable; they sometimes produce residual diphthongal transitions, but at other times produce full monophthongs. This indecisive attitude seems to reflect a lack of attention to phonetic details and minimal exposure to the original forms of loanwords.

The results of the study show that speakers' perceptions of monophthongization are influenced by factors such as age, foreign language experience, level of media exposure, and phonetic production habits. The different attitudes of younger speakers compared to older speakers indicate a sociolinguistic shift in which the new generation is more aware of phonetic variations influenced by global language contact. At the same time, the widespread acceptance of monophthongization shows that this process has become part of the phonological identity of modern Japanese.

Discussion

The results of acoustic analysis in this study show a clear pattern that the monophthongization of diphthongs in Japanese loanwords is not merely a random phonetic variation, but rather an ongoing phonological process in the Japanese speech community. Using acoustic parameters such as formants (F1 and F2), vowel duration, pitch, and energy structure on spectrograms, this study confirms that most diphthongs that theoretically exist in loanwords are realized as stable long vowels. The findings of this study strongly support the view that monophthongization in Japanese loanwords operates as a process of phonological adaptation rather than mere phonetic reduction. The consistent absence of formant transitions and the stabilization of vowel quality indicate that speakers are not simply reducing articulatory effort, but are restructuring diphthongal sequences to conform to the phonological constraints of Japanese (Aisyah et al., 2015; Irwin & Sohn, 2020). This suggests that the Japanese phonological system actively reorganizes incoming foreign forms, favoring stable vowel nuclei over dynamic vowel sequences.

The generational differences observed in this study further suggest that monophthongization represents an instance of ongoing sound change in the Labovian sense. Older speakers exhibit more stable monophthongal realizations, while younger speakers display greater variability and partial retention of diphthongal features. This stratified distribution indicates that competing phonetic norms coexist within the speech community, reflecting a change in progress rather than a completed phonological shift (Irwin & Sohn, 2020). Such variation aligns with sociolinguistic models of sound change, where linguistic forms are distributed unevenly across social groups and influenced by patterns of language exposure.

Taken together, these findings position diphthong monophthongization in Japanese loanwords as a contract-driven phonological adaptation process, shaped by both structural constraints and sociolinguistic dynamics. While exposure to foreign languages introduces diphthongal input, the Japanese phonological system systematically reshapes these forms, resulting in stable monophthongal outputs. At the same time, generational variation reveals that this process is not uniform, but socially stratified, reflecting an ongoing negotiation between external influence and internal phonological organization.

Reduction of Formant Movement

The data shows that in diphthongs such as /ai/ and /ei/, the formant movement that should mark the transition from one vowel to another is minimal or does not appear at all. In words such as **タイヤ** /taiya/, speakers in all age groups, especially adults and the elderly, tend to produce vowels with 1) F1 stable around the midpoint of the vowel /a/, 2) F2 stable without any rise towards the target /i/, and 3) Long vowel duration (often more than 0.20–0.25 seconds per vowel segment). The absence of a significant F1–F2 transition indicates that speakers do not direct their articulators towards the two vowel targets. This is consistent with Azzahro (2024) who found that the description of monophthongization as a process of articulatory simplification that commonly occurs in languages that structurally do not support full diphthongs.

Weak Transitions Persist

Although most diphthongs show a pattern of monophthongization, this study found that the diphthong /oi/, for example in the word **コイン** /koin/, often exhibits a slight formant transition—especially in younger speakers. However, this transition is not sufficient to be considered a full diphthong. An F2 change from around 1300 Hz to 1500–1600 Hz occurs, but this shift 1) is very brief, 2) does not reach the ideal /i/ target, and 3) is often omitted in the production of older speakers. This phenomenon indicates that some diphthongs may be at different stages of change, and monophthongization does not occur uniformly across all types

of diphthongs.

Duration as an Indicator of Sound Merging

Vowel duration also provides important indications regarding the process of monophthongization. In diphthongs that still retain a two-element vowel structure, the duration is usually longer because it includes the movement of the articulators toward the second vowel target. However, in this research data, the duration of vowel production in monophthongized diphthongs tends to be 1) short and stable, 2) does not show two phases of articulation, and 3) is consistent across speakers. This is clearly seen in the production of the word メイン /mein/, where younger speakers who retain the diphthong transition have a slightly longer duration than older speakers. Thus, duration can be an indicator of the position of a diphthong in its path of change towards a monophthong.

Single Energy vs. Tiered Energy

Spectrogram analysis provides visual evidence that reinforces the formant findings. In full diphthongs, two vowel energy structures are usually apparent, reflecting two articulatory targets. However, in most loanwords in this study, 1) only one continuous energy block appears, 2) there is no intensity shift at a specific frequency, 3) the two-layer waveform pattern characteristic of diphthongs is not apparent. This single energy structure is characteristic of monophthongization and indicates that speakers produce one long vowel rather than two fused vowels.

