Emotional Expression and Phonetic Features in Gidayū Narratives in *Hidakagawa Iriaizakura*

Shingo HAYANO[1]

Abstract

Gidayū-bushi, one of Japan's most emotionally charged narrative vocal arts, employs three distinct phonatory modes—*fushi* (melodic recitation), *ji* (narrative recitation), and *umiji* (non-harmonic emotional phonation). While previous scholarship has richly described musical form, performer lineage, and narrative aesthetics, little attention has been paid to the acoustic mechanisms through which emotional expression is produced. This study provides the first systematic phonetic analysis comparing the three vocal modes, based on an unaccompanied recording of *Hidakagawa Iriaizakura* performed by Takemoto Koshikō (2022). Fundamental frequency (*F0*) extraction, spectrographic analysis, and pitch variability measures were used to examine harmonicity, periodicity, and vocal stability across modes. Results reveal a three-layer system. *Fushi* exhibits high pitch and stable harmonic structure, while *ji* displays moderate variability with speech-like contours. In contrast, *umiji* is characterized by extreme *F0* excursions (80–500 Hz), non-periodic waveforms, glottal constriction, and spectral noise—indicating a controlled disruption of vocal stability rather than ornamentation. Statistical testing (*Kruskal–Wallis* and post-hoc *Dunn* tests) confirmed significant differences among modes. These findings demonstrate that emotional expression in Gidayū is grounded in measurable acoustic behavior, particularly the deliberate use of non-periodicity and instability in *umiji*. The study contributes an empirical framework bridging ethnomusicology, voice science, and affective performance studies.

Keywords:
Fundamental frequency (F0); harmonicity; non-harmonic phonation; traditional vocal arts; acoustic phonetics; *umiji*; pitch variability

[1] Tsuru University Faculty of Literature Professor

1. Introduction

Gidayū-bushi is one of the most acoustically sophisticated narrative performing arts in Japan. A chanter (Gidayū) constructs the emotional contour of the story primarily by alternating between fushi (melodic narration) and ji (narrative speech), while in highly affective scenes a third mode—umiji—emerges. Umiji has long been regarded as a core expressive technique in Gidayū performance, yet the acoustic properties of these three vocal modes have rarely been compared quantitatively. *Hidakagawa Iriaizakura* "Ferry Crossing" scene is among the most emotionally charged passages in the repertory, characterized by rapid alternation between vocal modes. This makes it an ideal site for phonetic analysis. The present study examines an unaccompanied recording by Takemoto Koshikō (member of the Gidayū Association). Previous scholarship on Gidayū and related Japanese narrative musics has been extensive but concentrated largely in musicology and performance studies. Tokita (2000) identifies formulaic melodic patterns shared across heikyoku, Gidayū-bushi, and Kiyomoto-bushi, demonstrating the patterned nature of Japanese narrative music. Tokita and Wasson (2015) provide a millennium-scale overview of Japanese singers of tales, delineating historical trajectories, stylistic variation, and genre-specific features. Within Japanese scholarship, Watanabe (2000) articulates the aesthetic principles of Gidayū-bushi, while Yasuda (2012) offers a detailed analysis of jōruri structure.

However, these studies overwhelmingly privilege melodic structure, narrative organization, and performance aesthetics, with little attention to the physical and acoustic features of the voice—including $F0$ behavior, harmonic structure, or stability/instability of vocal fold vibration. In particular, umiji—the hallmark of highly emotional Gidayū narration—has typically been described impressionistically rather than examined through acoustic measurement. Western voice science offers relevant but insufficient tools. Sundberg (1987) and Titze (2000) present detailed models of periodic vocal fold vibration and harmonic structure in classical singing, but these frameworks are premised on harmonic, periodic, and relatively stable phonation. As such, they do not directly account for the non-harmonic, non-periodic, and sharply fluctuating characteristics of umiji. Given this gap, the present study is positioned to contribute in two major ways:

1. To distinguish the three vocal modes of Gidayū—fushi, ji, and umiji—using phonetic indices such as $F0$, harmonic structure, and periodicity.
2. To quantify the acoustic nature of umiji, showing that its emotional expression derives from physical features—non-harmonicity, aperiodicity, abrupt shifts in $F0$, and glottal constriction—rather than from stylistic ornamentation alone.

The goal is to integrate insights from musicology and voice science to visually and acoustically model the vocal system of Gidayū narration. This approach not only refines our understanding of a major Japanese narrative genre but also contributes to broader discussions of emotional expression in voice, as well as analytical frameworks for Japanese vocal culture.

