LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR PHONETIK UND SPRACHVERARBEITUNG

An Analysis of Tune-Text Alignment in Two Varieties of Japanese

- Neutralization and Tonal Restructuring in Tokyo and Kagoshima Japanese Citation Forms -

Master's Thesis

Supervisor: Prof. Dr. Jonathan Harrington

Submission Date: 05.08.2025

Author: Katharina Neubert, 11961334

Address: Sommerstr. 35, 81543 Munich

k.neubert@campus.lmu.de

Phonetics and Speech Processing M.A.

SS 2025

Contents

Abstract

This thesis examines the alignment of tonal contours in Tokyo Japanese (TJ) and Kagoshima Japanese (KJ), focusing on how contact with TJ is influencing generational changes in KJ. Specifically, the study investigates tonal realization in isolated monomoraic and bimoraic monosyllables across three speaker groups: older KJ, younger KJ, and TJ speakers. Using acoustic data and Generalized Additive Mixed Models (GAMMs), the analysis tests whether tonal neutralization and phonetic convergence are occurring in the speech of younger KJ speakers.

Results show that both monomoraic and bimoraic words undergo neutralization in innovative KJ speech but in distinct ways. Monomoraic words show both phonological and phonetic convergence with TJ, surfacing with flat f_0 contours that closely resemble TJ unaccented patterns. Bimoraic words, by contrast, are reanalyzed as Type A (falling), but the pitch falls produced by younger KJ speakers differ phonetically from those of both older KJ and TJ speakers. These falls are typically shallower, earlier than KJ-old, and more variable in shape, pointing to a gradient and ongoing reorganization of phonetic implementation rather than abrupt phonological restructuring.

This study contributes to the understanding of tonal realignment in Japanese dialects by showing that younger KJ speakers are not simply adopting TJ forms wholesale but are instead reshaping their tonal system through incremental phonetic shifts shaped by contact and reinterpretation.

1. Introduction

The study of pitch accent systems offers valuable insight into how prosodic structures evolve under the influence of dialect contact, generational change, and broader processes of sound change. Kagoshima Japanese (KJ), a dialect spoken in the southernmost part of Kyushu, exhibits a typologically rare two-pattern syllable-based, word tone system that contrasts sharply with the multi-pattern, mora-based system of Tokyo Japanese (TJ) (Ishihara, 2004; Kubozono, 2007). Recent research suggests that increasing exposure to TJ through media, education, and mobility, may be influencing the prosodic system of KJ, particularly among younger speakers (Kubozono, 2018a, 2018b; Ota, 2020). This thesis investigates the nature of that change: Is it phonological, involving categorical restructuring of tonal categories, or phonetic, driven by gradual shifts in alignment and contour shape? How do such changes reflect broader mechanisms of contact-induced language change?

To answer these questions, this thesis examines the tonal behavior of monomoraic and bimoraic words in isolation across three speaker groups: older KJ speakers, younger KJ speakers, and Tokyo Japanese speakers. Specifically, the study tests three hypotheses: (I) that younger KJ speakers neutralize the pitch accent contrast in monomoraic words, mirroring TJ's unaccented pattern; (II) that bimoraic words are consistently realized with a pitch fall in isolation, reflecting a shift toward a uniform Type A contour; and (III) that tonal category shifts between KJ and TJ reflect not abrupt phonological replacement but phonetic drift, as predicted by listener-based models of sound change (e.g., Beddor, 2009, 2023).

Methodologically, the study employs Generalized Additive Mixed Models (GAMMs) to analyze f_0 contours, a technique well-suited for modeling the non-linear and speaker-specific nature of pitch movement over time. The results reveal both phonological restructuring and phonetic variability, providing evidence of an ongoing, gradient sound change shaped by both internal system constraints and external sociolinguistic pressures.

This thesis is organized into the following chapters: Chapter 2 provides the theoretical background, beginning with an overview of Autosegmental-Metrical (AM) theory and tonal alignment (section 2.1), followed by a detailed discussion of prosodic structure in Japanese dialects (section 2.2), with a focus on mora vs. syllable as the tone-bearing unit and the contrast between Tokyo and Kagoshima Japanese. The chapter concludes with a review of sound change as phonetic drift (section 2.3), including well-known models of sound change, and previous evidence of tonal change in KJ. Chapter 3 outlines the methodology used in the study, including speaker groups, materials, recording procedures, and the acoustic and statistical methods used to analyze f_0 contours. Chapter 4 presents the results in two parts: first, the behavior of monomoraic words (section 4.1), and second, the analysis of bimoraic words across speaker groups and evidence for gradient phonetic change (section 4.2). Finally, Chapter 5 discusses these findings in light of the sound change models introduced in the theoretical background, arguing that the prosodic changes observed in younger KJ speakers reflect both phonological restructuring and ongoing phonetic drift.

2. Theoretical Background

This chapter lays the theoretical foundation for the analysis of tonal change in Kagoshima Japanese (KJ). While the study is primarily based on phonetic analysis, it draws on key concepts from Autosegmental-Metrical (AM) theory to describe and categorize tonal contours. Section 2.1 introduces the AM framework, focusing on its treatment of tonal alignment and hierarchical prosodic structure, which proves useful for modeling pitch accent patterns in Japanese. Section 2.2 turns to the prosodic typology of Japanese dialects, examining how units such as morae and syllables function as tone-bearing domains across varieties. The contrast between Tokyo Japanese (TJ) and KJ serves as a foundation for understanding how tonal distinctions are structured, and potentially restructured, in contact settings. Finally, Section 2.3 addresses the mechanisms of sound change, emphasizing how gradient phonetic variation may give rise to categorical reanalysis. This section introduces theoretical models that help explain how subtle phonetic influences can gradually reshape the prosodic grammar of a dialect. Together, these components set the stage for the empirical analysis that follows, which investigates tonal realization in mono- and bimoraic forms among younger KJ speakers.

2.1. Autosegmental-Metrical (AM) Theory and Tonal Alignment

The study of intonation has historically been marked by deep theoretical divisions and unresolved questions, hindering the development of a unified framework for describing intonational phenomena. For many years, scholars disagreed not only on how intonation should be analyzed, but also on what aspects of it were most significant. However, since the mid-1970s, a convergence of research traditions has given rise to a more cohesive set of assumptions and methodologies. Among these, the Autosegmental-Metrical (AM) framework has emerged as a leading approach. This framework offers a systematic and cross-linguistically comparable phonological model in which pitch contours are represented as sequences of discrete tonal events, namely pitch accents and boundary tones (Ladd, 1996; 2008), associated with specific points in the segmental string.

The AM theory is grounded in the autosegmental insight that tonal and segmental representations operate on separate tiers (Ladd, 1996, 2008), and in the metrical principle that the units to which tones associate, such as syllables, morae, or feet, are language-specific and structurally constrained (e.g. Duanmu, 1992; Roseano et al., 2015). Within this framework, tones are drawn from a limited inventory of phonological primitives, typically High (H) and Low (L), which may combine to form complex pitch movements, such as rising or falling pitch accents (Ladd, 2008). Crucially, AM theory distinguishes between phonological representation and phonetic realization: tones are not transcriptions of fundamental frequency (f_0) values but abstract categories. Their phonetic expression is shaped by general principles of interpolation between a small set of specified tonal targets (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Pierrehumbert & Hirschberg, 1990). Phonetic representation, in this view, is an abstraction over "measurable properties of articulation, acoustics, and audition" (Pierrehumbert, 1990: 375).

A central feature of AM theory is its account of tonal alignment, which refers to the temporal coordination of tonal events with segmental material. As Ladd (2008) emphasizes, pitch accents like H* or L* are typically aligned with particular points in the speech stream (e.g., syllable onsets or vowel nuclei), and these alignment patterns can carry phonological significance. Differences in alignment timing can signal distinct intonational categories, and such distinctions can be perceived categorically by listeners (Ladd, 2008: 169–172). This is first evidenced in languages like Norwegian and Swedish, where distinct pitch accent types are differentiated by the alignment of their tonal peaks relative to the stressed syllable (Bruce, 1977; Haugen & Joos, 1952). The AM framework formalizes aspects of this through notational devices such as the star (*) convention, which indicates the 'starred tone' as the central tone aligned with the accented syllable (Pierrehumbert, 1980). However, this notation has been debated, as it can also imply a degree of intrinsic prominence beyond mere relational alignment, complicating its interpretation (Beckman & Pierrehumbert, 1986; Ladd, 2008).

To provide a nuanced understanding of how gradient phonetic variation relates to categorical phonological structures, AM theory makes a crucial distinction between *association* and *alignment* (Ladd, 2008). *Association* is an abstract phonological claim about which linguistic unit a tone or tonal group is linked to. These primary associations typically link pitch accents or boundary tones to higher-level prosodic constituents such as stressed syllables or intonational phrases. In contrast, *alignment* is a phonetic notion that captures the temporal coordination between tonal events, typically represented by f_0 movements, and specific segmental landmarks (e.g., consonant or vowel onsets).

Critically, AM theory also recognizes that some alignment patterns reflect phonologically meaningful distinctions. These are formalized as secondary associations, which specify how the individual tones that make up a pitch accent are anchored to finer-grained prosodic units (e.g., morae, syllable edges) (Pierrehumbert & Beckman, 1988). This capacity is further extended by AM theory's treatment of text-tune accommodation. Under conditions of temporal compression, like in fast speech, speakers may either compress tonal targets into shorter durations (e.g. Prieto, 1998) or truncate some targets entirely (e.g. Arvaniti et al., 1998). These adjustments vary across languages and contexts, reflecting lawful phonetic variation without implying phonological reanalysis.

A key contribution of the AM framework, particularly as developed by Pierrehumbert and Beckman (1988), is its hierarchical view of prosodic structure, in which tonal targets are associated with distinct levels of prosodic constituents, such as the prosodic word, accentual phrase, intermediate phrase, and intonational phrase. This model rejects a fully specified surface representation where every syllable or mora carries a tone (Haraguchi, 1977; Higurashi, 1983; Vance, 1987). Rather, tones are sparsely assigned with phonetic interpolation accounting for the resulting pitch contours. For instance, an H* pitch accent may associate with a metrically prominent syllable at the word level, while boundary tones (e.g., L%) and phrasal tones (e.g., H-) may align with the edges of larger constituents. This economy of tonal specification allows the model to capture a wide range of surface pitch phenomena without over-generating contrastive tones.

This hierarchical organization is particularly influential in the analysis of Japanese pitch accent, where lexical pitch accents interact with higher-level prosodic structure and post-lexical phenomena (Maekawa, 1997; J. Pierrehumbert & Beckman, 1988; Poser, 1984). These patterns are systematically captured in the X-JToBI annotation system (Maekawa et al., 2002), an extension of J_ToBI designed for Tokyo Japanese (Venditti, 2005). In this system, lexical pitch accent is marked as H*+L, indicating a sharp fall from a high tone on the accented mora to a following low tone.

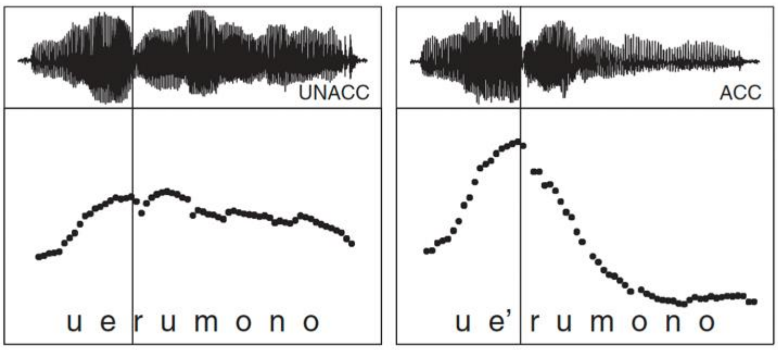


Figure 1 Spectrogram and f_0 contour of unaccented uerumono 'something to plant' accented ue'rumono 'the ones who are starved' phrases, uttered by the same speaker. The x-axis represents the time-course of the utterance, y-axis shows frequency (in Hz) of the f_0 contour. The vertical line marks the end boundary of the second mora in each phrase. Adapted from Venditti (2005).

Unlike in English, where pitch accents are assigned post-lexically to mark focus, Japanese pitch accent is lexically specified and remains fixed under focus. See Figure 1 for an example, where the pitch accent (marked with ‘) is the only distinction between the two phrases.

Instead, prominence is conveyed through post-lexical mechanisms such as pitch range expansion on focused constituents (often termed f_0 *reset* or *reset*) and compression on post-focal material (Venditti et al., 2012). These effects are distributed across two key prosodic domains: the Accentual Phrase (AP), a low-level prosodic grouping defined tonally by an initial rise, and the Intonation Phrase (IP), a higher-level prosodic grouping, within which pitch range is specified (Venditti, 2005). The IP often introduces a pitch reset and signals major boundaries. IPs are also marked by *boundary pitch movements* (BPMs) at their right edges, which can have various pragmatic functions (Igarashi, 2020; J. Pierrehumbert & Beckman, 1988; Venditti et al., 2012). Post-focal compression may also involve dephrasing, the reduction or deletion of boundary tones (L% H-) and accentual falls (J. Pierrehumbert & Beckman, 1988; Venditti et al., 2012). Together, these phenomena illustrate how the AM model, enriched by X-JToBI conventions, captures both lexical tonal structure and dynamic prosodic modulation in Japanese.

The flexibility of the AM model makes it a powerful tool for modeling the phonological structure of intonation across languages. However, its abstract representations do not always capture the full range of fine-grained phonetic variation, particularly in the alignment and realization of tones in short prosodic units. These issues become especially relevant when analyzing variation in the tonal realization of monomoraic and bimoraic words, as will be explored in the following chapters.

2.2. Prosodic Structure in Japanese dialects

While the AM model provides a flexible framework for describing tonal structure, it abstracts away from language-specific details that shape how tones are implemented phonetically. In Japanese, such details hinge crucially on prosodic units like the mora and the syllable, whose roles differ across dialects. This chapter focuses on the prosodic structure of Japanese dialects, beginning with an overview of the mora as the basic rhythmic and tonal unit (2.2.1). It then turns to the long-standing debate over whether the mora or the syllable serves as the primary tone-bearing unit (2.2.2), a distinction that carries important consequences for how tonal contrasts are realized. Finally, Chapter 2.2.3 compares two typologically divergent dialects, Tokyo Japanese and Kagoshima Japanese, illustrating how differences in prosodic structure condition the distribution and realization of tonal patterns. These contrasts lay the groundwork for understanding how dialect-specific prosodic grammars interact with external influences in cases of contact-induced change, as explored in later chapters.

2.2.1 The Mora as the Fundamental Prosodic Unit

Japanese is widely recognized as a mora-counting or mora-timed language. The mora functions as the basic, autonomous prosodic unit, dictating rhythm and prosodic measurement (Labrune, 2012). Native speakers' intuitions confirm that sequences like Okinawa [O-ki-na-wa] and Tokyo [to-o-kyo-o] are perceived as four beats or morae, despite having different numbers of segments or syllables. This rhythmic regularity is a hallmark of Japanese phonology.

