A quantitative exploration of the *kobushi* technique by Minami Kizuki

Rina Furusawa[a] and Shigeto Kawahara[b]
[a]*International Christian University and* [b]*Keio University*

Abstract

The *kobushi* singing technique is the one that is deployed in some traditional Japanese singing styles, and Minami Kizuki, a professional singer from the Amami Islands, applies kobushi to pop song musics. Next to nothing is known about the phonetic nature of kobushi, and thus, as a first step toward addressing this gap in the literature, this paper offers the first quantitative, phonetic exploration of kobushi. The current study found that (1) kobushi involves an abrupt F0 rise of about 70 Hz on average, which is implemented within about 30 ms, that (2) some kobushi type seems to be easier to be implemented with a neighboring voiceless obstruent, while the other type seems less compatible with a neighboring obstruent, and that (3) kobushi tends to appear on high musical notes, usually around and above 400 Hz, and that (4) intensity is not a defining characteristic of kobushi.

1 Introduction

This paper reports a quantitative study on the phonetic aspects of the *kobushi* singing technique, as sung by a professional singer Minami Kizuki. The kobushi technique is the one that is observed in some traditional Japanese singing styles, including the one commonly found in the Amami Islands, a part of the Ryukyu Archipelago being located in the southern part of Japan (Figure 1).

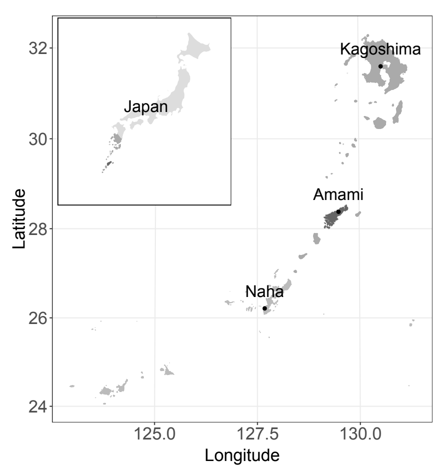

Figure 1: The location of the Amami Islands. Naha is the capital city of the Okinawa prefecture. Amami is a part of the Kagoshima Prefecture.

* First and foremost, we would like to thank Minami Kizuki for sharing her singing voices. We are also grateful to Reed Blaylock, Donna Erickson, Zen Nagatsuka, Yuka Hayashi, John Kingston, Yoichi Kitayama, Seunghun Lee, Jason Shaw and Julián Villegas for their comments on, as well as other kinds of support for, this project. This is a project in progress, and we expect that our understanding of kobushi will continue to grow (and change) in the future. Nevertheless, errors that remain in the current paper are ours.

Ms. Kizuki grew up in the southern part of the Amami Islands, and her original singing style is from the southern area. She applies the traditional singing style from Amami, including that of kobushi, to pop song musics, but very little is known about the phonetic properties of kobushi.[1] A preliminary study reported in Kawahara & Furusawa (2023) suggests that kobushi involves an abrupt rise and fall of F0, as exemplified in Figure 2. In this particular example, the F0 rises from 409 Hz to 482 Hz within about 15 ms.

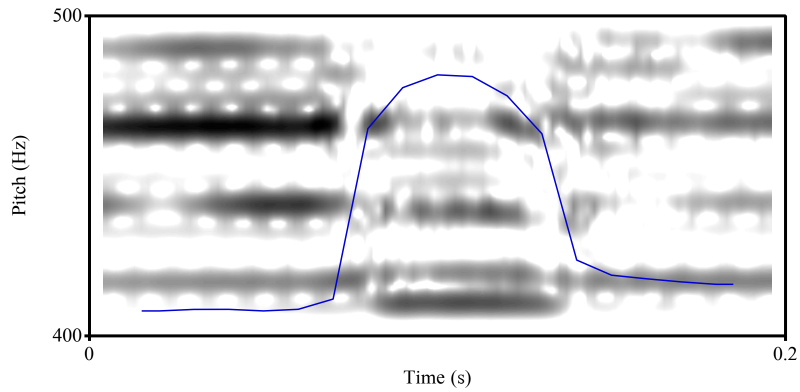

Figure 2: A sample illustration of kobushi, as it appears in the song "Aitsumugi." The rise in this example is from 409 Hz to 482 Hz, implemented within ca. 15 ms.