2. Literature Review: Narrative Music, Voice Studies, and Gidayū-Bushi

Studies of Japanese narrative music have primarily developed within musicology and ethnomusicology. Tokita (2000) offers one of the most comprehensive comparative analyses of heikyoku, Gidayū-bushi, and kiyomoto-bushi, demonstrating that these genres share a system of "formulaic musical material" that structures melodic contour and narrative flow. Rather than focusing on voice production itself, Tokita's work situates Gidayū-bushi within a broader typology of Japanese narrative song, emphasizing the patterned organization of melodic units across genres. Further contributions are found in *Japanese Singers of Tales* (Tokita & Wasson, 2015), which positions Gidayū-bushi within a thousand-year continuum of Japanese narrative performance. This volume highlights performer lineage, stylistic differentiation, and the socio-historical forces that shaped narrative vocal traditions. While highly valuable for contextualizing Gidayū-bushi, these studies primarily analyze form, repertoire, and performance practice—not the acoustic properties of the voice.

2.1. Studies on Jōruri, Gidayū Aesthetics, and Structural Analysis

Japanese-language scholarship provides rich discussions of the aesthetics and structural principles of jōruri narrative. Watanabe (2000) articulates the aesthetic foundations of Gidayū-bushi, emphasizing how fushi (melodic recitation), kotoba (text), and narrative semantics interact to create dramatic effect. Yasuda (2012) further analyzes the structural composition of jōruri, outlining the narrative functions, scene structures, and performative conventions that govern the genre. These studies are indispensable for understanding the formal and dramaturgical foundations of Gidayū-bushi. However, they do not address the acoustic or phonetic aspects of performance—particularly the mechanisms by which emotion is instantiated in vocal sound such as umiji, nor the transitions between fushi, ji, and other vocal modes.

2.2. Voice Science and Singing Studies Relevant to Gidayū-Bushi

Recent developments in digital performance analysis have begun to examine traditional Japanese vocal and narrative arts using empirical methods. Hayano and Don (2021), for

example, demonstrated that emotional intensity in ningyō jōruri can be visualized through motion capture analysis of performer movement, suggesting that embodied emotion follows measurable physical patterns rather than relying solely on aesthetic interpretation. While their study does not focus directly on the voice, it represents a methodological shift toward data-driven analysis of affective performance mechanisms—an approach that informs the present study's acoustic–phonetic examination of umiji, fushi, and ji in Gidayū-bushi. Research in voice science provides important theoretical resources for understanding vocal production. Sundberg's *The Science of the Singing Voice* (1987) and Titze's *Principles of Voice Production* (2000) outline the mechanisms of harmonic phonation, vocal fold vibration, and resonance. These works establish the physiological and acoustic bases of singing, including periodicity, harmonic structure, and vibrato (typically 5–7 Hz regular modulation). Nevertheless, these theories were constructed primarily for Western singing traditions, which rely on periodic, harmonic phonation with stable vibrato and controlled resonance shaping. Consequently, they do not account for the non-harmonic, irregular, and sometimes physiologically extreme vocal phenomena observed in Gidayū-bushi—particularly umiji, which exhibits sudden F_0 spikes, glottal constriction, and non-periodic waveforms.

2.3. Research Gap and the Need for Acoustic–Phonetic Analysis

Across both international and Japanese scholarship, a clear gap becomes apparent:

- Melodic and structural analyses are abundant.
- Acoustic analyses of voice production in Gidayū-bushi are virtually absent.
- No prior study quantitatively compares fushi, ji, and umiji through phonetic measures such as F_0, harmonicity, or periodicity.

Thus, there remains no systematic account of how vocal modes are acoustically differentiated or how emotional expression is encoded in the voice. This study addresses this gap by integrating phonetic analysis with narrative-music scholarship, offering the first data-driven model of Gidayū vocal production.

3. The Acoustic Structure of Fushi, Ji, and Umiji in Gidayū Narrative Performance

This chapter analyzes the three principal vocal modes of Gidayū-bushi—fushi (melodic recitation), ji (narrative recitation), and umiji (non-harmonic emotional phonation)—based on the unaccompanied recording of *Hidakagawa Iriaizakura*. By conducting F_0 extraction, harmonicity inspection, and spectrographic analysis, the chapter clarifies how these modes differ acoustically and how they interact to shape the emotional and narrative structure of the performance.