Each mora is traditionally considered to be isochronous, meaning it occupies a similar duration in time (e.g. Han, 1962; Homma, 1981), though this notion has been contradicted experimentally by many researchers (e.g. Beckman, 1982; Hoequist, Jr., 1983a, 1983b; Warner & Arai, 2001). The mora also serves as the phonetic support for pitch accent, with pitch changes aligning with moraic boundaries (Ishihara, 2004; Labrune, 2012).

Japanese morae typically consist of a consonant-vowel (CV) structure, such as /sa/ or /ko/, or a lone vowel (V), like /a/ or /i/. Additionally, there are "moraic segments" (Labrune, 2012: 132), which are also counted as full prosodic units. These include the moraic nasal (/N/), as in konpon [kompoN] 'basis', the first part of an obstruent geminate (/Q/), as in gakkari [gak'kari] 'disappointed' and the second part of a long vowel or diphthong (/R/), as in imōto /imoRto/ 'younger sister' or kūkō /kuRkoR/ 'airport'. These special morae are considered structurally "deficient" (Labrune, 2012: 162) because they lack one of the two prototypical consonant or vowel components.

2.2.2 Mora vs. Syllable as Tone-Bearing Unit (TBU): The Ongoing Debate

While the mora's role in Japanese rhythm is widely accepted, its status as the sole or primary tone-bearing unit (TBU) is a subject of ongoing debate, especially in relation to the role of the syllable (Ito & Mester, 2019; Kubozono, 1996; Labrune, 2012; McCawley, 1968). The TBU is the prosodic domain to which phonological tones are associated (Ladd, 1996; 2008). Some phonologists argue that the syllable functions alongside the mora as a relevant unit of tonal association often drawing on the traditional distinction between "light" (monomoraic, CV, V) and "heavy" (bimoraic, CVV, CVN, CVC) syllables (Labrune, 2012). Even when tones ultimately align with moraic timing, their phonetic realization may reflect syllabic structure.

This typological question has a long history. Sibata (1962), drawing on a framework reminiscent to that of Trubeckoj (1958), proposed that Japanese dialects can be classified as either mora-based or syllable-based, depending on their use of basic prosodic units. Building on this, McCawley (1978) introduced a more refined typology that distinguishes between two independent notions: (1) the unit of counting, which determines how accent or tone position is calculated (e.g., counting from the right edge of the word), and (2) the accent- or tone-bearing unit, the prosodic unit on which prominence is phonetically realized. According to this typology, Kyoto Japanese is a mora-counting, mora-language, where tonal position is calculated in morae and the H tone may appear on either the head or non-head mora of a syllable. In contrast, Tokyo Japanese is a mora-counting, syllable-language: although the position of the accent is computed using morae, the H tone is always realized on the head mora of the relevant syllable, indicating that the syllable serves as the tone-bearing domain. Kagoshima Japanese, by contrast, exemplifies a syllable-counting, syllable-language. Here, both the computation of tonal position and the realization of the H tone rely on the syllable: the H tone is linked to and manifested on a syllable counted from the right edge of the word. Interestingly, no known Japanese dialect appears to represent the hypothetical opposite type – a syllable-counting, mora-language – in which syllables determine tonal location, but morae bear the accent (Kubozono, 2012a).

Finally, some researchers have argued against the need to invoke syllables at all in the analysis of Japanese prosody. Labrune (2012), in particular, challenges syllable-based interpretations and contends that many phenomena attributed to the syllable can be accounted for entirely through reference to morae and foot structure. For instance, structural distinctions between vowel sequences and diphthongs (e.g., hae 'fly' vs. kai 'shell') are often cited in support of syllabic parsing, but such distinctions lack consistent phonological or phonetic support. On this view, the mora remains the most reliable and necessary unit in the phonological system of Standard Japanese.

2.2.3 Dialectal Variation: Tokyo Japanese vs. Kagoshima Japanese

Japanese dialects display considerable variation in their pitch accent systems, and this diversity has played a central role in typological studies of prosodic structure. Following Uwano (1999) and Kubozono (2012), these systems can be broadly classified into two types: multi-pattern systems, such as that of Tokyo Japanese (TJ), and N-pattern systems, exemplified by Kagoshima Japanese (KJ).

In multi-pattern systems, the number of possible tonal patterns increases with word length. For example, TJ nouns follow the so-called n+1 rule, whereby a monosyllabic noun can exhibit two tonal patterns, a disyllabic noun three, and so on. This structure reflects a pitch-accent system in which the presence and position of an abrupt pitch fall (from high to low) is lexically specified. TJ distinguishes accented and unaccented words: accented words exhibit a pitch drop after the accented mora, while unaccented words lack this drop entirely, even when cliticized particles are attached (Haraguchi, 1977; Kubozono, 2007; J. Pierrehumbert & Beckman, 1988).

However, tonal contrast in monomoraic nouns is generally neutralized in citation forms. Minimal pairs such as /ha/ ‘leaf’ (unaccented) and /ha/ ‘tooth’ (accented) surface with identical rising or flat contours, obscuring their underlying accentual difference (Haraguchi, 1977; Kubozono, 2018b; McCawley, 1968; Uwano, 2018; Vance, 1995). In isolation, monomoraic nouns tend to adopt a default LH or high-level pattern, and their tonal distinctions become recoverable only in phrasal contexts. Bimoraic monosyllables, such as /baN/ ‘evening’ (unaccented) and ‘order’ (accented), occupy a liminal space in the TJ pitch accent system. While their greater prosodic length compared to monomoraic words theoretically allows for a clearer realization of an HL contour, e.g., a high pitch on the first mora followed by a low pitch on the second, the extent to which they retain contrast in citation form is not uniform. Kubozono (2018b) observes that in TJ most bimoraic monosyllables surface with a falling pitch, regardless of their underlying accentual type, indicating a near-complete neutralization of contrast in citation contexts. Therefore, the tonal contrast in the words /tō/ (‘tower’ or ‘ten’) is lost and realized with a falling HL pitch contour. Conversely, Labrune (2012) suggests that in some Tokyo-related dialects, a contrast in surface melodies for heavy bimoraic syllables can exist, like /kai/ 'paddle' (accented on first part) and /kai/ 'shell' (accented on second part). This implies some form of contrast retention. As such, bimoraic monosyllables provide a partial, but not unproblematic, window into the realization of pitch accent in short prosodic domains.

Nonetheless, while neutralization is the dominant pattern, there is some evidence for incomplete neutralization or residual tonal contrast even among monomoraic words in TJ. Vance (1995) reports that while most speakers produce no consistent difference between accented and unaccented forms in isolation, at least one speaker exhibited f_0 distinctions that aligned with the expected accentual categories in production and perception. This suggests that neutralization may not be fully categorical for all speakers or contexts, and hints at gradient phonetic traces of underlying accent even when contrast is said to be neutralized.

In contrast to TJ, KJ represents an N-pattern system, specifically, a two-pattern system. All lexical items fall into one of two categories: Type A, where the high tone falls on the penultimate syllable, and Type B, where it falls onto the final syllable. Unlike TJ, this system is insensitive to word length: tonal categories are fixed, and do not multiply with the number of syllables (Hirayama, 1951; Igarashi, 2014; Ishihara, 2000; Kubozono, 2004, 2007). The classification of KJ as a syllable-counting, syllable-language, as outlined in Chapter 2.2.2, has important implications for its tonal organization. In KJ, tonal melodies such as HL (Type A) or H (Type B) are assigned at the word level, as opposed to specific syllables or morae, and surface as consistent tonal patterns that hold even in monomoraic words (Hayata, 1999; Ishihara, 2004). Therefore, Type A items such as hi 'sunshine' and ki 'spirit' are realized with a falling (HL) contour, while corresponding Type B minimal pair items like hi 'fire' and ki 'tree' have a high level (H) tone (Hirayama, 1960; Ishihara, 2004; Kubozono, 2012a). The HL contour in Type A words is compressed into a single syllable, producing a perceptible fall even on monomoraic forms (Ito & Mester, 2019). This robust preservation of tonal contrast reflects KJ's typological classification as a "word tone dialect" (Hayata, 1999; Matsuura, 2014), in which tonal information is lexically specified for the entire word, rather than assigned to a specific segment, as is the case in the accent-based system of TJ.

However, the status of tonal contrast in short forms, particularly light monosyllabic words, is not without complications. Kibe (1997), as cited in Ishihara (2004), reports that some Type A and Type B minimal pairs may both surface with a high-level pitch, suggesting a loss of contrast in these forms. Ishihara (2004) further observes inconsistencies in the realization of Type A light monosyllables, proposing that pitch deletion rules may be necessary to account for their surface forms. These findings imply that maintaining a full two-tone (HL) contour on a single syllable or mora may sometimes be phonologically or phonetically unstable, even in a word tone system like KJ.

Supporting this view, Kubozono (2018b) also acknowledges the tonal neutralization in short forms noted by Kibe but adds a generational perspective: while older KJ speakers reportedly maintain the tonal contrast in monomoraic citation forms, realizing Type A with a fall and Type B with a level pitch, this distinction is increasingly lost among younger speakers.

A striking comparison arises from Koshikijima Japanese, a dialect closely related to KJ. Despite this genealogical proximity, Koshikijima exhibits complete tonal neutralization in monomoraic forms when spoken in isolation. This has been attributed to constraints like NOCONTOUR-μ (Ito & Mester, 2019), which prohibit the realization of contour tones on single morae. Unlike KJ, which relies on syllables as tone-bearing units, Koshikijima Japanese is a mora-counting, mora-language (Kubozono, 2012a, 2012b), and this prosodic typology appears to significantly influence whether tonal contrasts are maintained or lost in short prosodic domains.

The differing treatment of tonal contrast in these related dialects offers compelling evidence that prosodic structure fundamentally shapes the phonological behavior of tone. Such variation not only reflects dialectal divergence, but also raises broader questions about how phonetic realization shapes phonological systems. These questions will be addressed in the following chapter.

2.3. Sound Change as Phonetic Drift

Sound change has traditionally been treated as the domain of historical linguistics, but recent research emphasizes its close ties to synchronic phonetic variation and perceptual processing. Rather than viewing sound change as a purely categorical phenomenon, many current models argue that it emerges from the accumulation and reinterpretation of gradient phonetic patterns. This chapter explores how such variation can lead to phonological restructuring, beginning with general principles of phonologization (2.3.1) and continuing with theoretical models that account for the listener's role in sound change (2.3.2). The final section applies these insights to ongoing tonal change in Kagoshima Japanese (2.3.3).

2.3.1 From Phonetic Variation to Phonological Restructuring

Sound change is a fundamental aspect of linguistic evolution, shaping how languages develop and diverge across generations (Yu, 2023). A long-standing question in phonological theory concerns the nature of this change: does it occur through abrupt, categorical shifts in underlying representations, or through gradual phonetic drift that accumulates over time into structural reanalysis?

Historically, phonological processes such as neutralization were conceived as categorical, involving a loss or merger of contrast between discrete phonological categories. However, growing evidence from acoustic analyses suggests that neutralization is often incomplete: speakers may continue to produce subtle, sub-phonemic differences even in contexts where a contrast is supposedly neutralized (Port & O'Dell, 1985). For instance, studies have shown that durational differences between stop categories in a neutralizing context, like word-final devoicing in German or Dutch (Kleber et al., 2010; Warner et al., 2004), persist and are often pervasive. Vowel duration, in particular, stands out as a crucial cue for preserving the voicing contrast to some extent, even in languages like Polish where it is not the primary cue word-medially (Slowiaczek & Dinnsen, 1985). These residual cues can persist for generations, rendering apparent mergers phonetically gradient (Bukmaier et al., 2014). This challenges the idea of categorical replacement processes in assimilation (Pouplier et al., 2011). The auditory impression might be of a categorical change, but acoustic signals can still show remnants of the old or dialectal form, indicating a gradual shift. These findings align with physiological studies showing that even when a segment is no longer perceptible to listeners, subtle articulatory traces of that segment may still be present in speech (Pouplier, 2007; Pouplier & Hardcastle, 2005)

Similarly, sound change as a historical process is increasingly viewed not as an abrupt replacement of one category with another, but as a phonetically motivated drift that accumulates incrementally over time to redefine a phonological category

(Harrington, 2007, 2012; Harrington et al., 2019; Jansen & Mompean, 2023; Kataoka, 2009; Pittayaporn, 2018; Schertz & Clare, 2020; Stevens & Harrington, 2022; Yu, 2023). This perspective aligns with usage-based theories of speech perception, where phonological categories are defined by density distributions in an acoustic-perceptual space, continuously expanded by new remembered exemplars (Kleber et al., 2010; J. B. Pierrehumbert, 2001). This allows for a gradual shift in the distribution of variants along a continuum due to various factors, including external influences like dialect leveling (Bukmaier et al., 2014; Harrington et al., 2012).

This diachronic view is supported by apparent-time studies, which compare different age groups at a single point in time assuming that differences between older and younger speakers reflect ongoing sound change (Bailey, 2003; Sankoff & Blondeau, 2007). Such studies consistently reveal generational shifts in the acoustic realization of segments and tones. These shifts are often observable as small but consistent differences between age cohorts in various phonetic properties.

In Afrikaans, for instance, apparent-time studies reveal that younger speakers increasingly rely on f_0, not voice onset time (VOT) to signal voicing contrasts, suggesting a shift from a segmental to a suprasegmental cue system. This reflects a phonetic reorganization in which a historical voicing distinction in plosives is being replaced by tonal differences on the following vowel (Coetzee et al., 2018). A parallel development is observed in Seoul Korean: while older speakers distinguish lenis and aspirated plosives in phrase-initial position primarily through VOT (with aspirated stops showing longer VOT), younger speakers increasingly rely on f_0 cues. Specifically, historically lenis plosives are followed by low f_0, whereas aspirated plosives are followed by high f_0 (Bang et al., 2018; Kang, 2014; Kirby, 2013; Oh, 2011). Female speakers tend to exhibit larger f_0 distinctions and appear to lead the change, indicating a gender-linked trajectory in the ongoing merger of VOT categories (Kang, 2014; Oh, 2011). Taken together, these developments point to an incipient stage of tonogenesis, where tonal cues begin to supplant laryngeal features as primary contrastive markers.

These properties, over time, accumulate to redefine a phonological category, leading to phonological restructuring or reanalysis (Beddor, 2009, 2012). This process, termed phonologization, involves a non-distinctive phonetic pattern becoming a language-specific phonological pattern (Garrett & Johnson, 2013; Hyman, 1976). The idea is that phonetic biases, which are inherent to speech production and perception (e.g., motor planning, aerodynamic constraints, gestural mechanics, perceptual parsing), introduce non-random and directional variation into the pool of synchronic phonetic variation (Garrett & Johnson, 2013). Although many such variants are typically "filtered out" or compensated for by listeners (e.g. Fowler, 2005), some are not, leading to a shift in representation. When a listener perceives a phonetic effect but attributes it to a segment other than its original source, or assigns a different weighting to co-varying articulatory properties, it can initiate sound change even without misperception. This systematic yet variable behavior of speakers and listeners is seen as the foundation of gradual change. The changes are often gradual because the distribution of variants along a phonetic continuum is incrementally shifted due to internal (phonetic) or external (sociolinguistic) factors.