To the best of our knowledge, there has been no quantitative or phonetic studies of kobushi. As a first step toward addressing this gap in the literature on the phonetics of singing voice (Sundberg, 1987), this paper examines the phonetic properties of kobushi, based on a cappella sounds for two songs—"Aitsumugi" and "Anata-ni aete yokatta"—that were sung by Ms. Kizuki. Since next to nothing is known about the specific properties of kuboshi, our study is necessarily an exploratory one. As such, we refrain from conducting any statistical hypothesis testing (Winter, 2019).

The specific questions about kobushi that are addressed in this paper are listed in (1).

(1) Specific questions addressed in the paper
  a. What is the general magnitude of the kobushi F0 rises?
  b. How fast are the kobushi rises implemented?
  c. How may the kobushi F0 movement interact with the neighboring consonants?
  d. Are there any musical notes that are generally compatible with kobushi?
  e. Does kobushi involve systematic changes in intensity?

The first two questions are descriptively oriented and arose out of our genuine intellectual curiosity. Changes in F0, as observed in normal speech, do not often show changes as radical as 73 Hz, completed within 15-20 ms, as the one in Figure 2 (see Xu & Sun 2004 and references cited therein). Therefore, we were interested in these general properties of kobushi rises—how large can the kobushi rises be and how fast can F0 move? These results may shed light on the question of what the upper limit is on how fast human beings can change their F0 (in the context of singing).

The third question is more theoretically-driven. As background, Kawahara & Furusawa (2023) have identified that there are two kinds of kobushi—one appears at the right edge of a vowel, i.e. near a VC-transition, and the other appears at the left edge of a vowel, i.e. near a CV-transition. An example phrase that contains both types of kobushi is provided in Figure 3, which also happens to contain vibrato toward the end. In Figure 3, the right-aligned kobushi appears at the end of the first vowel [a] during its transition to [k],

[1] There are likely to be various kinds of kobushi, and only some of them will be discussed in the main body of this paper. According to our auditory impressions, it seems that there can be non-trivial differences between those kinds of kobushi that are used in traditional singing and those that are used in pop songs. A systematic comparison of kobushi between the two genres is left for future studies. Since Ms. Kizuki sings in both styles, a within-singer comparison is possible.

whose fall phase is partially made invisible due to its overlap with the closure phase of [k]. The left-aligned kobushi appears near the left edge of the long [e], with its full rise-fall manifestation. Interestingly, Kawahara & Furusawa (2023) found no tokens of kobushi that appeared in the mid section of the vowels.