3.1 Data and Recording Context

The analysis in this study is based on an unaccompanied performance of *Hidakagawa Iriaizakura* ("Ferry Crossing" scene), one of the most emotionally charged and frequently studied segments of the Gidayū repertoire. The recording was produced by Takemoto Koshikō, a certified member of the Gidayū Association, at the Nishikawa Koryū rehearsal studio on December 18, 2022. The performance style reflects contemporary professional Gidayū practice and was delivered without shamisen accompaniment in order to isolate the vocal signal and prevent harmonic masking effects commonly introduced by instrumental resonance. The recording was captured in a controlled rehearsal environment with minimal room reverberation. Audio was recorded at a 44.1 kHz sampling rate with 16-bit resolution using a cardioid condenser microphone positioned approximately 30–40 cm from the performer's mouth to balance clarity with natural dynamic behavior during expressive vocal projection. No compression, equalization, or post-processing was applied beyond normalization to a consistent amplitude level for analysis. The absence of instrumental accompaniment and minimal acoustic interference allow the phonatory characteristics of fushi, ji, and umiji to be observed with precision, particularly in relation to harmonic structure, periodicity, and extreme pitch excursions.

3.2 Acoustic Analysis Procedure

Acoustic analysis was conducted using Praat (Version 6.1). To accommodate the wide phonatory range characteristic of Gidayū vocal performance—particularly the extreme pitch excursions observed in umiji—the pitch tracking parameters were set to a floor of 75 Hz and a ceiling of 600 Hz. This elevated ceiling was necessary because conventional settings used for conversational speech or classical singing (typically 300–400 Hz) failed to capture the highest register produced during emotional intensification. Waveform displays and wide-band spectrograms (Hamming window, 0.005 s window length) were examined to assess harmonicity, periodicity, and the presence of noise components. Extracted $F0$ values were segmented according to the three vocal modes (fushi, ji, and umiji) and submitted to descriptive statistical analysis, including mean $F0$, standard deviation, and pitch range (minimum–maximum values). To statistically evaluate whether the three vocal modes represent acoustically distinct categories rather than continuous stylistic variation, a Kruskal–Wallis test was applied to the $F0$ distributions. Where significant differences were observed, Dunn's post-hoc comparisons with Bonferroni correction were conducted. In addition, kernel density estimation (KDE) was performed to visualize the distributional shape and spread of $F0$ patterns across modes.

This procedure allowed for the comparison of harmonic stability in fushi, moderate variability in ji, and the extreme, non-linear pitch fluctuations characteristic of umiji.

3.3 Results: Mode-Specific Acoustic Profiles
3.3.1 Fushi: The Melodic Mode
Fushi appears in passages where emotional intensity rises and the performer moves into a melodic, quasi-singing style. Acoustically, fushi is characterized by elevated pitch ranges (approximately 300–450 Hz) and smooth, stable $F0$ contours. In the segment "hime wa aru ni mo arareba koso," the mean $F0$ is approximately 320 Hz with a relatively small standard deviation of 35 Hz, demonstrating a stable harmonic structure (Figure 1). Although fushi does not employ Western-style vibrato (5–7 Hz periodic modulation), the voice retains harmonicity and produces a controlled melodic contour that contributes to the emotional uplift of the narrative.
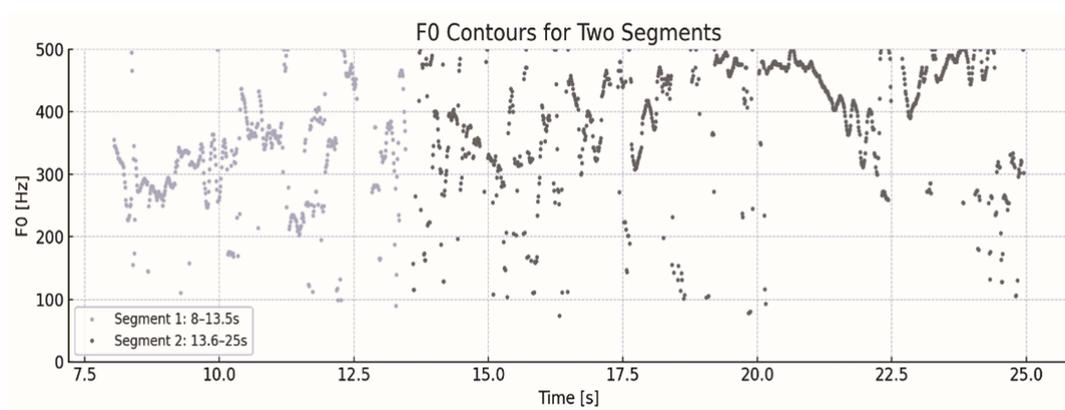