These findings highlight how systematic phonetic variation provide the raw material for sound change. Yet theories differ in how they model the transition from variation to change. Some, like Ohala's listener-based model, argue that categorical sound change arises from abrupt misinterpretations of phonetic input. Others, such as Beddor's perceptual reweighting model, emphasize gradual shifts in how listeners parse and prioritize coarticulatory cues. The following section explores these competing accounts in more detail, examining how different models conceptualize the initiation of sound change from both perceptual and articulatory perspectives.

2.3.2. Mechanisms of Reanalysis/From Coarticulation to Contrast

A central mechanism underlying phonologization is coarticulation: the systematic overlap of articulatory gestures that produces context-dependent phonetic variation. While listeners usually compensate for such variation (Fowler, 2005), it can under certain conditions be misattributed or reinterpreted as intrinsic to a particular segment. This shift from phonetic nuance to phonological contrast lies at the heart of many sound change theories (Beddor, 2009; Lindblom et al., 1995; Ohala, 1993). The following section examines three influential models that place coarticulation at the center of reanalysis: Ohala's misperception-based account, which frames change as a perceptual error; Lindblom's adaptive view of articulation, which highlights speaker-listener dynamics and the role of communicative economy in shaping phonetic structure; and Beddor's model of cue reweighting and parsing variability, which emphasizes listener-driven reinterpretation.

Ohala's highly influential theory, often termed the "listener as a source of sound change," posits that many sound changes originate from the listener's misparsing of coarticulatory cues (Ohala, 1981, 1989, 1993). He argued that the speech signal is "inherently ambiguous" (Ohala, 1981: 178) due to vocal tract constraints and gestural overlap, which introduce 'noise' or 'distortions'. While listeners typically compensate for these predictable coarticulatory variations, in rare instances, particularly with inexperienced listeners or learners, this compensation might fail or be incomplete. Through this 'hypocorrection' the listener interprets a contextually-induced phonetic effect as an inherent, distinctive sound feature, misattributing it from its original source. This reinterpretation, even if a 'mini sound change' within an individual's grammar, can then be publicly manifested when that listener becomes a speaker, producing the novel variant. For Ohala, this ‘perceptual switch’ is considered abrupt and categorical (Harrington et al., 2019; Kirby & Harrington, forthcoming). He noted that sound change is rare precisely because listeners are usually very adept at normalizing for context (Mann & Repp, 1980). Ohala also drew a sharp distinction between the *initiation* of sound change (the phonetic origins) and its *actuation* or *spread*, suggesting that the latter is often a matter of chance and not necessarily governed by phonetic principles (Ohala, 2012).

Lindblom's H&H theory provides another significant framework for understanding sound change, emphasizing speaker-listener interaction and the adaptive nature of speech (Lindblom, 1990; Lindblom et al., 1995). In H&H theory, motor control is purpose-driven: coarticulation is seen as an economical form of motor behavior (known as *hypospeech*), which the articulatory system defaults to when unconstrained. However, coarticulation, which may render contrastive differences less distinct, is avoided in listener-oriented speech (known as *hyperspeech*), that is, speech in which listeners’ informational needs are estimated to be high. Lindblom and colleagues (1995) agree with Ohala that unnormalized coarticulated variants can serve as the raw material for sound change. However, their key difference is that attention to such phonetic detail need not be the result of misperception or listener inexperience. Instead, listeners may occasionally be aware of the unnormalized or unprocessed acoustic form of the input signal, for example, when intelligibility demands are low or when sociolinguistic information encoded in the phonetic detail is more important (Beddor, 2009: 787). This "decontextualization" of speech is central to Lindblom's and Ohala's models, though the reason for decontextualization differs: Ohala links it to misinterpretation of coarticulation, while Lindblom attributes it to engagement of a listening mode usually reserved for semantically unpredictable speech during hypoarticulated speech. In Lindblom's model, the listener-turned-speaker may purposefully produce the new variant, even if the original speaker's intent was correctly apprehended. This approach suggests that listeners' sporadic access to the unprocessed phonetic form yields new pronunciations and puts language users in a "state of readiness for phonetic and phonological innovations" (Lindblom et al., 1995). Whether the new variant is accepted and used by others depends on a combination of articulatory, perceptual, and social factors (Yu, 2023).

Beddor's model offers a different, yet related, perspective on listener-based sound change, moving beyond Ohala’s strict notion of misperception (Beddor, 2009, 2012; Beddor et al., 2013). Her approach is "firmly grounded in the view that the systematic acoustic consequences of overlapping articulations are perceptually informative variants" (Beddor, 2009: 788). Beddor argues that sound change can occur even when listeners accurately perceive the input signal. Instead of misperception, the emphasis is on parsing variability or differential cue weighting.

Beddor's model crucially appeals to the idea that multiple phonological representations or different perceptual weightings of acoustic properties can be consistent with the variants found in everyday speech (Beddor, 2023). A listener might not arrive at the exact same representation or weighting as the speaker or other listeners. A key aspect of her theory is the role of articulatory covariation and perceptual equivalence (or trading relations) between a coarticulatory source and its effect (Beddor, 2009). In this scenario, the initial stage of phonologization often involves an inverse relationship: when the coarticulatory source is reduced or its alignment changes, the coarticulatory effect becomes more salient and reinterpreted as distinctive.

For example, in the phonologization of vowel nasalization (VN > Ṽ), Beddor (2009) hypothesized that shorter nasal consonants (the source) co-occur with more extensively nasalized vowels (the effect) due to variable alignment of the velum gesture. Listeners, experiencing this covariation, may perceive the nasality on the vowel and consonant as equivalent sources of information, or in contexts with very short nasals, prioritize the vocalic information. This systematic difference in perceptual weighting can lead to a gradual shift in the language's phonological categories (Gao & Kirby, 2024; Kirby & Harrington, forthcoming; Zellou & Cohn, 2024). This approach also contributes to the understanding that listener behaviors contributing to change are more systematic than sporadic (Beddor, 2023).

In both Ohala's and Lindblom's model, sound change at the level of the individual is phonetically abrupt (Harrington, 2012). Beddor's model and exemplar-based models (J. B. Pierrehumbert, 2001) tend to support a more gradual change, as the lexicon can store multiple phonetic forms associated with a word, and these forms are continuously updated through use. Ultimately, the models by Ohala, Beddor, and Lindblom offer complementary insights into the initiation of sound change. Ohala focuses on an "errorful" misperception, Beddor on differential cue weighting and parsing variability (even with accurate perception), and Lindblom on adaptive articulation in response to communicative needs and the occasional uptake of hypoarticulated forms. All these models recognize that the constant generation of phonetic variation, driven by inherent biases in speech production and perception, provides the "seeds" for sound change, even if only a small fraction of this variation eventually phonologizes and spreads through a community.

In sum, the study of sound change must account for both gradient phonetic processes and discrete phonological outcomes. These two dimensions are not mutually exclusive; rather, they represent different phases in a single dynamic process. As this thesis will show, the ongoing tonal reorganization in innovative KJ speakers exemplifies how contact-induced influence, such as increased sensitivity to pitch falls in Tokyo Japanese, may begin as subtle phonetic adjustment but ultimately lead to structural reinterpretation, mediated by the native prosodic grammar of the dialect.

2.3.3. Evidence of Ongoing Tonal Change in Kagoshima Japanese

The models discussed in the previous section offer a framework for understanding how low-level phonetic variation can gradually give rise to phonological restructuring. This section focuses on a specific instance of this process: the tonal reorganization currently underway in KJ.

Recent research has revealed that the pitch accent system of KJ has undergone considerable change over the past few decades, particularly among younger speakers (Kubozono, 2007, 2018b, 2018a). These changes appear to be driven by dialect contact with TJ, largely through media exposure, education, and broader socio-economic pressures (Kubozono, 2007, 2018a; Ota, 2020; Ota et al., 2016; Ota & Takano, 2014; Takano & Ota, 2017). Younger KJ speakers are frequently described as bilingual listeners, exhibiting perceptual awareness of both prosodic systems (Kubozono, 2007, 2018a). However, rather than fully adopting TJ forms, they appear to integrate select tonal features from TJ into their native grammar in systematic and prosodically constrained ways.

A key observation is that younger KJ speakers are sensitive to the presence or absence of a pitch fall, which serves as the primary cue distinguishing accented and unaccented words in TJ. This perceptual sensitivity has led to three notable patterns of tonal shift in KJ (Kubozono, 2018a, 2018b):

- Words traditionally Type A in KJ (with a pitch fall) tend to shift to Type B (no pitch fall) if they are unaccented in TJ (Kubozono, 2018a: 295).
- Conversely, words traditionally Type B in KJ (no pitch fall) tend to shift to Type A (with a pitch fall) if they are accented in TJ (Kubozono, 2018a: 292).
- In monosyllabic words, bimoraic forms (CVV, CVN, CVC) tend to adopt a falling contour (Type A), while monomoraic forms (CV, V) tend toward a flat realization (Type B), directly reflecting TJ accentual tendencies (Kubozono, 2018b: 45).

While TJ's influence is pervasive, its prosodic features are integrated selectively into the KJ system. Crucially, these innovations are reported to remain constrained by KJ's native prosodic grammar. Core features of the dialect, such as the two-pattern tonal system and syllable-based pitch assignment, remain largely intact (Kubozono, 2018a). Thus, what emerges is not a complete convergence with TJ's multi-pattern mora-based system, but rather a restructured KJ system that mirrors TJ's binary [±pitch fall] distinction through the lens of KJ's phonological architecture.

Kubozono (2018a: 313) underscores this point, noting that the surface forms produced by younger speakers are "not identical" to those of TJ. Even when a pitch fall is introduced to reflect a TJ-like accent, its alignment and phonetic realization remain consistent with KJ's syllable-counting structure, for example, falling on the penultimate syllable in Type A and the final syllable in Type B. As Kubozono (2018b: 37) emphasizes, young speakers selectively adopt the presence of a pitch fall while ignoring other prosodic cues in TJ, such as the exact position of the pitch rise or fall. The result is what Ota (2020: 92) describes as a “hybrid system”, in which TJ tonal features are filtered through KJ’s native phonological framework.

This evolving system raises important questions about how innovative speakers map new tonal categories onto existing prosodic templates. The present study aims to investigate these changes more closely by focusing on pitch accent realization in short prosodic forms, with three specific goals:

I. To test whether monomoraic words (CV) in younger KJ speakers consistently surface with a flat (Type B) pitch contour, reflecting a TJ-like neutralization as predicted by Kubozono (2018b).
II. To test whether bimoraic monosyllables (e.g., CVV, CVN) maintain a falling (Type A) pattern in isolation, as predicted by Kubozono (2018b).
III. To test whether tonal changes in younger KJ speakers reflect not only categorical phonological restructuring (i.e., lexical reanalysis of accent types), but also gradient phonetic drift in the shape and alignment of f_0 contours.

In addition to the analysis of younger KJ speakers, this study includes comparative data from older KJ speakers, focusing on monomoraic forms in citation. This allows us to assess whether the traditional Type A vs. Type B contrast remains intact in older generations and provides a diachronic reference point against which current changes can be measured.

Furthermore, a parallel analysis is conducted for monomoraic words in Tokyo Japanese, in order to test whether tonal neutralization in citation forms is as complete as often assumed. Following earlier findings by Vance (1995), this study explores the possibility of incomplete neutralization, where subtle phonetic differences in f_0 realization might still reflect underlying accentual distinctions in isolation.

By triangulating these three areas – younger KJ, older KJ, and TJ citation forms – this study aims to clarify the phonetic and phonological dynamics of tonal change and stability in short prosodic domains across dialects and generations.

3. Methodology

3.1. Speaker Groups

The data was collected as part of the current project SoundAct (2023-2028) at the Institute of Phonetics and Speech Processing (IPS), LMU Munich, which has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (grant agreement No. 101053194).

Three speaker groups were recorded: the control group of Tokyo Japanese (TJ) speakers with native fluency in Standard Japanese, raised and educated in the Tokyo region, older Kagoshima Japanese speakers (KJ-old) over 50 years old from the Kagoshima prefecture (mean age 68.1 years), with lifelong exposure to the local dialect, and younger Kagoshima Japanese speakers (KJ-young) aged approximately 26.2 years, born and raised in the Kagoshima prefecture but exposed to standard Japanese via media and schooling. All participants were native speakers of Japanese with no known speech or hearing impairments. The final dataset included 15 younger KJ speakers, 20 older KJ speakers and 35 TJ speaker. The precise numbers are detailed in the appendix.

3.2. Materials and Stimuli

A list of 46 lexical items was used, selected to test pitch accent alignment across two prosodic word types:

- Monomoraic words (32 items): CV/V-structured content words, such as /ha/ 'tooth'/'leaf' or /ki/ 'tree'/'feeling'
- Bimoraic words (14 items): Included CVN (e.g., /kan/ 'can'/'instinct') and CVV: (e.g., /tō/ 'ten'/'tower') words

Words were classified according to their accent status in both dialects, using prior lexicon-based descriptions:

- TJ-accented (TJ-A) vs. unaccented (TJ-U)
- KJ Type A (KJ-A) vs. KJ Type B (KJ-U)

To simplify the analysis and highlight broader contour parallels, KJ Type B words were grouped under the label KJ-U. This classification was motivated by the fact that KJ Type B and TJ-unaccented words both exhibit flat f_0 contours, whereas KJ Type A and TJ-accented words both involve a prominent pitch peak. While KJ-U and TJ-U contours do differ in alignment details, grouping them together emphasizes the contrast between contour vs. flat patterns, which is the focus of this alignment study. Stimuli were presented in isolation, without particles or carrier phrases. The full list of stimuli is included in the appendix.

3.3. Recording Procedure

Recordings were conducted in quiet environments using high-quality microphones and digital recorders. Each participant was instructed to produce each word three times in a natural but clear citation form. They were encouraged to speak at a natural pace. Recordings were saved at 44.1 kHz, 16-bit resolution. The audio was manually segmented and labeled in emuR (Jochim et al., 2016) with each token checked for clarity, fluency, and consistency. Tokens with disfluencies or extreme pitch artifacts were excluded.

3.4. Acoustic Analysis

3.4.1. f_0 Extraction and Smoothing

f_0 contours were extracted using REAPER (Talkin, 2015), an autocorrelation-based pitch tracker that is robust to microprosodic perturbations. Contours were extracted from each token and smoothed using Discrete Cosine Transform (DCT) filtering to remove microfluctuations. Semitones were calculated with reference to 100Hz. A 6-coefficient approximation was used to retain the overall pitch contour shape while minimizing microvariation due to segmental effects. The resulting contours were time normalized to 11 points per speaker. This allowed for direct comparison of contours across tokens with different durations, independent of their absolute timing.