Evidence of Ongoing Phonetic Change

One important finding in the acoustic analysis is how production patterns differ based on age. Adult and elderly speakers show more stable production and more frequently monophthongize diphthongs in loanwords. In contrast, younger speakers more often retain traces of diphthongal transitions, especially in words derived from modern English. This phenomenon shows that phonetic change does not occur uniformly, but is influenced by generational factors and exposure to foreign languages. The younger generation, who are more exposed to English through global media, seem to be more sensitive to the original form of loanwords, so their production shows a more "foreign" character than the older generation.

Monofthongization as a Phonological Adaptation of Japanese

In general, the acoustic results of this study support the view that Japanese is in the process of adapting loanwords to better suit its phonological structure. Since Japanese does not recognize full diphthongs as phonemic, speakers naturally simplify vowel movements and move toward more economical articulation. This process is consistent with the theory of sound change: 1) Reduction and assimilation, 2) Target undershoot (speakers do not reach the second articulation target), and 3) Vocalic centralization (formant values move toward central vowels). These patterns have been found in many studies of cross-linguistic phonetic change.

Speaker Perceptions and Social Impacts on the Phenomenon of Diphthong Monophthongization

Research findings related to speaker perceptions show that diphthong monophthongization in loanwords is influenced not only by phonetic mechanisms, but also by social experiences, exposure to foreign languages, and the language attitudes of individual speakers. Based on a perception survey of 50 participants, it was found that speakers were divided into three main attitude groups: those who consider monophthongization to be natural, those who see it as an influence of globalization and the media, and those who do not have a definite opinion.

The largest group consisted of speakers who considered monophthongization to be a natural process in the phonological development of the Japanese language. They argue that

sound changes such as the simplification of diphthongs into long vowels are natural because Japanese does not inherently recognize full diphthongs. Speakers in this group are generally adults and the elderly. Their attitudes are consistent with the stable vowel production patterns found in acoustic analysis, indicating a correlation between phonetic practice and language attitudes.

The second group, which consists mostly of young speakers, views the phenomenon of monophthongization as a result of the influence of globalization, digital media, and contact with foreign languages, especially English. They are aware that the original form of loanwords in the source language has clear diphthongs, so the change to monophthongs is considered a form of linguistic adaptation. Young speakers in this group demonstrate higher language awareness, mainly because they hear the original pronunciation more often through the internet, games, music, and movies. This makes them more sensitive to phonological differences between Japanese and foreign languages. The last group consists of speakers who do not have a clear opinion. They tend to be unaware of the difference between diphthongs and monophthongs or do not see this phenomenon as important. Interestingly, the acoustic production patterns of this group also tend to be unstable—they sometimes produce residual diphthongal transitions, but at other times produce full monophthongs. This indecisive attitude seems to reflect a lack of attention to phonetic details and minimal exposure to the original forms of loanwords.

The results of the study show that speakers' perceptions of monophthongization are influenced by factors such as age, foreign language experience, level of media exposure, and phonetic production habits. The different attitudes of younger speakers compared to older speakers indicate a sociolinguistic shift in which the new generation is more aware of phonetic variations influenced by global language contact. At the same time, the widespread acceptance of monophthongization shows that this process has become part of the phonological identity of modern Japanese.

CONCLUSION

This study shows that the monophthongization of diphthongs in Japanese loanwords is a phenomenon that occurs not only at the phonetic level, but is also influenced by social factors and the linguistic environment of the speakers. From the results of acoustic-phonetic analysis using Praat, it can be seen that most diphthongs such as /ai/, /ei/, and /oi/ are realized as stable single vowels. The formant movement that should mark the transition from one vowel to another has almost completely disappeared in many utterances, especially among adult and elderly speakers. The stability of F1 and F2 values, the tendency for vowel duration to be even, and the absence of layered energy patterns in the spectrogram reinforce the idea that this monophthongization process has become part of the vowel production pattern in contemporary Japanese.

On the other hand, sociolinguistic findings reveal that speakers' perceptions of this phenomenon are not uniform. Some speakers view monophthongization as a natural development in Japanese phonology—a simplification of articulation that is consistent with the sound structure of the language. However, other speakers, especially younger generations, associate this change with the influence of the media and exposure to foreign languages. They are aware of the original form of loanwords in the source language, so they are more sensitive to the existence of diphthongs even though in practice they still use monophthong forms. Meanwhile, another group of speakers occupy a neutral position, indicating that variations in language awareness and experience play a major role in shaping attitudes towards sound changes.

Thus, this study shows that monophthongization cannot be understood solely as an independent phonetic process. This change is the result of interactions between the phonological tendencies of the Japanese language, intergenerational social dynamics, and the

intensity of contact with foreign languages. These findings not only enrich our understanding of sound changes in modern Japanese, but also provide a basis for the development of language teaching, phonological research, and speech-based technology that requires actual acoustic data from native speakers. This study repositions monophthongization as a contract-driven and socially mediated phonological process rather than a purely articulatory simplification. Overall, the monophthongization of diphthongs in loanwords reflects how language continues to adapt to social changes and increasingly intense global interactions.