Figure 1.
（姫はあるにもあらばこそ　ええ聞ませぬ聞こえませぬ安珍さま）

3.3.2 Ji: The Narrative Mode
Ji serves as the primary narrative layer, delivering verbal content with less melodic pressure. Pitch stabilizes in the mid–upper range (approximately 250–300 Hz), following natural speech-like modulation. In the segment "kyō totemo chichiue no goiken…," the mean $F0$ is about 288 Hz, with wider variation than in fushi but without the aperiodicity observed in umiji (Figure 2). Ji maintains narrative continuity and functions as the structural backbone between fushi and emotionally charged passages.
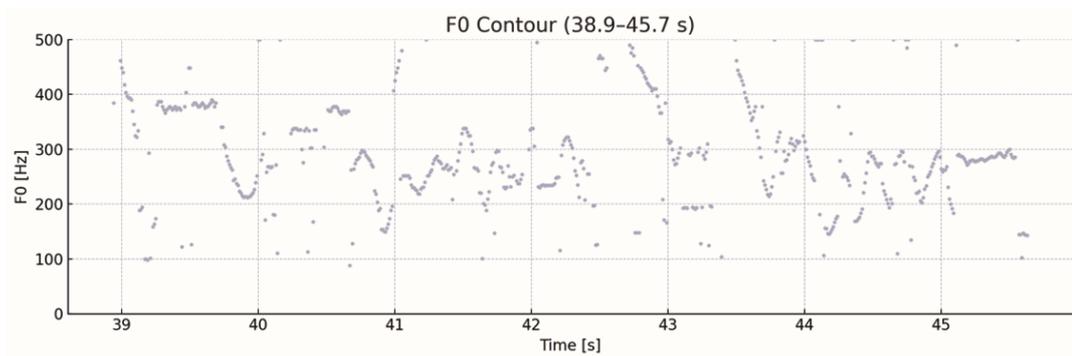
Figure 2.

（今日とても父上の御意見ごもつともとは思へども）

3.3.3 Umiji: Non-harmonic Emotional Phonation

Umiji is the most distinctive vocal mode in Gidayū-bushi, marked by non-harmonicity, irregular $F0$ fluctuation, glottal constriction, and rapid $F0$ shifts. In the segments "anchin-samaa…" (Figure 1) and "rinne wa aa…" (Figure 3), maximum $F0$ values reach 490–510 Hz while minima plunge to 80–90 Hz, yielding extremely large standard deviations (~90 Hz). Spectrograms reveal a breakdown of harmonic structure, unlike the orderly vertical striations seen in fushi. Umiji thus differs fundamentally from Western vibrato: rather than periodic modulation, it exhibits non-linear spikes and drops, reflecting intense emotional tension.
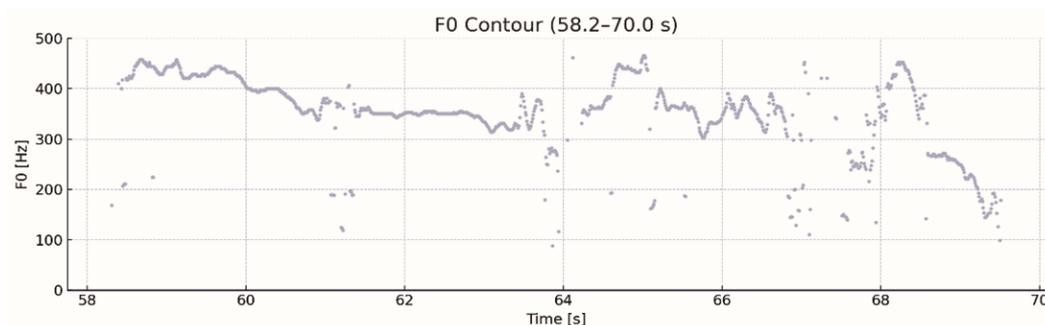

Figure 3.