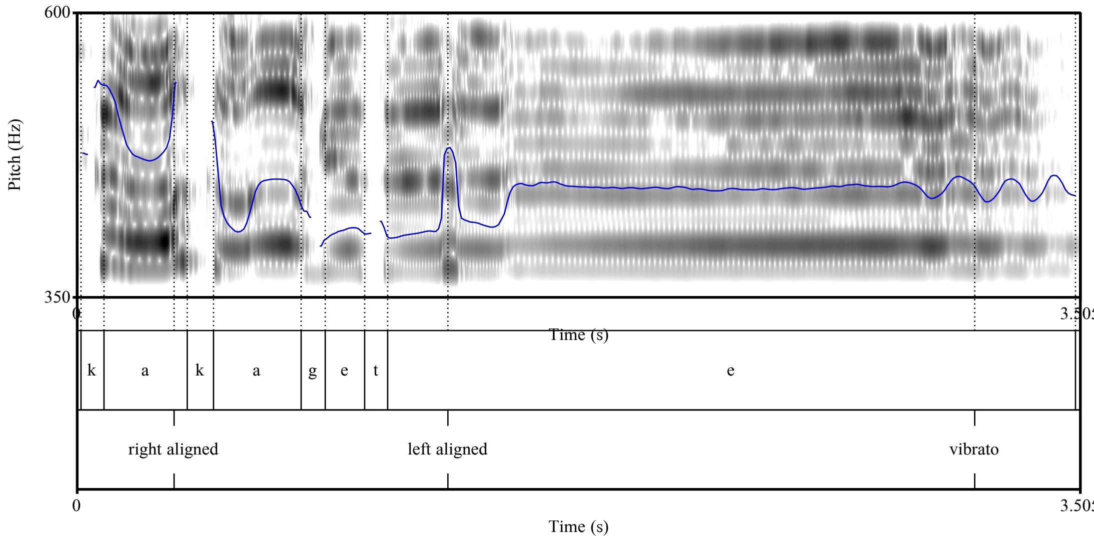

Figure 3: The two types of kobushi, as they appear in the song "Aitsumugi." Vibrato appears at the end of the phrase as well.

Kawahara & Furusawa (2023) have also found, based on the analyses of the lyrics of two songs ("Aitsumugi" and "Anata-ni aete yokatta"), that right-aligned kobushi very often appears together with a following voiceless obstruent (34/58 tokens = 58.6%). It looks as if Ms. Kizuki is taking advantage of the rise of F0 that is caused by voiceless obstruents—in other words, she creates right-aligned kobushi by exaggerating the F0 perturbation effect caused by voiceless obstruents (for the F0 perturbation effect in natural speech, see e.g. Gao & Arai 2019; Kirby 2018; Kingston & Diehl 1994 among many others). On the other hand, Kawahara & Furusawa (2023) found that left-aligned kobushi rarely appears when the preceding consonant is an obstruent, either voiced or voiceless (4/28 tokens = 14.2%). It looks like Ms. Kizuki is avoiding implementing left-aligned kobushi in the region when F0 is perturbed by obstruents, whether the F0 is depressed by voiced ones or raised by voiceless ones.

In our conversation, Ms. Kizuki told us that left-aligned kobushi is what she is really intending to convey as kobushi to the listeners, while right-aligned kobushi may be something different (although the difference between the two may not always be clear-cut to her in some cases). Therefore, it makes sense that left-aligned kobushi appears in vocalic intervals whose F0 is not perturbed. We also note here that left-aligned kobushi appears in intervals where listeners' superior temporal gyrus (STG) regions are most sensitive (Oganian & Chang, 2019), i.e. the interval during which the amplitude rises in CV-transitions. This observation aligns with the feeling that Ms. Kizuki has—left-aligned kobushi is *the* kobushi that she is intending to express. On the other hand, Ms. Kizuki told us that she has the feeling that right-aligned kobushi should not be too conspicuous. With this distinction in mind, to iterate the phonological generalizations about these two types of kobushi (Kawahara & Furusawa, 2023), left-aligned kobushi is dispreferred—or phonologically "marked"—when it co-ocurs with an obstruent, whereas right-aligned kobushi appears very often—or phonologically "unmarked"—with a voiceless obstruent. Given these observations, we wanted to explore how these co-occurrence restrictions may affect the phonetic realizations of kobushi.

The fourth question arose from the impressionistic observation that we made in Kawahara & Furusawa (2023) that kobushi may appear in musical notes that are relatively high. In our conversation, Ms. Kizuki also had that intuition, suggesting that it is easier for her to implement kobushi with high musical notes. We thus wanted to explore whether kobushi indeed appears mostly in high musical notes, and if so, how high these notes should be.

The final question is concerned with the issue of whether kobushi manifests itself in acoustic dimensions other than F0, and in this paper we focus on intensity to address this question. The impressionistic observation by Kawahara & Furusawa (2023) was that some kobushi tokens involve a rise in intensity, whereas other tokens involve a fall in intensity, and therefore, intensity cannot be a defining characteristic of kobushi. We sought to examine this impressionistic observation in a quantitative manner.

2 Methods

As a first step, the first author transcribed all the kobushi-like F0 movements of the two songs—"Aitsumugi" and "Anata-ni aete yokatta"—on the musical scores and checked with Ms. Kizuki, for each token, whether it is left-aligned kobushi or right-aligned kobushi. This process allowed us to make sure that we understand where right-aligned and left-aligned kobushi appear in each song (although not all potential tokens of kobushi are realized in every rendition of the songs). Next, in order to address the questions in (1), we asked Ms. Kizuki to sing the two songs only with [ta]-syllables and only with [ma]-syllables.[2] We chose [ta] and [ma] to explore the interaction between an obstruent on the one hand and right-aligned and left-aligned kobushi on the other.[3] There were 28 tokens of left-aligned kobushi in the [m]-condition, 33 tokens of left-aligned kobushi in the [t]-condition, 42 tokens of right-aligned kobushi in the [m]-condition, and 32 tokens of right-aligned kobushi in the [t]-condition.