3.4.2. Alignment Measures

The primary goal was to analyze the alignment of tonal targets in relation to the segmental string. Instead of labeling discrete landmark, which due to the nature of of mono-/bimoraic words posed difficult, if not impossible (bimoraic words are often words with final long vowel, diphthong or nasal), the analysis used continuous-time modeling of f_0 shape across normalized time. Alignment was assessed by comparing the overall contour shape, the location of maximum f_0 (peak) and slope and timing of the fall. These were compared across word type (monomoraic/bimoraic), accent class (A=accented/U=unaccented), and speaker group (Tokyo Japanese speakers, older Kagoshima Japanese speakers, younger Kagoshima Japanese speakers).

3.4.3. Statistical Modeling

To model f_0 contour shape over time and account for multiple levels of variation, this study employed Generalized Additive Mixed Models (GAMMs), using the `bam()` function from the mgcv package in R (Wood, 2017). f_0 trajectories, like other phonetic contours, tend to display complex, non-linear patterns over time. Traditional parametric approaches, such as linear regression, are often ill-suited to modeling these patterns accurately, as they rely on overly simplified assumptions about the shape of the data (Hastie & Tibshirani, 1990; Sóskuthy, 2021). GAMMs overcome this limitation by incorporating 'smooth terms' into their models. These smooths allow for continuous-time modeling of pitch contours, with normalized time (0–1) used as a continuous predictor. Fixed effects in the models included dialect group (`TJ`, `KJ` or `dialect2`), capturing overall group-level differences in f_0 trajectories. The models also included by-group smooths over time (`s(times, by = group)`), allowing contour shapes to vary non-linearly across dialects. Random smooths (factor smooths) were specified for both speakers and words (`s(times, speaker, bs = "fs", m = 1)` and `s(times, KAN, bs = "fs", m = 1)`), modeling item- and speaker-specific temporal deviations from the group-level contours. In addition, random intercepts were included for speakers and lexical items, with some nested by group, to account for overall speaker- and item-level variability (`s(KAN, bs = "re")`, `s(speaker, TJ, bs = "re")`, `s(KAN, dialect2, bs = "re")`, etc.). Table 1 provides an overview of the fixed effects and random smooths included in the first GAMM formula, illustrating how speaker, word, and group-specific variations were modeled. Specific formulas for all models are attached in the Appendix. To assess statistical differences between groups across the time course of the word, pairwise comparisons of f_0 trajectories were performed using the `plot_diff()` function from the itsadug package (van Rij et al., 2022). The resulting difference plots may also be found in the Appendix.

This GAM-based approach allows for nonlinear, continuous-time modeling of pitch contours, making it well suited for capturing gradient alignment differences that may not align with discrete phonological categories.

Component	Type	Description
TJ	Fixed effect	Main effect of dialect group (Tokyo Japanese)
s(times, by = TJ)	By-group smooth (fixed effect)	Smooth over time, estimated separately for each level of TJ
s(KAN, bs = 're')	Random intercept	Random intercept for lexical items (words)
s(speaker, TJ, bs = 're')	Random intercept (nested by TJ)	Random intercepts for speakers nested within TJ groups
s(times, speaker, bs = 'fs',m = 1)	Random smooth (factor smooth)	Allows speaker-specific non-linear deviations over time
s(times, KAN, bs = 'fs', m = 1)	Random smooth (factor smooth)	Allows word-specific non-linear deviations over time
s(speaker, bs = 're')	Random intercept	Random intercept for speakers

Table 1 Components of the Generalized Additive Mixed Model (GAMM) Formula for monomoraic TJ words.

4. Results

This chapter presents the f_0 alignment results from monomoraic and bimoraic words produced in isolation by three speaker groups: Tokyo Japanese (TJ), older Kagoshima Japanese (KJ-old), and younger Kagoshima Japanese (KJ-young). Both pitch accent category (accented vs. unaccented) and alignment behavior are examined, focusing on whether KJ-young speakers pattern with TJ, retain traditional KJ alignment, or show intermediate patterns. All results are based on normalized f_0 contours modeled with Generalized Additive Mixed Models (GAMMs).

4.1. Monomoraic Words

This section begins by establishing the baseline behavior in monomoraic words in TJ and in older KJ speakers, before turning to younger KJ speakers to test, whether tonal contrast is also neutralized in their production of monomoraic forms.

4.1.1. Incomplete Neutralization in Tokyo Japanese (TJ)

In line with previous research, Tokyo Japanese is expected to show neutralization of pitch accent in monomoraic words when spoken in isolation (Haraguchi, 1977; Kubozono, 2018b; McCawley, 1968; Uwano, 2018). That is, even lexically accented items (e.g., /ha/ ‘tooth’, /ki/ ‘tree’) are often realized with flat or slightly rising contours, similar to unaccented forms. This is predicted by the X-JToBI model and has been observed in citation-form speech (Maekawa et al., 2002; Venditti, 2005).

However, the present results in Figure 2 reveal an unexpected pattern: instead of full neutralization, TJ speakers produced systematically distinct contours for lexically accented vs. unaccented monomoraic words. Specifically, f_0 contours for TJ-accented items show a small but consistent rise followed by a fall, or at least a higher overall pitch compared to unaccented items. In contrast, unaccented words tended to maintain a flatter contour. This pattern is supported by the significance bars shown at the bottom of Figure 2, which indicate time intervals where the contours for accented and unaccented items differ significantly. Significance bars were derived from plot_diff() comparisons[1] of the fitted GAMM smooths. Bars indicate time intervals where differences between speaker groups are statistically significant at the 95% level.

[1] Difference plots, formulas and statistic model summaries for all GAMMs attached in the appendix.

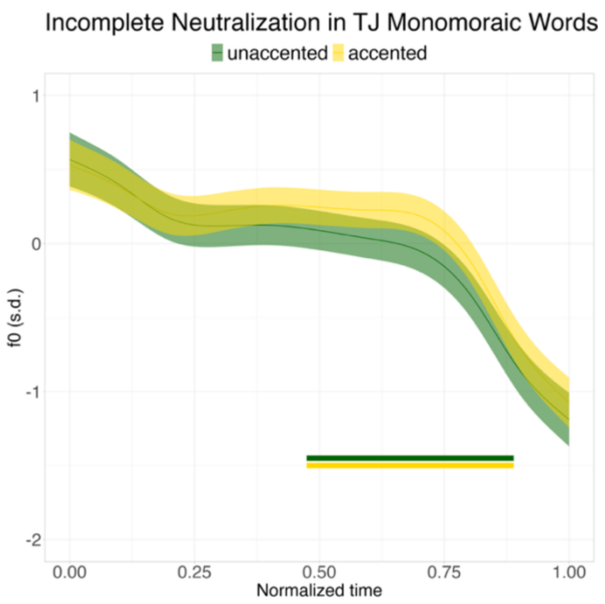


Figure 2 Comparison of monomoraic words that are TJ-A (yellow)/TJ-U (green). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure shows the interval over which the categories are significantly different.

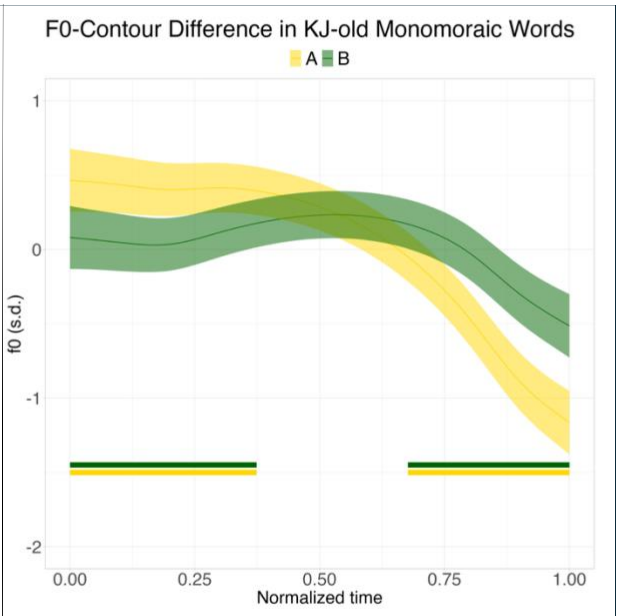


Figure 3 Comparison of monomoraic words that are KJ Type A (green) and KJ Type B (yellow). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure shows the interval over which the categories are significantly different.

Note that across all GAMM plots, a downward f_0 movement occurs near the end of the word – even in forms that are expected to be flat, such as unaccented (TJ) or Type B (KJ) items. This pattern likely reflects the presence of a low boundary tone (L%), given that all stimuli were produced in isolation. Consequently, the most informative differences in pitch shape occur earlier in the contour, particularly before the final portion of the word.

4.1.2. Contrast Maintenance in Kagoshima Japanese – Older Speakers (KJ-old)

Older KJ speakers, on the other hand, produced monomoraic words with clear and consistent differences between Type A and Type B accent classes, reflecting the well-documented two-pattern system in Kagoshima Japanese. Specifically, Type A words were realized with a falling f_0 contour, typically beginning high and falling within the word, while Type B words show an overall higher f_0 and a flatter overall shape as can be observed in Figure 3. These patterns indicate a robust distinction between the two accent classes at the phonetic level. The presence of a clear f_0 fall in Type A and its absence in Type B provides strong evidence that older KJ speakers preserve the traditional pitch-accent contrast, even in prosodically minimal monomoraic words.

4.1.3. Neutralization in Kagoshima Japanese – Younger Speakers (KJ-young)

This section examines how younger speakers of Kagoshima Japanese (KJ-young) produce monomoraic words that differ in pitch accent classification across dialects, focusing on whether tonal neutralization occurs and how their f_0 realizations compare to both traditional KJ (KJ-old) and Tokyo Japanese (TJ) patterns.

(a) KJ-A and TJ-U monomoraic words

KJ-young speakers pattern closely with Tokyo Japanese speakers in their production of monomoraic words. Specifically, items that were produced with a clear pitch fall by KJ-old speakers (i.e., KJ-A) are realized by KJ-young with a flat f_0 pattern, mirroring the unaccented form observed in TJ. As shown in Figure 4, no significant differences were found between TJ and KJ-young speakers across the time span of these words, as indicated by the absence of significance bars (grey+red) for these groups. This suggests that KJ-young speakers have adopted Tokyo-like lexical accent categories and realize them phonetically in a manner consistent with the unaccented monomoraic pattern of TJ.

(b) KJ-U and TJ-A monomoraic words

In the subset of monomoraic words that are classified as KJ-U and TJ-A, all three speaker groups are expected to produce flat f_0 contours. This is because pitch accent is almost completely neutralized in monomoraic words in Tokyo Japanese, meaning that accented TJ-A items surface as unaccented in isolation.

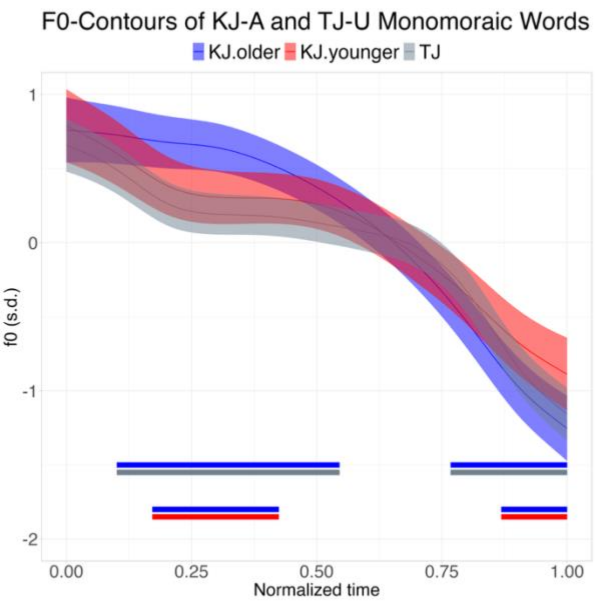


Figure 4 Comparison of monomoraic words that are KJ-A and TJ-U for the speaker groups KJ-older (blue), KJ-younger (red) and TJ (grey). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure show the intervals over which the categories are significantly different.

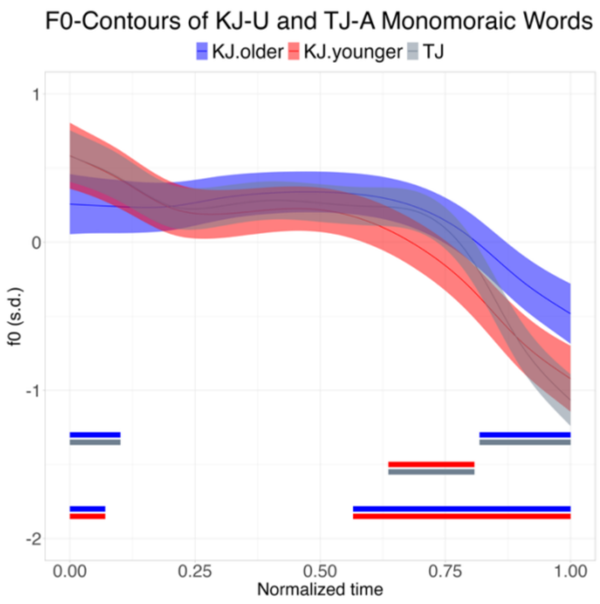


Figure 5 Comparison of monomoraic words that are KJ-U and TJ-A for the speaker groups KJ-older (blue), KJ-younger (red) and TJ (grey). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure show the intervals over which the categories are significantly different.

If younger KJ speakers are adopting the Tokyo accent system, we would expect them to reclassify these items as unaccented and produce a Type B contour (i.e., flat) consistent with the unaccented pattern in KJ. This prediction is supported by the data: as shown in Figure 5, all three groups – TJ, KJ-old, and KJ-young – produce flat contours, and statistical modeling confirms no significant difference in f_0 in the middle of the word. However, there is a subtle but important distinction in the contour shape. The f_0 trajectory produced by KJ-young speakers more closely resembles the TJ contour than the KJ-old shape, as can be seen in the significance lines, as blue+red lines are the longest.

4.1.4. Summary of Findings

Analyses of monomoraic words across speaker groups reveal distinct patterns of tonal realization. In Tokyo Japanese, speakers do not fully neutralize pitch accent in isolation as previously assumed, as accented and unaccented items still differ subtly in f_0, indicating incomplete neutralization. In contrast, older Kagoshima Japanese speakers maintain a clear phonetic distinction between Type A and Type B patterns, with Type A showing a falling contour and Type B remaining flat. However, younger KJ speakers show (near-)complete neutralization of this contrast: comparisons show that KJ-young speakers closely align with TJ speakers in both the realization of KJ-A words (flattened, TJ-like) and KJ-U words (all flat), suggesting convergence toward a Tokyo-like prosodic system and a restructuring of lexical tone categories.