（可愛いという輪廻は離れず）

3.3.4 The Three-Layer Structure of the Three Modes

The acoustic data justify a three-layer model (Figure 4). At the base level, ji functions as the narrative mode, characterized by mid–high pitch and moderate fluctuation. Fushi occupies an intermediate layer, marked by melodic shaping, harmonic stability, and elevated pitch with a relatively stable contour. Umiji forms an emergent upper layer,

defined by non-harmonic, non-periodic phonation accompanied by extreme pitch fluctuations, including abrupt spikes and drops. These modes alternate in response to narrative and emotional demands, collectively constituting the multilayered expressive system of Gidayū performance.
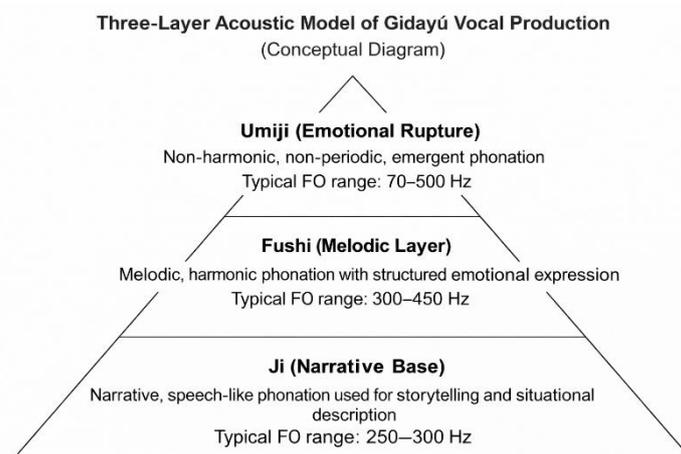

Three-Layer Acoustic Model of Gidayú Vocal Production
(Conceptual Diagram)

Umiji (Emotional Rupture)
Non-harmonic, non-periodic, emergent phonation
Typical F0 range: 70–500 Hz

Fushi (Melodic Layer)
Melodic, harmonic phonation with structured emotional expression
Typical F0 range: 300–450 Hz

Ji (Narrative Base)
Narrative, speech-like phonation used for storytelling and situational description
Typical F0 range: 250—300 Hz

Figure 4.

4. Emotional–Acoustic Interaction: The Phonetic Basis of Umiji

Gidayū narration reaches its most distinctive sonic form in umiji, a vocal mode that differs fundamentally from both melodic fushi and narrative ji. While fushi builds a relatively stable, harmonic melodic line and ji maintains a controlled narrative contour, umiji is produced under conditions of strong expiratory pressure, irregular vocal fold vibration, and abrupt $F0$ excursions. In this sense, umiji is not merely a stylistic color but an acoustic manifestation of emotional overload.

4.1 Non-harmonic Voice Quality and the Limits of Vibrato-based Models

The clearest hallmark of umiji is its non-harmonic voice quality. In typical singing, the spectrum exhibits a clear series of integer-multiple harmonics above the fundamental frequency ($F0$), resulting in vertically aligned striations in the spectrogram. By contrast, in umiji segments such as "anchin-samaa…" and "rinne wa aa…" the harmonic series breaks down; the vertical harmonic stripes become blurred, and noise components intrude into the formant regions. This indicates that vocal fold vibration has become aperiodic and partially chaotic. These observations place umiji outside the explanatory range of conventional Western voice science, which has mainly modeled periodic phonation and regular vibrato. Classical vibrato is typically characterized as a quasi-sinusoidal modulation of $F0$ at around 5–7 Hz with relatively constant depth. In umiji, however, $F0$

does not oscillate periodically. Instead, it shows irregular spikes and drops: the contour contains sudden upward and downward movements, with local segments in which periodicity is effectively lost. Umiji is therefore better described as a form of non-harmonic, non-periodic phonation at the physiological limits of vocal control rather than as a variant of vibrato. In the "rinne wa aa…" segment (see Figure 3), *F0* repeatedly surges toward 500 Hz and then collapses toward the 80–90 Hz range within very short time spans. Such spike–drop patterns cannot be captured by models that assume regular modulation around a central *F0*. They reflect a state in which the chanter deliberately drives the vocal system close to breakdown, using the instability itself as an expressive resource.