While the F0 movement of kobushi often looks like a trapezoid with four gestural landmarks, like the gestural representation that is assumed in many studies in Articulatory Phonology (Browman & Goldstein 1989; Gafos 2002; Shaw & Gafos 2015 and see especially Karlin 2022), we decided to focus on the rise phase of kobushi for two reasons. One is that for right-aligned kobushi, the fall phase is often partially covered and made invisible by a consonantal gesture, as observed in Figure 3. For left-aligned kobushi also, we found cases in which the fall is rather gradual, and therefore it was not easy to pin down the onset of the fall.

We annotated the onset and offset of the rise using Praat (Boersma & Weenink, 1999–): an illustrative example is given in Figure 4, where the label "l_m2" means "[m] that precedes the second left-aligned kuboshi" and "l_k2" means "the second token of left-aligned kobushi". We created a 10 ms analysis window around the annotated kobushi-related boundaries and asked Praat to extract the maximum F0 around the target of kobushi, the time of that maximum F0, the minimum F0 around the onset of kobushi, and the time of that minimum F0. The rise time was defined as the time of the maximum F0 minus the time of the minimum F0. The rise magnitude was calculated as the maximum F0 value minus the minimum F0 value. We also calculated the distance between the onset of the kobushi rise and the offset of the preceding consonant. The last measure was calculated to examine whether the [t] vs. [m] difference affects the implementation of left-aligned kobushi (see §3.3 for details). Consonantal intervals were determined based on the significant weakening of formant energies (primarily, the second formant); [t]'s release burst was included in the consonant intervals, because otherwise, [t]'s burst may artificially lengthen the distance between the consonant offset and the kobushi onset.

[2] As much as we would like to make the sound materials publicly available following the spirit of the open science research practice in phonetics and linguistics (Garellek et al., 2020; Winter, 2019), since Ms. Kizuki is a professional singer, we cannot make the phonetic data available due to copyright issues. Interested readers can listen to the songs on YouTube: `https://www.youtube.com/watch?v=hCRV9iS87ck` and `https://www.youtube.com/watch?v=mA_5lF6OEQw`

[3] We specifically chose [ma] because it is more or less standard to use this syllable for research using reiterant speech (see e.g. Kelso et al. 1985). We did not use [p] for the obstruent condition, because [p] at least in Japanese has distinctive sound symbolic values (Kumagai, 2022), and it would have been difficult—if not impossible—to sing a song only with [pa].

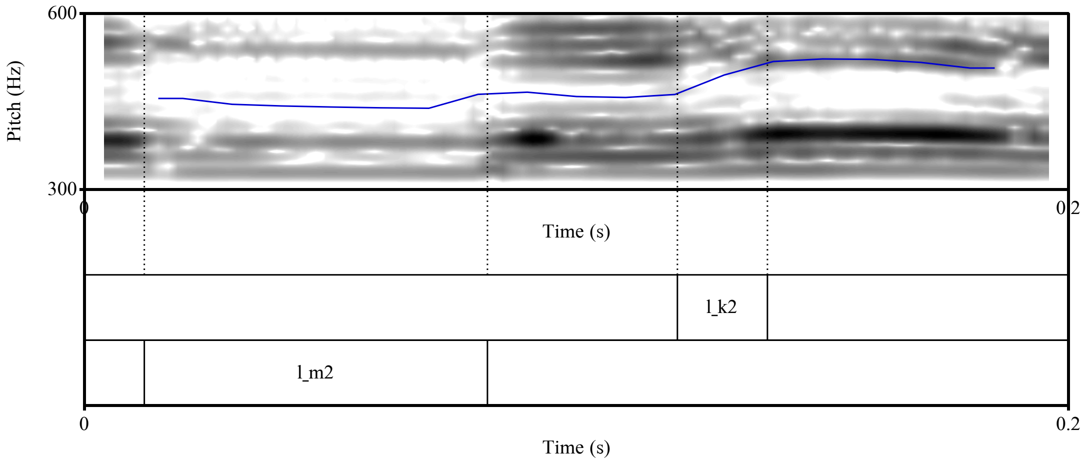

Figure 4: A sample illustration of the annotation convention.