4.2. Bimoraic Words

Bimoraic words offer a clearer window into alignment differences across speaker groups as, unlike monomoraic words, they do not undergo pitch accent neutralization in Tokyo Japanese, making it possible to compare f_0 contours across all three speaker groups.

4.2.1. Alignment Patterns in Accented Bimoraic Words

This section investigates alignment behavior in accented bimoraic words across speaker groups to assess whether younger KJ speakers maintain traditional KJ phonetic realization or show evidence of change toward Tokyo Japanese. Two hypotheses guide this analysis: *(1)* If TJ speakers time the f_0 fall to the boundary between the two morae, their contours should show a sharper fall early in the word compared to KJ-old speakers. In KJ, which lacks moraic structure, the fall may be distributed more evenly across the syllable. *(2)* If no significant phonetic change has occurred across generations in KJ, younger KJ speakers should resemble older KJ speakers in their realization of KJ-A forms.

To isolate alignment differences from lexical variation, the analysis focuses on accented items across all three groups:

- KJ-A words produced by KJ-old speakers
- TJ-A words produced by TJ speakers
- KJ-A/TJ-A words produced by KJ-young speakers

Note that for KJ-young, only those KJ-A words that are also accented in TJ are included. This restriction accounts for the fact that KJ-young speakers may produce flat f_0 contours for KJ-A words that correspond to unaccented forms in TJ, suggesting lexical reanalysis. Figure 6 presents the results. As expected, TJ speakers produce early f_0 peaks followed by steep falls, with the pitch drop occurring near the mora boundary. KJ-old speakers, by contrast, exhibit slightly later and lower peaks, with a more gradual fall across the word. The two groups diverge significantly at the onset of the segment, but their contours converge in the later portion of the word *(1)*. KJ-young speakers display a third, distinct pattern. Their contours begin falling at roughly the same time as those of TJ speakers, but the fall is less steep, and the pitch range is compressed. This results in a more linear overall shape. Although statistically closer to KJ-old in parts of the contour, the phonetic realization of KJ-young speakers differs noticeably from both reference groups.

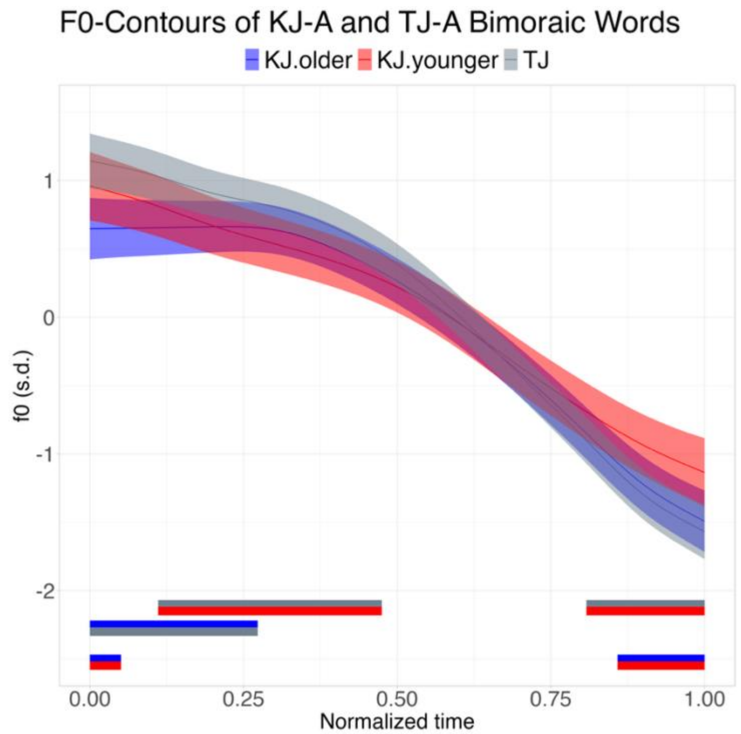


Figure 6 Comparison of bimoraic words that are KJ-A (produced by KJ-old), TJ-A (produced by TJ), and KJ-U/TJ-A (produced by KJ-young) for the speaker groups KJ-older (blue), KJ-younger (red) and TJ (grey). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure show the intervals over which the categories are significantly different.

These results suggest that KJ-young speakers are not simply replicating either the TJ or KJ-old alignment pattern. Instead, they appear to be developing a novel realization that is neither fully Tokyo-like nor traditionally Kagoshima. This shift points to a gradual phonetic reorganization, potentially signaling the emergence of a new alignment norm in the speech of younger KJ speakers.

4.2.2. KJ-young in Reanalyzed Lexical Categories

In addition to accent-matched comparisons, two further analyses explore how KJ-young speakers realize words that differ in lexical pitch accent across dialects. These conditions provide critical insight into the shift observed in KJ-young.

(a) KJ-A (KJ-old) vs. TJ-U vs. KJ-young

In this comparison, the relevant lexical items are accented in KJ (Type A) and unaccented in TJ. If KJ-young speakers are aligning with Tokyo Japanese, they should reclassify these items as unaccented and produce flat contours, phonetically resembling Type B realizations in KJ-old. However, the data show that KJ-young speakers produce contours that are intermediate. As shown in Figure 7, their f_0 contours are more linear and less dynamic than those of TJ or KJ-old. The contour does not exhibit the flattening of TJ unaccented items, nor the typical pitch drop observed in Type A words in KJ-old speakers.

Statistically significant differences between KJ-young and TJ can be observed in the first and last quarters of the trajectories between KJ-young and KJ-old. This suggests that KJ-young speakers are influenced by the TJ unaccented category, but that the phonetic realization is not yet stable or target-like. The in-between shape likely reflects a transitional stage in phonetic convergence, rather than an abrupt categorical re-mapping from KJ-A to TJ-U.

(b) KJ-U (KJ-old) vs. TJ-A vs. KJ-young

Here, the lexical items are unaccented in KJ (Type B) and accented in TJ. If KJ-young speakers have adopted the Tokyo system, they should produce falling contours, reflecting the TJ-accented pattern. This prediction is largely confirmed: as depicted in Figure 8, KJ-young speakers do produce falling contours for these words. However, their contour shape differs from both reference groups. The TJ fall is steeper and peaks earlier, while the KJ-old contour remains mostly flat, as expected. KJ-young speakers again show a more linear fall, shallower than TJ but still clearly falling, unlike KJ-old. Statistical comparisons confirm that all three groups differ significantly in f_0 for most of the time span of the word. Only during a brief mid-word interval do the contours converge, showing no significant difference.

As in the previous comparison, the f_0 trajectory of KJ-young speakers is more linear and lacks the curvature found in the other two groups, reinforcing the interpretation that KJ-young are phonetically approximating the TJ realization, but still retain prosodic characteristics of KJ. Together, these two comparisons reinforce the interpretation that KJ-young speakers are shifting both lexically and phonetically, but that the phonetic implementation is progressing more gradually, with alignment and contour shape still developing toward the Tokyo target.

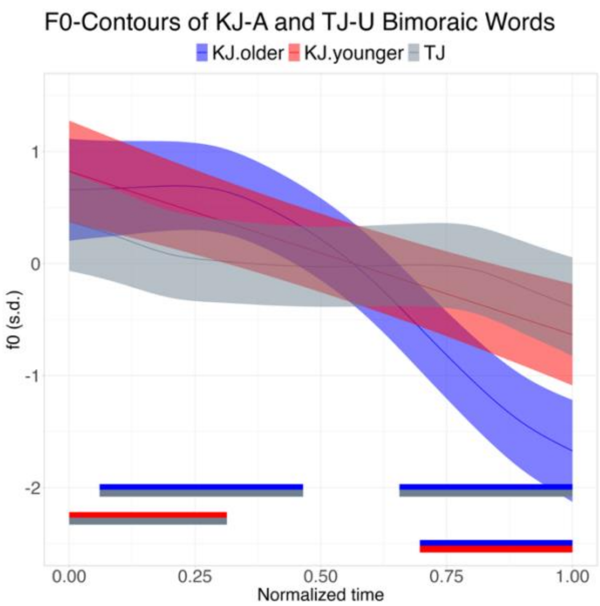


Figure 7 Comparison of bimoraic words that are KJ-A and TJ-U for the speaker groups KJ-older (blue), KJ-younger (red) and TJ (grey). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure show the intervals over which the categories are significantly different.

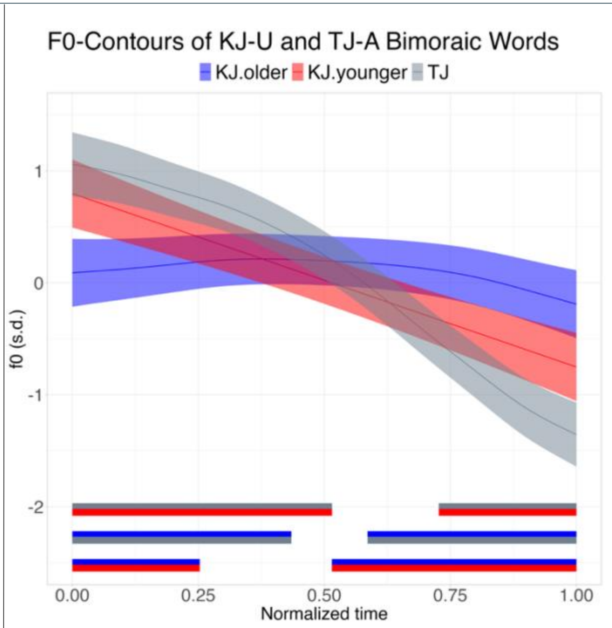


Figure 8 Comparison of bimoraic words that are KJ-U and TJ-A for the speaker groups KJ-older (blue), KJ-younger (red) and TJ (grey). Bands show pointwise 95%-confidence intervals. The colored bars at the bottom of the figure show the intervals over which the categories are significantly different.

4.2.3. Summary: Evidence for Gradient Phonetic Change

The bimoraic word data provide strong support for ongoing change in the prosodic system of younger Kagoshima speakers. In an accent-matched comparison, TJ speakers exhibit compact, steep HL falls characteristic of early-aligned, mora-based pitch accents, whereas KJ-old speakers produce later peaks and gradual, less compressed pitch falls, consistent with a broader or syllable-based alignment domain. In unaccented words both TJ and older KJ speakers show a flat f_0 contour. KJ-young speakers fall between these two patterns in both alignment timing and contour shape. Across both accent-matched and cross-accent comparisons, KJ-young speakers consistently produce intermediate f_0 contours: they are more linear and less steep than accented bimoraic words produced by both TJ and older KJ speakers. Moreover, innovative KJ speakers show falling contours in both conditions, even though a flat contour was expected for TJ-U words. Therefore, bimoraic monosyllables seem to be neutralizing to falling f_0 contours. These findings support the conclusion that KJ-young are undergoing a gradient phonetic shift in alignment, likely influenced by dialect contact and increasing exposure to Tokyo Japanese, but without full phonological restructuring of the pitch accent system.

5. Discussion

This study has explored the prosodic evolution of Kagoshima Japanese (KJ) in younger speakers through a detailed investigation of tonal patterns in monomoraic and bimoraic words in isolation. The study tested three core hypotheses to evaluate whether younger speakers show signs of pitch accent neutralization (I & II), phonetic drift, or tonal category shifts under the influence of Tokyo Japanese (TJ) (III). Using Generalized Additive Mixed Models (GAMMs) to analyze f_0 contours, this study provides compelling evidence of both phonological restructuring and fine-grained phonetic variation, pointing toward a gradual, contact-induced sound change.

(a) Monomoraic Neutralization and Lexical Reanalysis

The first hypothesis predicted that innovative KJ speakers would neutralize the pitch accent contrast in monomoraic words (CV, V), producing flat contours regardless of whether the word was historically Type A or Type B (Kubozono, 2018b). The results confirm this prediction. Younger KJ speakers show neutralization of this contrast: both KJ-A and KJ-B words are produced with flat contours, closely resembling the unaccented realizations observed in TJ, where monomoraic forms in isolation are nearly completely neutralized. This pattern reflects not only a shift in tonal category assignment but also a strong phonetic convergence with TJ. In other words, younger KJ speakers do not merely reinterpret tonal categories within the framework of their native dialect, they may also adopt the characteristic pitch contour shapes of the mora-timed TJ system, instead of preserving the syllable-based alignment patterns of traditional KJ. This contradicts the notion that the two-pattern system of KJ is simply being filtered through a new contrast (accented vs. unaccented). The findings indicate a deeper restructuring of the prosodic system, both in terms of phonological categories and phonetic realization. In contrast, older KJ speakers maintain a clear phonetic distinction between Type A and Type B, preserving the traditional pitch fall in Type A words. Altogether, the results suggest that younger speakers have adopted a Tokyo-like model of pitch accent neutralization in isolated monomoraic forms, leading to both phonological reanalysis and phonetic convergence.

(b) Bimoraic Neutralization and Phonetic Drift

The second hypothesis proposed that isolated bimoraic monosyllables (e.g., CVV, CVN) would lose the Type A/B contrast in younger KJ speakers and be realized as a falling pattern by innovative speakers (Kubozono, 2018b). On a phonological level, the results support this hypothesis: younger speakers consistently produce a falling f_0 contour in bimoraic words, indicating a reanalysis of these items as Type A. However, the phonetic realization of this fall diverges notably from that of the older generation. While older KJ speakers exhibit relatively later and more gradual pitch falls aligned with a syllable-based prosodic system, younger KJ speakers produce earlier, shallower falls. These contours are more linear and compressed in pitch range, marking a shift in the phonetic implementation of the same tonal category. In contrast, TJ speakers display steep, early-aligned HL falls characteristic of mora-based alignment. Notably, though all contours show a fall, they are not uniform across the KJ-young group: there is some variation in shape and alignment even within this group (compare red contours in Figure 6 vs. in Figures 7&8). This internal variation suggests that younger speakers are not producing a single stable phonetic target, but rather a range of falling contours, indicating that the phonetic implementation of these tonal categories is still undergoing change. Compared to both speaker groups KJ-old and TJ speakers, KJ-young contours remain distinct in both timing and shape. This points not to a complete adoption of TJ phonetics, but to a gradual phonetic reorganization within the KJ system. Instead of simply mapping TJ tonal categories onto their native system, younger KJ speakers appear to be developing new phonetic realizations through processes of phonetic drift and variation.

To summarize, the findings across both monomoraic and bimoraic monosyllables in isolation point to a process of tonal neutralization among younger KJ speakers. In monomoraic forms, this neutralization is both phonological and phonetic: younger speakers not only reclassify traditionally accented items as unaccented but also produce contours that closely mirror those of TJ unaccented forms. This indicates a strong convergence with the Tokyo system, reflecting both a shift in tonal category and a corresponding phonetic realignment. In bimoraic forms, the picture is more complex. While younger KJ speakers consistently produce a pitch fall, the phonetic realization of this fall departs notably from both older KJ and TJ patterns. The contours differ in timing, shape, and steepness, and even show variability across items within the younger speaker group itself.