4.2 Large-amplitude F0 Excursions and Emotional Climax

Across several umiji segments in the corpus (e.g., Figure 1 "anchin-samaa…," Figure 3 "kawaii to iu rinne wa aa…"), *F0* excursions reach amplitudes of approximately ±40–90 Hz. Mean *F0* tends to be higher than in fushi or ji, but more crucial is the magnitude of instantaneous change: maximum values approach 490–510 Hz, while minima can drop suddenly to 70–90 Hz. In "kawaii to iu rinne wa aa…" (Figure 3), a recurring pattern is observed in which the voice is prolonged in a high-tension state—maintaining a high *F0*—and then abruptly falls. Aerodynamically, this reflects a shift from extreme subglottal pressure with stretched vocal folds to rapid glottal constriction and partial cessation of vibration. Psychophysiologically, such behavior has been associated with anger, grief, and intense longing—high-arousal affective states in which the vocal system is driven beyond its ordinary operating range. These extreme excursions are therefore not incidental by-products of heightened emotion but intentional manipulations of the phonatory system. Seen in this light, the large-amplitude *F0* movements function as acoustic markers of emotional climax. They are not ornamental modifications added to a melodic framework. Rather, they constitute a transformation of the phonatory regime itself, in which physical instability becomes a semiotic carrier of emotional meaning. The instability—the spike–drop pattern, the sudden collapse of regular periodicity, and the extreme fluctuation in *F0*—operates as an embodied sign of emotional rupture, dramatically heightening the narrative moment.

4.3 Harmonic Long Tones and Local Emergence of Umiji (Role of Figure 5)

Importantly, umiji does not occur continuously throughout an emotionally loaded passage. Even within highly charged lines, harmonic and non-harmonic segments alternate. This

can be clearly seen in the segment "itoshii yukashii koishii to yumutsutsu ni mo wasurekane" (Figure 5), whose contour is presented in Figure 5. At the beginning of this phrase, the word "itoshii" is sustained as a relatively stable long tone with a well-organized harmonic structure. The spectrogram shows clear vertical harmonic stripes, and the *F0* contour, while not perfectly flat, remains within a modest range of fluctuation. This example demonstrates that a long vowel is not necessarily non-harmonic; there are also "calm" harmonic long tones in Gidayū narration. Towards the latter half of the same segment, however, especially on "koishii," local regions of non-harmonicity and irregular *F0* motion begin to appear. The voice temporarily departs from the harmonic pattern, introducing fragments of umiji-like instability. Figure 5 thus plays a crucial role in clarifying that umiji is not identical with "sustained sound" in general. Rather, umiji emerges locally within phrases as an intrusive, non-harmonic episode inserted into an otherwise harmonic melodic or narrative context. By contrasting Figure 5 with Figure 3, we can therefore distinguish three situations: (i) stable harmonic long tones (fushi-like), (ii) fully non-harmonic umiji with continuous irregularity, and (iii) mixed cases in which harmonic and non-harmonic segments alternate within a single phrase.
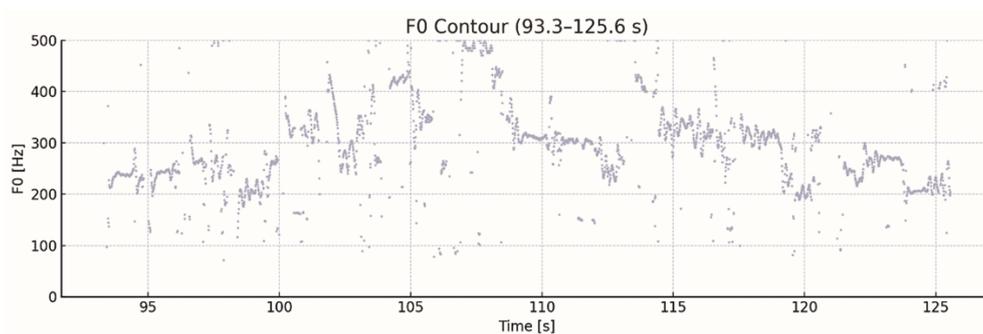

Figure 5.
（愛しい床しい恋しいと夢現にも忘れかね）

4.4 Temporal Organization of Emotion: Accumulation and Rupture

Umiji also has a characteristic temporal organization. It is not simply "screaming at high pitch," but a phenomenon that tends to appear at the moment when melodic and narrative equilibrium can no longer be maintained. In other words, umiji marks the rupture point where the accumulated emotional tension surpasses the capacity of regular fushi or ji. This process is clearly illustrated in the passage "omoikirarenu anchin-sama kikoenu waina to mi o modae 'watsu' to bakari ni koe o age…," whose *F0* contour is shown in Figure 6. The first part of the phrase is delivered in a relatively controlled narrative voice: *F0* fluctuates within a mid–high range, and the harmonic structure remains largely intact.