To reiterate, all the analyses presented in this paper are exploratory, and therefore, we did not perform any statistical testing (Winter, 2019).

3 Results

3.1 *The rise magnitude* Figure 5 shows the magnitude of F0 rise of kobushi—how much F0 rises during kobushi—in Hz. The figure is organized in terms of (1) the preceding consonant type ([m] vs. [t]), (2) the kobushi types (left-aligned or right-aligned) and (3) the two songs. The figure consists of violin plots, which show the normalized probability distributions of the values within each facet. Transparent lightblue circles show each token; thick red circles represent the averages within each facet; the error bars around the averages stand for the bootstrap 95% confidence intervals. This and other result figures in the paper were created using `ggplot` package (Wickham, 2016) with R (R Development Core Team, 1993–).

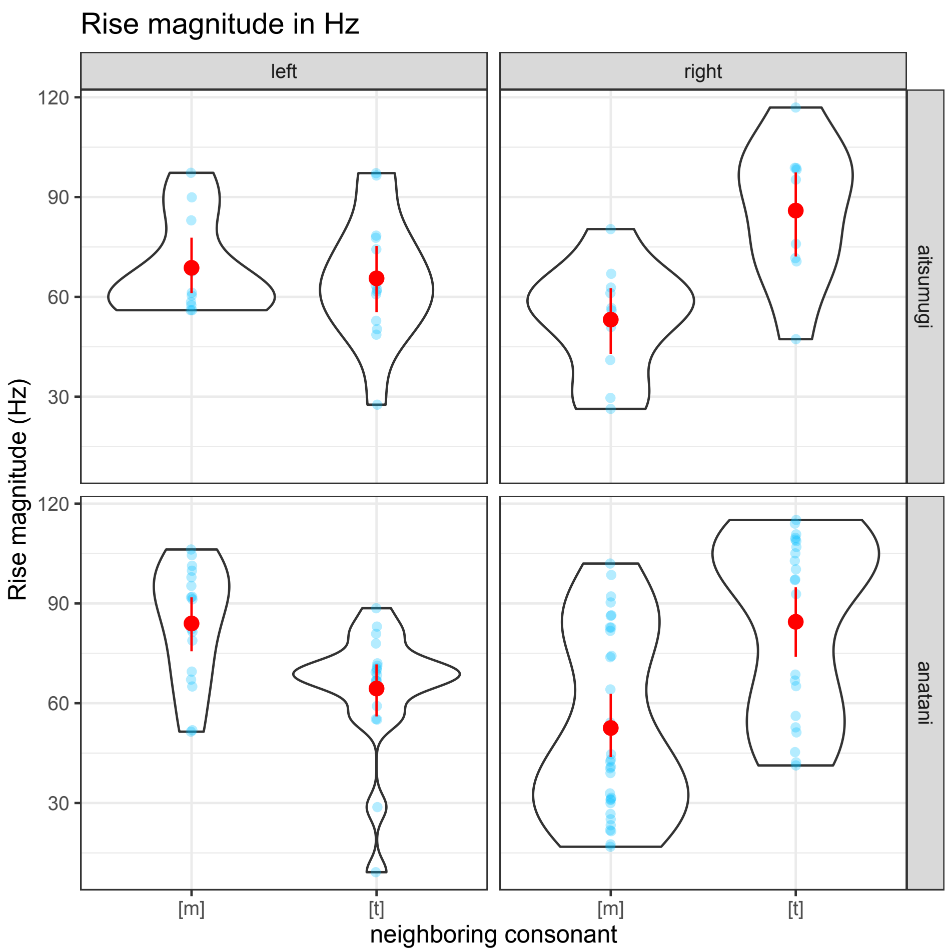

Figure 5: The magnitude of the kobushi rise in Hz, organized in terms of (1) the preceding consonant type ([m] vs. [t]), (2) the kobushi types (left-aligned or right-aligned) and (3) the two songs. The violins show normalized probability distributions of the data within each condition; each transparent circle shows each token (slightly jittered to avoid overlap); the red circles represent the averages within each facet; the error bars stand for the bootstrapped 95% confidence intervals.