These patterns align closely with Beddor's model of sound change as a process of perceptual cue reweighting and gradual phonetic drift. In particular, the observation that KJ-young speakers produce pitch falls that differ acoustically from both KJ-old and TJ speakers suggests that the shift is not a matter of abrupt misperception or categorical replacement. Rather, younger speakers appear to be reinterpreting and reimplementing tonal categories through subtle adjustments in alignment and contour shape. This results in intermediate forms, yet the produced contours do not match the traditional KJ realization of Type A. Such discrepancies provide clear evidence for a transitional phonetic state in which speakers are no longer faithfully reproducing older patterns, but have not yet converged on a new stable form. This trajectory exemplifies the kind of listener-driven, phonetically gradient restructuring described in Beddor’s framework, where small shifts in acoustic realization accumulate across generations and reshape the phonological system from within.

Interestingly, the data also challenge the assumption of full neutralization in TJ. While traditionally thought to neutralize accent in monomoraic forms (Haraguchi, 1977; Kubozono, 2018b; McCawley, 1968; Uwano, 2018; Vance, 1995), TJ speakers in this study exhibited consistent, albeit subtle, pitch differences between accented and unaccented forms. This persistence of subtle phonetic differences indicates that neutralization in TJ is incomplete, a finding consistent with a broader trend in phonetic literature across various languages. Many studies have demonstrated that phonological contrasts, even when traditionally considered neutralized, often retain sub-phonemic acoustic cues in production that can sometimes be perceived by listeners. For instance, research on final devoicing in German and Dutch shows that underlying voicing contrasts can be incompletely neutralized, with measurable durational and other acoustic differences persisting in neutralized contexts, and listeners sometimes exploiting these cues for discrimination (Bukmaier et al., 2014; Kleber et al., 2010; Port & O’Dell, 1985; Warner et al., 2004). Specifically for TJ, evidence comes from studies like Vance (1995), which found that one Tokyo speaker consistently produced a distinction in f_0 between final-accented and unaccented forms in isolation, and that this speaker could also perceive this distinction in their own speech with greater than chance accuracy. The concept of "incomplete neutralization" suggests that phonetic detail is an integral part of speech communication. Even when a phonological contrast is said to be neutralized, articulatory traces or subtle acoustic remnants of the underlying distinction may still be observable (Bukmaier et al., 2014; Pouplier, 2007; Pouplier & Hardcastle, 2005). These fine phonetic details can be perceptually salient and can even drive sound change over time, often through processes where listeners may "misinterpret" coarticulatory effects as inherent features of a segment. The subtle f_0 differences observed in TJ are therefore part of a well-documented phenomenon where phonetic cues, though not perceived as distinct categories, contribute to the intricate acoustic landscape of a language's phonological system.

6. Conclusion

This study presents evidence that younger KJ speakers are undergoing a tonal shift driven by contact with TJ. While both monomoraic and bimoraic words show neutralization and category shifts, bimoraic words diverge phonetically from both reference systems. The observed patterns exemplify phonetic drift, where change begins in the gradient realization of prosodic structure and may eventually result in phonologization.

While this thesis has focused on isolated mono- and bimoraic words to examine tonal alignment patterns in Tokyo and Kagoshima Japanese, a promising direction for future research lies in analyzing the same words when followed by monosyllabic particles such as /ga/ or /mo/. In polysyllabic words, the H tone in KJ surfaces on the penultimate syllable in Type A words and on the final syllable in Type B words. Thus, when additional syllables are added, the tonal target shifts rightward within the prosodic phrase. In contrast, Tokyo Japanese speakers produce the pitch fall consistently within the word, regardless of following particles. These alignment differences in phrasal contexts offer valuable insight into the interaction between tonal structure and prosodic phrasing. Additionally, analyzing monomoraic words with particles may help resolve ambiguities in isolation-based alignment patterns. For instance, TJ tonal contrasts re-emerge in phrases, providing more reliable cues for lexical accent categories.

Another valuable direction for future research would be to investigate the perceptual side of this tonal reorganization. While this thesis has focused on production, exploring how listeners from different age groups in Kagoshima perceive and categorize tonal patterns, especially those that fall between traditional KJ and TJ contours, could provide crucial insight into whether phonetic drift is accompanied by shifts in perceptual norms.

References

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, *26*(1), 3–25. https://doi.org/10.1006/jpho.1997.0063

Bailey, G. (2003). Real and apparent time. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 312–332). Blackwell Publishing Ltd. https://doi.org/10.1111/b.9781405116923.2003.00018.x

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005

Beckman, M. (1982). Segment Duration and the 'Mora' in Japanese. *Phonetica*, *39*(2–3), 113–135. https://doi.org/10.1159/000261655

Beckman, M., & Pierrehumbert, J. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. JSTOR.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language*, *85*(4), 785–821. https://doi.org/10.1353/lan.0.0165

Beddor, P. S. (2012). Perception grammars and sound change. In M.-J. Solé & D. Recasens (Eds.), *Current Issues in Linguistic Theory* (Vol. 323, pp. 37–56). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.323.06bed

Beddor, P. S. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics*, *98*, 101228. https://doi.org/10.1016/j.wocn.2023.101228

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350–2366. https://doi.org/10.1121/1.4794366

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund.

Bukmaier, V., Harrington, J., & Kleber, F. (2014). An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: Evidence for a gradient sound change. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00828

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. https://doi.org/10.1016/j.wocn.2017.09.009

Duanmu, S. (1992). An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman Area*, *15*(1). https://doi.org/10.32655/ltba.15.1.06

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, *33*(2), 199–213. https://doi.org/10.1016/j.wocn.2004.10.003

Gao, J., & Kirby, J. (2024). Laryngeal contrast and sound change: The production and perception of plosive voicing and co-intrinsic pitch. *Language*, *100*(1), 124–158. https://doi.org/10.1353/lan.2024.a922001

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of Sound Change* (pp. 51–97). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573745.003.0003

Han, M. (1962). The Feature of Duration in Japanese. *Onsei No Kenkyuu [Studies in Phonetics]*, *10*, 65–80.

Haraguchi, S. (1977). *The tone pattern of Japanese: An autosegmental theory of tonology*. Kaitakusha.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *Laboratory Phonology*, *9*, 125–144.

Harrington, J. (2012). The relationship between synchronic variation and diachronic change. In *The Oxford Handbook of Laboratory Phonology* (pp. 321–332). Oxford University Press.

Harrington, J., Kleber, F., & Reubold, U. (2012). The production and perception of coarticulation in two types of sound change in progress. *Speech Planning and Dynamics*, 39–62.

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2019). The phonetic basis of the origin and spread of sound change. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 401–426). Routledge.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall/CRC.

Haugen, E., & Joos, M. (1952). Tone and Intonation in East Norwegian. *Acta Philologica Scandinavica*, *22*, 41–64.

Hayata, T. (1999). *Oncho no Taiporoji [The Typology of Tone]*. Taishukan Shoten.

Higurashi, Y. (1983). *The Accent of Extended Word Structures in Tokyo Standard Japanese, The*. EDUCA. https://books.google.de/books?id=VauwAAAAIAAJ

Hirayama, T. (1951). *Kyushu hogen oncho no kenkyu [Studies on tone in Kyushu dialects]*. Gakkai no Shishin-sha.

Hirayama, T. (1960). *Zenkoku Akuseno Ziten [All Japan Accent Dictionary]*. Meiji Shoin.

Hoequist, Jr., C. (1983a). Durational Correlates of Linguistic Rhythm Categories. *Phonetica*, *40*(1), 19–31. https://doi.org/10.1159/000261679

Hoequist, Jr., C. (1983b). Syllable Duration in Stress-, Syllable- and Mora-Timed Languages. *Phonetica*, *40*(3), 203–237. https://doi.org/10.1159/000261692

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, *9*(3), 273–281. https://doi.org/10.1016/s0095-4470(19)30971-4

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies presented to Joseph H. Greenberg* (pp. 401–418). Amma Libri.

Igarashi, Y. (2014). *Typology of intonational phrasing in Japanese dialects* * (pp. 464–492). https://doi.org/10.1093/acprof:oso/9780199567300.003.0015

Igarashi, Y. (2020). *Intonation in Japanese dialects*.

Ishihara, S. (2000). Linguistic-Tonetic Differences in Target-Tone Realisation: Standard vs Kagoshima Japanese. *Proceedings of the 8th Australian International Speech Science and Technology Conference*, 192–197.

Ishihara, S. (2004). *An acoustic-phonetic descriptive analysis of Kagoshima Japanese tonal phenomena*. [The Australian National University].

Ito, J., & Mester, A. (2019). Pitch accent and tonal alignment in Kagoshima Japanese. *The Linguistic Review*, *36*(1), 1–24. https://doi.org/10.1515/tlr-2018-2005

Jansen, S., & Mompean, J. A. (2023). GOOSE -fronting in Received Pronunciation across time: A trend study. *Language Variation and Change*, *35*(1), 55–77. https://doi.org/10.1017/S0954394523000017

Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2016). *emuR: Main Package of the EMU Speech Database Management System* (p. 2.5.0) [Dataset]. https://doi.org/10.32614/CRAN.package.emuR

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76–90. https://doi.org/10.1016/j.wocn.2014.03.005

Kataoka, R. (2009). A Study on Perceptual Compensation for /u/-fronting in American English. *Annual Meeting of the Berkeley Linguistics Society*, *35*(1), 156. https://doi.org/10.3765/bls.v35i1.3607

Kirby, J. (2013). The role of probabilistic enhancement in phonologization. In A. C. L. Yu (Ed.), *Origins of Sound Change* (pp. 228–246). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573745.003.0011

Kirby, J., & Harrington, J. (forthcoming). Acoustic-perceptual factors in the actuation of sound change. *The Wiley Blackwell Companion to Diachronic Linguistics*.

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *38*(2), 185–196. https://doi.org/10.1016/j.wocn.2009.10.001

Kubozono, H. (1996). Syllable and Accent in Japanese–Evidence from Loanword Accentuation. *The Bulletin of the Phonetic Society of Japan*, *211*, 71–82.

Kubozono, H. (2004). *Tone and Syllable in Kagoshima Japanese*. Kobe Papers in Linguistics. https://doi.org/10.24546/80010004

Kubozono, H. (2007). Tonal change in language contact: Evidence from Kagoshima Japanese. In T. Riad & C. Gussenhoven (Eds.), *Tones and Tunes: Typological Studies in Word and Sentence Prosody* (Vol. 1, pp. 323–352). Mouton de Gruyter. https://doi.org/10.1515/9783110207569.323

Kubozono, H. (2012a). Varieties of pitch accent systems in Japanese. *Lingua*, *122*(13), 1395–1414. https://doi.org/10.1016/j.lingua.2012.08.001

Kubozono, H. (2012b). Word-level vs. Sentence-level prosody in Koshikijima Japanese. *The Linguistic Review*, *29*(1). https://doi.org/10.1515/tlr-2012-0005

Kubozono, H. (2018a). Bilingualism and Accent Changes in Kagoshima Japanese. In H. Kubozono & M. Giriko (Eds.), *Tonal Change and Neutralization* (pp. 279–329). De Gruyter. https://doi.org/10.1515/9783110567502-011

Kubozono, H. (2018b). Postlexical Tonal Neutralizations in Kagoshima Japanese. In H. Kubozono & M. Giriko (Eds.), *Tonal Change and Neutralization* (pp. 27–57). De Gruyter. https://doi.org/10.1515/9783110567502-003

Labrune, L. (2012). *The phonology of Japanese*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199545834.001.0001

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511808814

Ladd, R. (1996). *Intonational Phonology*. Cambridge University Press.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista Di Linguistica 7*, 5–37.

Maekawa, K. (1997). Effects of focus on duration and vowel formant frequency in Japanese. *Proceedings of the International Conference on Spoken Language Processing*.

Maekawa, K., Kikuchi, H., Igarashi, Y., & Venditti, J. (2002). X-JToBI: An extended J-ToBI for spontaneous speech. *7th International Conference on Spoken Language Processing (ICSLP 2002)*, 1545–1548. https://doi.org/10.21437/ICSLP.2002-468

Mann, R., & Repp, B. (1980). Influence of vocalic context on the perception of [ʃ]-[s] distinction: I. Temporal factors. *Perception & Psychophysics*, *28*, 213–229.

Matsuura, T. (2014). *Nagasaki hogen kara mita go oncho no kozo [The Structure of Word Tone in Nagasaki Dialect]*. Hitsuzi Shobo.

McCawley, J. (1968). *The Phonological Component of a Grammar of Japanese*. Mouton.

McCawley, J. (1978). What Is a Tone Language? In *Tone* (pp. 113–131). Elsevier. https://doi.org/10.1016/b978-0-12-267350-4.50009-1

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, *39*(1), 59–67. https://doi.org/10.1016/j.wocn.2010.11.002

Ohala, J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrik, & M. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203). Chicago Linguistic Society.

Ohala, J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language Change* (pp. 173–198). DE GRUYTER MOUTON. https://doi.org/10.1515/9783110853063.173

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–278). Longman.

Ohala, J. (2012). The listener as a source of sound change. An Update. In M.-J. Solé & D. Recasens (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 21–36). John Benjamins.

Ota, I. (2020). *Tonal Variation of Kagoshima Japanese and Its Relevant Factors*. 国際基督教大学. https://doi.org/10.34577/00004630

Ota, I., Nikaido, H., & Utsugi, A. (2016). Tonal variation in Kagoshima Japanese and factors of language change. In M.-H. Côté, R. Knooihuizen, & J. Nerbonne

(Eds.), *The future of dialects: Selected papers from* Methods in Dialectology XV (pp. 389–398). Language Science Press.