As the line progresses, however, tension builds up; at the exclamation "watsu," *F0* suddenly jumps from around 180 Hz to nearly 420 Hz and then drops into the 70 Hz range. This abrupt spike–drop pattern corresponds to a momentary collapse of phonatory stability—the vocal equivalent of a physical convulsion. From the listener's perspective, this sequence of gradual accumulation followed by instantaneous rupture creates a strong sense of emotional impact. It is precisely this time course, rather than pitch height alone, that allows umiji to convey the extremity of the character's state.
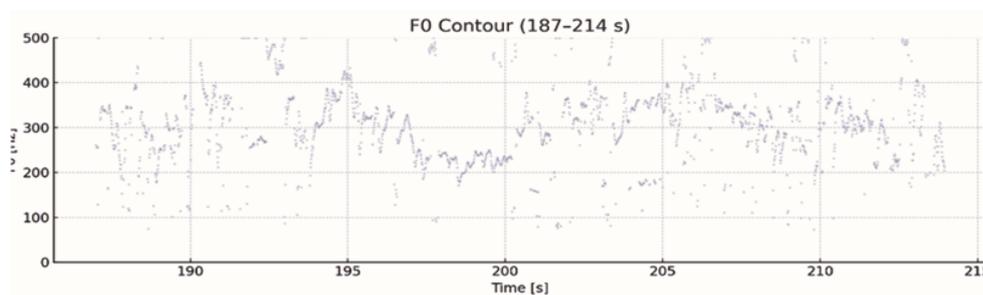

Figure 6.

（思い切られぬ安珍さま聞こえぬわいなと身をもだえ「わつ」とばかりに声をあげ嘆く涙の雨車軸）

4.5 Functional Role of Umiji within the Three-layer System

The analyses above allow us to re-situate umiji within the three-layer model of Gidayū vocal production. Fushi provides the melodic framework, shaping the overall emotional contour through relatively stable, harmonic pitch movement. Ji maintains the narrative skeleton, advancing the storyline with controlled, speech-like intonation. Umiji, in contrast, is a third, acoustically distinct mode that appears at structurally and emotionally critical points. By exploiting non-harmonicity, aperiodicity, and large *F0* excursions, umiji converts the imbalance of the vocal system itself into a semiotic resource. In doing so, it sonically marks turning points in the narrative: a sudden outburst of resentment, an intensification of longing, or an irreversible decision. The combined use of fushi, ji, and umiji thus allows Gidayū narration to construct a finely layered interaction between story, emotion, and voice—an interaction that cannot be captured by melodic analysis alone.

5. Statistical Validation of the Three-Layer Acoustic Model

The descriptive statistics summarized in Table 1 demonstrate clear differences among the three vocal modes—fushi (melodic recitation), ji (narrative speech), and umiji (non-harmonic emotional voicing)—particularly in mean *F0*, maximum range, and variance. To examine whether these acoustic differences form statistically separable categories, a

kernel density estimation (Figure 7) was performed using all valid extracted *F0* values from each mode.

Table 1. Acoustic Features of the Three Vocal Modes in Gidayū-bushi
(Based on Analysis of Hidakagawa Recording)

| Segment (Example)	Vocal Mode	Mean F0	Max F0	Min F0	SD (Pitch Variability)	Acoustic Characteristics
hime wa aru ni mo arareba koso	fushi	~320 Hz	~410 Hz	~210 Hz	~35 Hz	Stable harmonic structure; smooth melodic contour; controlled pitch movement
kyō totemo chichiue no goiken	ji	~288 Hz	~360 Hz	~190 Hz	~88 Hz	Moderately variable F0; speech-like phrasing; controlled but less melodic
anchiii-samaa	umiji	~295 Hz	~500 Hz	~90 Hz	~95 Hz	Non-harmonic spectrum; abrupt pitch rise and fall; glottal constriction observed
kawaii to iu rinne wa aa...	umiji	~300 Hz	~510 Hz	~80 Hz	~90 Hz	Large amplitude non-periodic fluctuations; irregular vocal fold vibration
omoikirarenu anchina-sama	umiji	~305 Hz	~490 Hz	~85 Hz	~92 Hz	Alternation of strained high pitch and collapse-like pitch fall; emotional intensity maximized
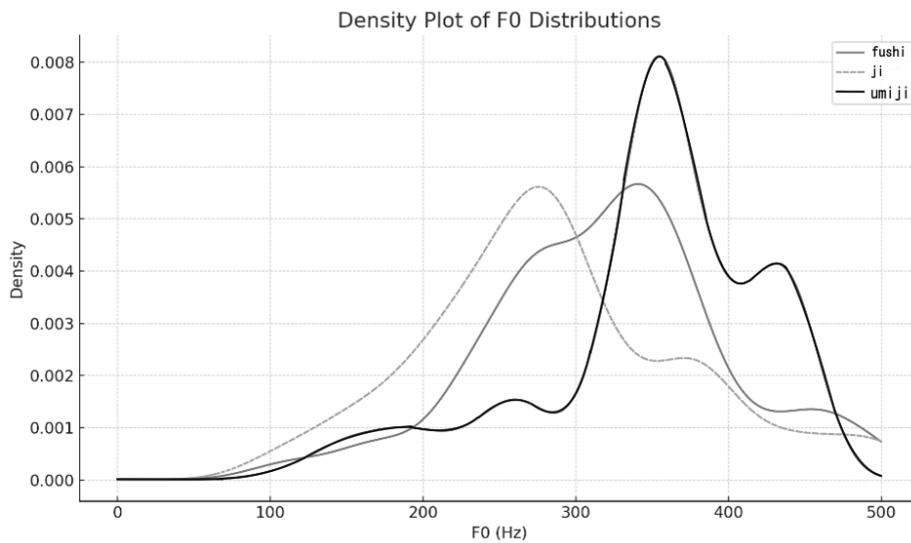
Figure7