ACKNOWLEDGEMENT

The author wishes to express sincere appreciation to all parties who contributed to the completion of this research. Special gratitude is extended to the research team in Japan, whose assistance in collecting primary data and coordinating on-site recording procedures was invaluable. The author also acknowledges the contributions of the translation team, whose support in linguistic verification and data refinement greatly facilitated the accuracy of the analysis. The author further extends heartfelt thanks to all participants, who generously devoted their time and cooperation throughout the study. Their willingness to be involved made this research possible. The support and collaboration of every individual and institution involved are deeply appreciated.

FUNDING

This study received no external funding. The research was planned, conducted, and completed using the authors' personal resources and available institutional support, with no financial assistance from public, commercial, or nonprofit organizations. All stages of the work, including preparation, data collection, data analysis, interpretation of findings, and manuscript writing, were carried out independently.

INFORMED CONSENT STATEMENT

Participation in this study was fully voluntary, and informed consent was obtained from all participants before the research began. Each participant received clear information about the study's purpose, procedures, possible risks, and expected benefits in language that was easy to understand. Participants were also told that their personal information would be kept confidential and that their responses would be used only for research purposes. They were informed of their right to refuse participation or withdraw at any time without penalty or negative consequences.

DATA AVAILABILITY STATEMENT

The data used in this study are not publicly available because protecting participant privacy and maintaining confidentiality are essential ethical responsibilities. Restricting public access also helps ensure compliance with applicable research ethics standards and data protection requirements. However, the data may be made available to qualified researchers upon reasonable request for purposes such as verification or further analysis. Any request will be reviewed carefully on a case by case basis. Data sharing will only be considered when the proposed use is ethically appropriate and consistent with participant consent. Approval from the relevant institutional ethics review board is required before any data can be shared.

REFERENCES

Aisyah, S. F., Az-Zahra, N., & Azzahro, N. (2025). Monoftongisasi diftong pada dialek Kansai berbahasa Jepang (kajian fonologi). *Caruban Proceeding*, 73–82. https://carubanproceeding.ipbcirebon.ac.id/index.php/carubanproceeding/article/view/38

Azzahro, N. (2024). Monoftongisasi diftong dalam fonologi bahasa Jepang: Analisis pola bunyi konteks penggunaan sehari-hari. *Jurnal Ilmiah Bina Bahasa, 17*(1), 12–22. https://doi.org/10.33557/binabahasa.v17i1.3054

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3). https://www.praat.org

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press. https://doi.org/10.7551/mitpress/8979.001.0001

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in vowel formants of American English. *The Journal of the Acoustical Society of America, 126*(5), 2603–2618. https://doi.org/10.1121/1.3224703

Hayes, B. (2011). *Introductory phonology*. Wiley-Blackwell. https://doi.org/10.1002/9781444318159

Holmes, J., & Wilson, N. (2017). *An introduction to sociolinguistics* (5th ed.). Routledge. https://doi.org/10.4324/9781315728431

Irwin, M., & Sohn, H. (2020). *The Oxford handbook of Japanese linguistics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195307347.001.0001

Johnson, K. (2012). *Acoustic and auditory phonetics* (3rd ed.). Wiley-Blackwell. https://doi.org/10.1002/9781118592886

Kubozono, H. (1999). Mora and syllable. *Phonology, 16*(1), 33–63. https://doi.org/10.1017/S0952675799003710

Kubozono, H. (Ed.). (2015). *Handbook of Japanese phonetics and phonology*. De Gruyter Mouton. https://doi.org/10.1515/9781614511984

Labov, W. (2001). *Principles of linguistic change: Social factors*. Wiley-Blackwell. https://doi.org/10.1002/9780470756598

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Cengage Learning.

Liebe-Harkort, M. (2019). English loanwords in Japanese: Phonological adaptation patterns. *SKASE Journal of Theoretical Linguistics, 16*(1), 1–17. http://www.skase.sk/Volumes/JTL40/pdf_doc/01.pdf

Mahboob, A., & Cruz, P. (2013). English and the Asian language contact ecology. *Interchange, 44*, 83–91. https://doi.org/10.1007/s10780-013-9191-x

Mawarni, N., Syarfina, T., Zulhantiar, P. A., & Rangkuti, R. (2024). Analysis of Javanese language prosody (acoustic phonetic study). *Fonologi: Jurnal Ilmuan Bahasa dan Sastra Inggris, 2*(4), 63–75. https://doi.org/10.61132/fonologi.v2i4.1136

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801624

Trudgill, P. (2008). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199534213.001.0001

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806827

Virdaus, N. (2023). *Analisis akustik bunyi konsonan hambat /d/ pada pemelajar bahasa Jepang penutur asli bahasa Jawa*.

Wardhaugh, R., & Fuller, J. M. (2021). *An introduction to sociolinguistics* (8th ed.). Wiley-Blackwell. https://doi.org/10.1002/9781119733721

Yoneyama, K. (2015). The role of perception in Japanese phonology. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 700–726). De Gruyter Mouton. https://doi.org/10.1515/9781614511984-023