Ota, I., & Takano, S. (2014). The Media Influence On Language Change In Japanese Sociolinguistic Contexts. In J. Androutsopoulos (Ed.), *Mediatization and Sociolinguistic Change* (pp. 171–204). DE GRUYTER. https://doi.org/10.1515/9783110346831.171

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. MIT.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Typological Studies in Language* (Vol. 45, p. 137). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie

Pierrehumbert, J., & Beckman, M. (1988). Japanese Tone Structure. In *The Journal of the Association of Teachers of Japanese* (Vol. 23). https://doi.org/10.2307/489057

Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In *Intentions in Communication* (pp. 271–312). The MIT Press. https://doi.org/10.7551/mitpress/3839.003.0016

Pittayaporn, P. (2018). *Tonal Change and Neutralization* (H. Kubozono & M. Giriko, Eds.; pp. 249–278). De Gruyter Mouton. https://doi.org/10.1515/9783110567502-010

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, *13*(4), 455–471. https://doi.org/10.1016/s0095-4470(19)30797-1

Poser, W. J. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese* [MIT]. https://www.degruyter.com/document/doi/10.1515/jjl-1990-0120/html

Pouplier, M. (2007). Tongue Kinematics during Utterances Elicited with the SLIP Technique. *Language and Speech*, *50*(3), 311–341. https://doi.org/10.1177/00238309070500030201

Pouplier, M., & Hardcastle, W. (2005). A Re-Evaluation of the Nature of Speech Errors in Normal and Disordered Speakers. *Phonetica*, *62*(2–4), 227–243. https://doi.org/10.1159/000090100

Pouplier, M., Hoole, P., & Scobbie, J. M. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Laboratory Phonology*, *2*(1), 1–33. https://doi.org/10.1515/labphon.2011.001

Prieto, P. (1998). The scaling of the L values in Spanish downstepping contours. *Journal of Phonetics*, *26*(3), 261–282. https://doi.org/10.1006/jpho.1998.0074

Roseano, P., Mar Vanrell, M. D., & Prieto, P. (2015). Intonational phonology of Friulian and its dialects. In *Intonation in Romance* (1st ed., pp. 101–139). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199685332.003.0004

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /R/ in Montreal French. *Language*, *83*(3), 560–588. https://doi.org/10.1353/lan.2007.0106

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, *11*(2), e1521. https://doi.org/10.1002/wcs.1521

Sibata, T. (1962). On'in [Phonology]. In Kokugogakkai (Ed.), *Hogengaku Gaisetsu (Introduction to Dialectology)* (pp. 137–161). Musashinoshobo.

Slowiaczek, L. M., & Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, *13*(3), 325–341. https://doi.org/10.1016/s0095-4470(19)30763-6

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Stevens, M., & Harrington, J. (2022). Individual variation and the coarticulatory path to sound change: Agent-based modeling of /str/ in English and Italian. *Glossa: A Journal of General Linguistics*, *7*(1). https://doi.org/10.16995/glossa.8869

Takano, S., & Ota, I. (2017). A sociophonetic approach to variation in Japanese pitch realizations: Region, age, gender and stylistic parameters. *Asia-Pacific Language Variation*, *3*(1), 5–40. https://doi.org/10.1075/aplv.3.1.02tak

Talkin, D. (2015). *REAPER: Robust Epoch and Pitch EstimatoR* [Computer software]. Google Inc. https://github.com/google/REAPER

Trubeckoj, N. S. (1958). *Grundzüge der Phonologie (Principles of Phonology)*. Vandenhoeck & Ruprecht. https://books.google.de/books?id=Mw9ZAAAAMAAJ

Uwano, Z. (1999). Classification of Japanese accent systems. In S. Kaji (Ed.), *Proceedings of the Symposium "Cross-Linguistic Studies on Tonal Phenomena, Tonogenesis, Typology, and Related Topics"* (pp. 151–186).

Uwano, Z. (2018). Accentual Neutralization in Japanese Dialects. In H. Kubozono & M. Giriko (Eds.), *Tonal Change and Neutralization* (pp. 129–155). De Gruyter. https://doi.org/10.1515/9783110567502-006

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (Version R package version 2.4.1) [Dataset].

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany (N.Y.) : State university of New York press. http://lib.ugent.be/catalog/rug01:000140813

Vance, T. J. (1995). Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics*, *23*(4), 487–499. https://doi.org/10.1006/jpho.1995.0035

Venditti. (2005). The J_ToBI model of Japanese intonation. *Prosodic Typology: The Phonology of Intonation and Phrasing*. https://doi.org/10.1093/acprof:oso/9780199249633.003.0007

Venditti, J., Maekawa, K., & Beckman, M. E. (2012). Prominence Marking in the Japanese Intonation System. In *The Oxford Handbook of Japanese Linguistics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195307344.013.0017

Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America*, *109*(3), 1144–1156. https://doi.org/10.1121/1.1344156

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*(2), 251–276. https://doi.org/10.1016/s0095-4470(03)00032-9

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman & Hall/CRC. https://doi.org/10.1201/9781315370279

Yu, A. C. L. (2023). The Actuation Problem. *Annual Review of Linguistics*, *9*(1), 215–231. https://doi.org/10.1146/annurev-linguistics-031120-101336

Zellou, G., & Cohn, M. (2024). Apparent-time variation in the use of multiple cues for perception of anticipatory nasal coarticulation in California English. *Glossa: A Journal of General Linguistics*, *9*(1). https://doi.org/10.16995/glossa.10831

Appendix

1. Speakers

speaker	age	group	speaker	age	group
KJ.0003	41	KJ.younger	TJ.0001	33	TJ
KJ.0004	31	KJ.younger	TJ.0002	67	TJ
KJ.0005	33	KJ.younger	TJ.0003	68	TJ
KJ.0006	66	KJ.older	TJ.0004	69	TJ
KJ.0008	75	KJ.older	TJ.0005	71	TJ
KJ.0009	79	KJ.older	TJ.0006	65	TJ
KJ.0010	67	KJ.older	TJ.0007	69	TJ
KJ.0011	62	KJ.older	TJ.0008	47	TJ
KJ.0013	75	KJ.older	TJ.0009	61	TJ
KJ.0014	68	KJ.older	TJ.0010	60	TJ
KJ.0015	19	KJ.younger	TJ.0011	28	TJ
KJ.0016	19	KJ.younger	TJ.0012	18	TJ
KJ.0017	58	KJ.older	TJ.0013	71	TJ
KJ.0018	80	KJ.older	TJ.0014	71	TJ
KJ.0019	65	KJ.older	TJ.0015	20	TJ
KJ.0020	65	KJ.older	TJ.0016	31	TJ
KJ.0021	70	KJ.older	TJ.0017	30	TJ
KJ.0022	71	KJ.older	TJ.0018	70	TJ
KJ.0030	72	KJ.older	TJ.0019	23	TJ
KJ.0031	73	KJ.older	TJ.0020	20	TJ
KJ.0032	59	KJ.older	TJ.0021	23	TJ

KJ.0033	66	KJ.older	TJ.0022	37	TJ
KJ.0034	63	KJ.older	TJ.0023	54	TJ
KJ.0035	18	KJ.younger	TJ.0024	23	TJ
KJ.0036	19	KJ.younger	TJ.0025	34	TJ
KJ.0037	19	KJ.younger	TJ.0026	20	TJ
KJ.0038	19	KJ.younger	TJ.0027	25	TJ
KJ.0039	40	KJ.younger	TJ.0028	20	TJ
KJ.0040	65	KJ.older	TJ.0029	19	TJ
KJ.0041	63	KJ.older	TJ.0030	22	TJ
KJ.0042	42	KJ.younger	TJ.0031	21	TJ
KJ.0043	20	KJ.younger	TJ.0032	22	TJ
KJ.0044	22	KJ.younger	TJ.0033	23	TJ
KJ.0045	25	KJ.younger	TJ.0034	24	TJ
KJ.0046	20	KJ.younger	TJ.0035	54	TJ

2. Stimuli

1.1. Monomoraic words

prompt	**Traditional KJ**	**TJ**	**gloss**	**translation**	**KAN**
ba1	Type B	unaccented	場	place	b a
chi1	Type B	accented	地	land	tS i
chi2	Type A	unaccented	血	blood	tS i
e1	Type B	accented	絵	picture	e
e2	Type A	unaccented	柄	handle	e
go1	Type A	accented	五	five	g o
go2	Type B	ambigue	碁	Go (game)	g o
ha1	Type A	unaccented	葉	leaf	h a
ha2	Type B	accented	歯	tooth	h a
hi1	Type A	unaccented	日	sunshine	C i

hi2	Type B	accented	火	fire	C i
i1	Type A	unaccented	胃	stomach	i
ka1	Type A	unaccented	蚊	mosquito	k a
ka2	Type B	accented	課	department	k a
ki1	Type A	unaccented	気	feeling	k i
ki2	Type B	accented	木	tree	k i
ma1	Type A	ambigue	魔	demon	m a
ma2	Type B	unaccented	間	break	m a
me1	Type B	accented	目	eye	m e
me2	Type B	accented	芽	sprout	m e
mi1	Type A	unaccented	身	oneself	m i
ni1	Type B	accented	荷	load	J i
sa1	Type A	unaccented	差	difference	s a
shi1	Type A	unaccented	詩	poem	S i
shi2	Type A	accented	市	city	S i
wa1	Type B	accented	輪	circle	w a
wa2	Type A	accented	和	peace	w a
ya1	Type A	accented	矢	arrow	j a
yo1	Type A	accented	世	world	j o
yo2	Type B	accented	夜	evening	j o
yu1	Type B	accented	湯	hot water	j u
za1	Type B	ambigue	座	position	z a

1.2. Bimoraic words

prompt	Traditional KJ	TJ	gloss	translation	KAN
ban1	A	unaccented	晩	evening	b a N\
ban2	B	accented	番	order	b a N\
kai1	A	accented	下位	lower rank	k a i

kai2	B	accented	貝	shellfish	k a i
kan1	A	accented	缶	can	k a N\
kan2	A	unaccented	勘	instinct	k a N\
kyuu1	B	accented	九	nine	k_j u:
kyuu2	A	unaccented	急	urgent	k_j u:
tai1	A	accented	タイ	Thailand	t a i
tai2	B	accented	鯛	sea bream	t a i
too1	A	accented	十	ten	t o:
too2	B	accented	塔	tower	t o:
zyuu1	A	accented	銃	gun	dZ u:
zyuu2	B	accented	十	ten	dZ u:

3. Analysis Code

All acoustic data processing, smoothing, normalization, and statistical modeling were implemented in R using a custom analysis pipeline written in an R Markdown script.

- File name: monosyllables.Rmd
- Software environment: R version 4.x, with the following key packages:
  - ggplot2 (Wickham, 2016) for contour plotting
  - tidyverse (Wickham et al., 2019) for data wrangling
  - REAPER (Talkin, 2015) for f_0 extraction
  - mgcv (Wood, 2017) for fitting Generalized Additive Models
  - itsadug (van Rij et al., 2022) for GAMM visualization and pairwise comparisons

The full .Rmd file includes:

- Preprocessing steps (e.g., z-normalization, DCT smoothing, time normalization)
- GAM model definitions for each condition
- Pairwise comparison routines using plot_diff()
- Figure generation code for all contour plots shown in Chapter 4

4. Summary Statistics and Difference plots

Monomoraic, TJ A/U

Formula:

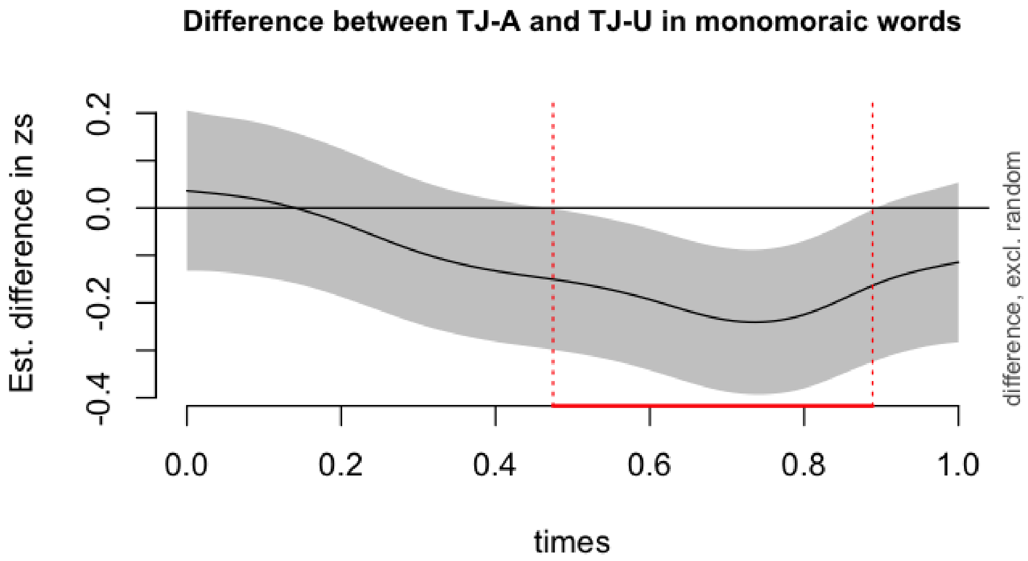
zs ~ TJ + s(times, by = TJ) + s(KAN, bs = "re") + s(speaker,

TJ, bs = "re") + s(times, speaker, bs = "fs", m = 1) + s(times,

KAN, bs = "fs", m = 1) + s(speaker, bs = "re")

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	p-value
(Intercept)	0.03012	0.04849	0.621	0.5344
TJunaccented	-0.11756	0.06960	-1.689	0.0912
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):TJaccented	8.467	8.549	46.234	< 2e-16 ***
s(times):TJunaccented	8.219	8.356	37.846	< 2e-16 ***
s(KAN)	12.939	27.000	0.920	< 2e-16 ***
s(speaker,TJ)	39.695	68.000	29.843	< 2e-16 ***
s(times,speaker)	267.028	314.000	29.843	< 2e-16 ***
s(times,KAN)	163.896	259.000	13.144	< 2e-16 ***
s(speaker)	9.556	34.000	0.544	2.98e-05 ***



Monomoraic, KJ-old A/B

Formula:

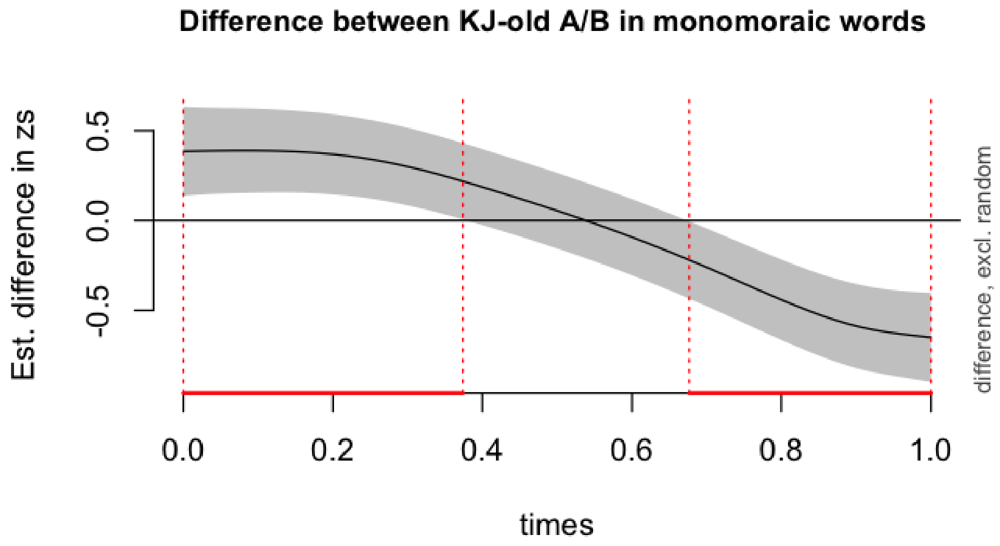
zs ~ KJ + s(times, by = KJ) + s(KAN, bs = "re") + s(speaker,