The distribution plot reveals that:

- fushi forms a narrow, unimodal peak in the upper-mid range (≈300–360 Hz).
- ji occupies a lower mid-range field with a broader but still coherent distribution.
- umiji displays a multi-peaked structure with extreme spread (≈80 Hz–500 Hz), reflecting non-harmonic instability and abrupt pitch excursions.

A Kruskal–Wallis test confirmed significant differences among the three distributions ($H = 228.10$, $p < 2.93 \times 10^{-50}$). A post-hoc Dunn test further revealed that all pairwise contrasts (fushi vs. ji, ji vs. umiji, fushi vs. umiji) were significant after Bonferroni correction ($p < .001$). Effect size using epsilon-squared ($\varepsilon^2 = 0.094$) indicates a large practical divergence among modes, supporting the interpretation that these differences represent distinct phonetic systems rather than gradual stylistic variation.

6. Future Directions

Although this study provides the first acoustic–phonetic evidence distinguishing fushi, ji, and umiji as three systematically differentiated vocal modes in Gidayū narrative performance, several research questions remain open. The present analysis was based on a single performer and a single performance instance, and therefore future work should expand to multiple performers, schools, and performance contexts in order to determine whether the observed three-layer structure represents an individual style, a genre-wide pattern, or a lineage-dependent phenomenon. Further acoustic measures such as jitter, shimmer, spectral tilt, and harmonic-to-noise ratio would enable a more precise characterization of the mechanisms underlying non-harmonic and non-periodic production in umiji, particularly in relation to vocal fold behavior. Listener perception studies could also examine how audiences cognitively distinguish the three modes, and whether the extreme acoustic instability of umiji universally correlates with heightened emotional interpretation. In addition, applying the present findings to machine learning models may contribute to the development of synthetic or assistive Gidayū voices for preservation, pedagogy, and cultural transmission. Ultimately, the results of this study underscore the need for an interdisciplinary framework in which narrative musicology, voice science, and computational modeling jointly advance the understanding of Gidayū as both an artistic and phonetic system.

Conflict of Interest

The author declares no conflict of interest.

References

Hayano, S., & Don, R. (2021). *A study of emotional expression through the movement of circles in Ningyō Jōruri*: *An analysis using motion capture.* The Tsuru University Review, 94,1－10. (In Japanese)

Sundberg, J. (1987). *The science of the singing voice*. DeKalb, IL: Northern Illinois University Press.

Titze, I. R. (2000). *Principles of voice production* (2nd ed.). National Center for Voice and Speech.

Tokita, A. (2000). *The nature of patterning in Japanese narrative music: Formulaic musical material in Heikyoku, Gidayū-bushi, and Kiyomoto-bushi*. Musicology Australia, 23(1), 99–122. https://doi.org/10.1080/08145857.2000.10415643

Tokita, A., & Wasson, D. (Eds.). (2015). *Japanese singers of tales: Ten centuries of performing narrative*. Leiden, The Netherlands: Brill.

Watanabe, T. (2000). *Gidayū-bushi no bigaku* [The aesthetics of Gidayū-bushi]. Tokyo: Iwanami Shoten. (In Japanese)

Yasuda, B. (2012). *Jōruri no kōzō* [The structural foundations of jōruri narrative]. Tokyo: Perikansha. (In Japanese)