KJ, bs = "re") + s(times, speaker, bs = "fs", m = 1) + s(times,

KAN, bs = "fs", m = 1)

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	p-value
(Intercept)	-0.009687	0.069715	-0.139	0.889
KJB	0.029226	0.098588	0.296	0.767
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):KJA	7.183	7.392	29.268	< 2e-16 ***
s(times):KJB	6.794	7.031	11.486	< 2e-16 ***
s(KAN)	14.367	30.000	0.919	< 2e-16 ***
s(speaker,KJ)	32.938	38.000	6.596	< 2e-16 ***
s(times,speaker)	144.922	179.000	22.717	< 2e-16 ***
s(times,KAN)	222.770	286.000	17.204	< 2e-16 ***



Monomoraic, KJ-young KJ-A/TJ-U

Formula:

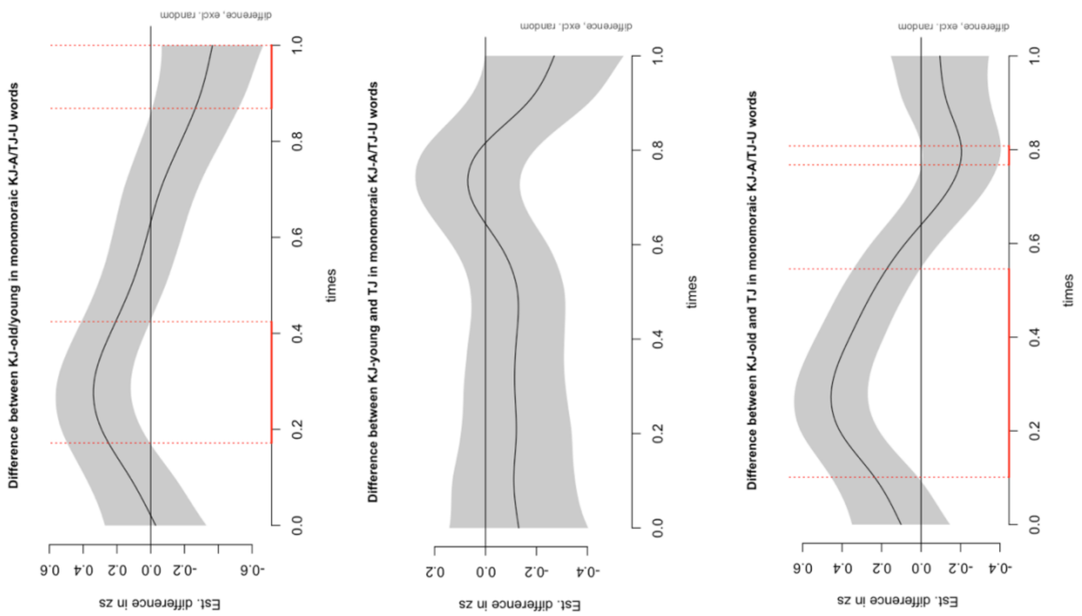
zs ~ dialect2 + s(times, by = dialect2) + s(times, KAN, bs = "fs",

m = 1) + s(times, speaker, bs = "fs", m = 1) + s(KAN, dialect2,

bs = "re")

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	Pr(>|t|)
(Intercept)	0.08769	0.06142	1.428	0.153
dialect2KJ.younger	-0.01760	0.07780	-0.226	0.821
Dialect2TJ	-0.12051	0.06529	-1.846	0.065 .
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):dialect2KJ.older	6.472	6.821	31.576	< 2e-16 ***
s(times):dialect2KJ.younger	6.404	6.765	15.414	< 2e-16 ***
s(times):dialect2TJ	8.089	8.238	36.923	< 2e-16 ***
s(times,KAN)	67.212	89.000	10.082	< 2e-16 ***
s(times,speaker)	530.190	627.000	9.566	< 2e-16 ***
s(KAN,dialect2)	10.781	27.000	0.859	0.00147 **



Monomoraic, KJ-young KJ-U, TJ-A

Formula:

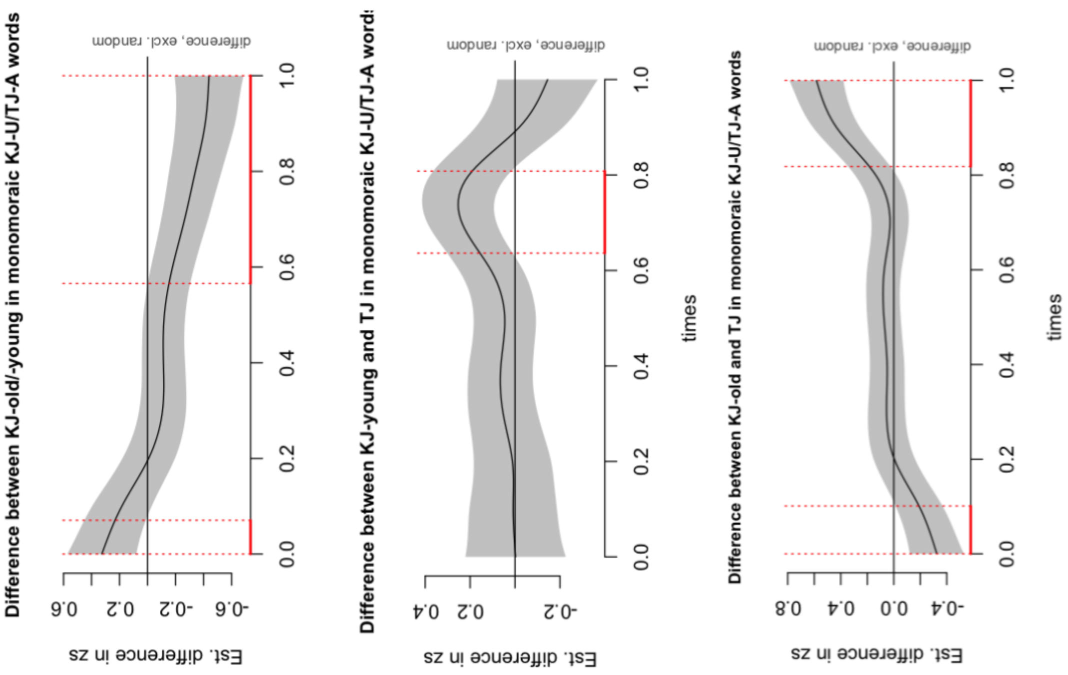
zs ~ dialect2 + s(times, by = dialect2) + s(times, KAN, bs = "fs",

m = 1) + s(times, speaker, bs = "fs", m = 1) + s(KAN, dialect2,

bs = "re")

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	Pr(>|t|)
(Intercept)	0.13869	0.05567	2.491	0.01273 *
dialect2KJ.younger	-0.13365	0.04732	-2.824	0.00474 **
Dialect2TJ	-0.08316	0.04044	-2.056	0.03977 *
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):dialect2KJ.older	6.213	6.596	11.523	< 2e-16 ***
s(times):dialect2KJ.younger	6.865	7.206	16.865	< 2e-16 ***
s(times):dialect2TJ	8.415	8.519	45.195	< 2e-16 ***
s(times,KAN)	87.813	107.000	21.966	< 2e-16 ***
s(times,speaker)	511.177	627.000	7.792	< 2e-16 ***
s(KAN,dialect2)	12.212	33.000	0.789	0.00111 **



Bimoraic, Accented all

Formula:

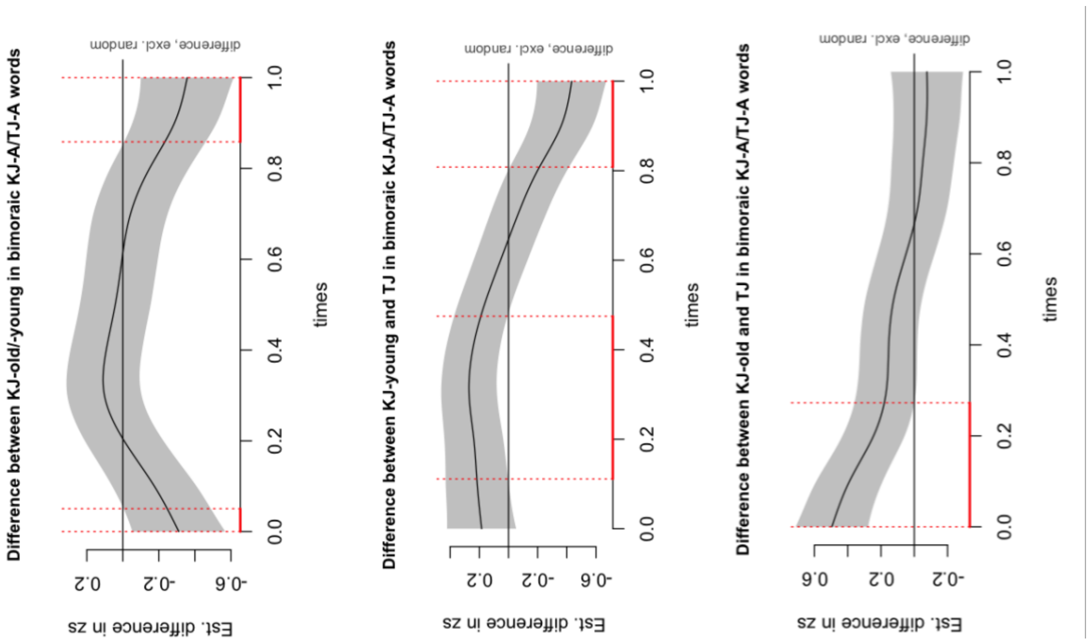
zs ~ dialect2 + s(times, by = dialect2) + s(times, KAN, bs = "fs",

m = 1) + s(times, speaker, bs = "fs", m = 1) + s(KAN, dialect2,

bs = "re")

Summary statistics:

A. parametric coefficients	**Estimate**	***SE***	**t-value**	**Pr(>|t|)**
(Intercept)	-0.05912	0.07265	-0.814	0.4158
dialect2KJ.younger	0.10008	0.08414	1.189	0.2343
Dialect2TJ	0.12737	0.07327	1.738	0.0822 .
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):dialect2KJ.older	6.741	7.139	39.287	< 2e-16 ***
s(times):dialect2KJ.younger	5.236	5.803	27.743	< 2e-16 ***
s(times):dialect2TJ	7.342	7.606	62.440	< 2e-16 ***
s(times,KAN)	98.830	116.000	31.506	< 2e-16 ***
s(times,speaker)	488.120	627.000	5.566	< 2e-16 ***
s(KAN,dialect2)	10.093	22.000	1.906	< 2e-16 ***



Bimoraic, KJ-A/TJ-U

Formula:

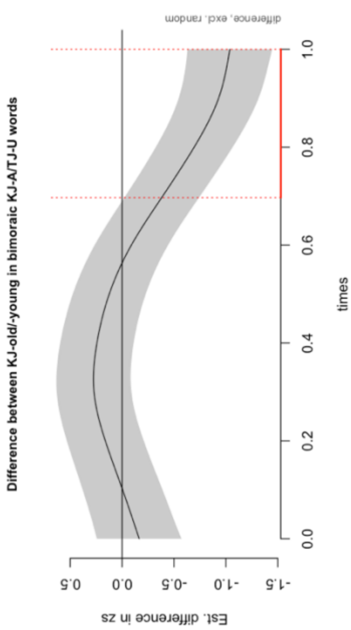
zs ~ dialect2 + s(times, by = dialect2) + s(times, KAN, bs = "fs",

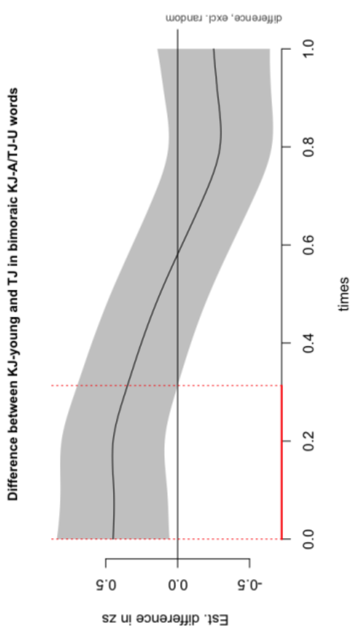
m = 1) + s(times, speaker, bs = "fs", m = 1) + s(KAN, dialect2,

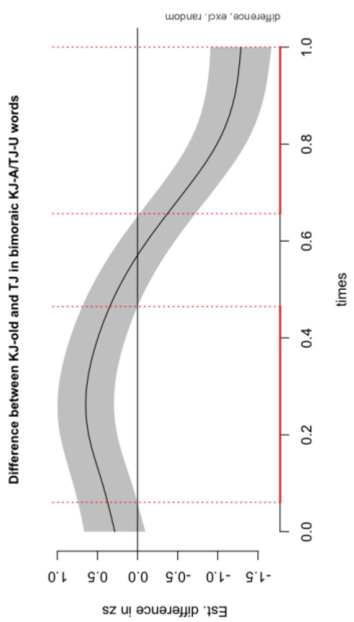
bs = "re")

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	Pr(>|t|)
(Intercept)	-0.1339	0.1775	-0.754	0.451
dialect2KJ.younger	0.2288	0.1691	1.353	0.176
Dialect2TJ	0.1343	0.1657	0.810	0.418
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):dialect2KJ.older	6.148	6.813	20.362	< 2e-16 ***
s(times):dialect2KJ.younger	1.000	1.000	24.815	6.91e-07 ***
s(times):dialect2TJ	5.663	6.301	3.784	0.000796 ***
s(times,KAN)	22.038	26.000	157.681	0.006223 **
s(times,speaker)	383.768	627.000	2.212	< 2e-16 ***
s(KAN,dialect2)	4.190	6.000	14.411	< 2e-16 ***







Bimoraic, KJ-U/TJ-A

Formula:

zs ~ dialect2 + s(times, by = dialect2) + s(times, KAN, bs = "fs",

m = 1) + s(times, speaker, bs = "fs", m = 1) + s(KAN, dialect2,

bs = "re")

Summary statistics:

A. parametric coefficients	Estimate	*SE*	t-value	Pr(>|t|)
(Intercept)	0.09900	0.10499	0.943	0.346
dialect2KJ.younger	-0.07462	0.09057	-0.824	0.410
Dialect2TJ	-0.05014	0.08395	-0.597	0.550
B. smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(times):dialect2KJ.older	3.738	4.188	3.446	0.00682 **
s(times):dialect2KJ.younger	1.000	1.000	47.642	< 2e-16 ***
s(times):dialect2TJ	6.046	6.576	29.466	< 2e-16 ***
s(times,KAN)	29.028	44.000	75.676	< 2e-16 ***
s(times,speaker)	450.735	627.000	3.759	< 2e-16 ***
s(KAN,dialect2)	7.296	12.000	4.971	< 2e-16 ***