We observe first of all that the average rise magnitudes are about or above 50 Hz in each condition, and there are many tokens of rises that are much larger. The grand average is 68.6 Hz (s.d. = 25.3). In terms of semi-tones, the average rise magnitude was 2.46 st (s.d. = 0.81).[4]

We can also identify intriguing patterns in Figure 5 regarding the interaction between the preceding consonant type and the rise magnitude. First, for left-aligned kobushi (left facets), whose co-occurrence with a preceding obstruent is underrepresented in actual songs (Kawahara & Furusawa, 2023), the rise magnitude is smaller after [t] than after [m] (64.9 Hz vs. 78.0 Hz). This tendency is more clearly observed in “Anata-ni aete yokatta” (64.4 Hz vs. 84.0 Hz) than in “Aitsumugi” (65.6 Hz vs. 68.7 Hz), although we are yet unsure of where this difference comes from. Second, those kobushi rise tokens that are smaller than 30 Hz appear only after [t]. These observations seem to be compatible with a phonological restriction that Kawahara & Furusawa (2023) have identified: left-aligned kobushi is dispreferred—or phonologically marked—after obstruent consonants. This markedness effect is likely due to the fact that F0 is perturbed after obstruents (e.g. Gao & Arai 2019; Kirby 2018; Kingston & Diehl 1994): in order to convey the existence of left-aligned kobushi, those vowels whose F0 is perturbed may not be optimal.

On the other hand, the right-aligned kobushi shows the opposite pattern—right-aligned kobushi, which

[4] 1 semi-tone = $12 \times \log_2 \frac{max}{min}$

is likely to make use of F0 perturbation by voiceless obstruents, shows higher rises before [t] than before [m] (84.9 Hz vs. 52.7 Hz). In other words, it is compatible with the idea that Ms. Kizuki is making use of F0 perturbation to create right-aligned kobushi, i.e. right-aligned kobushi is created by exaggerating the rise of F0 that is caused by a voiceless obstruent. (For right-aligned kobushi, there seem to be no substantial inter-song differences: "Aitsumugi" = 53.2 Hz vs. 85.9 Hz, "Anata-ni aete yokatta" = 52.6 Hz vs. 84.5 Hz).

This general conclusion may support the view that F0 perturbation is, at least to some extent, under the speaker's control rather than purely automatic/physiological (see e.g. Gao & Arai 2019; Kingston & Diehl 1994; Luo 2018; see also Kirby 2018 and Kirby & Ladd 2016 for discussion). If F0 perturbation were not controllable by speakers, Ms. Kizuki would not have been able to deploy that mechanism to create right-aligned kobushi in the first place. This conclusion is admittedly based on data from a single speaker, but it nevertheless shows that F0 perturbation is controllable by an individual.

3.2 *The speed in semi-tone* Figure 6 shows the speed of the kobushi rise in semi-tone per second. We use semi-tone rather than Hz to compare the current results with those of Xu & Sun (2004), who examined the maximum speed of pitch change with native speakers of Chinese and English. We acknowledge that what we are examining—a particular singing style by a professional singer—is quite different from what Xu & Sun (2004) have investigated—speech production patterns elicited by presenting synthesized stimuli (see also Sundberg 1979 for a comparison between singers and non-singers). Therefore, it is probably not fair to treat the two sets of data on the same footing, but we nevertheless find this comparison interesting and potentially informative.

In Figure 6, we observe that the average speed of Ms. Kizuki's kobushi is well above 50 st/sec in every condition. Although Xu & Sun (2004) found interesting effects of gender and language, abstracting away from these influences, the maximum speeds that were found in their study were around 40 st/sec, and what we are observing in Figure 6 is that Ms. Kizuki is going well above those values (the grand average is 84.0 st/sec, s.d. = 38.6).

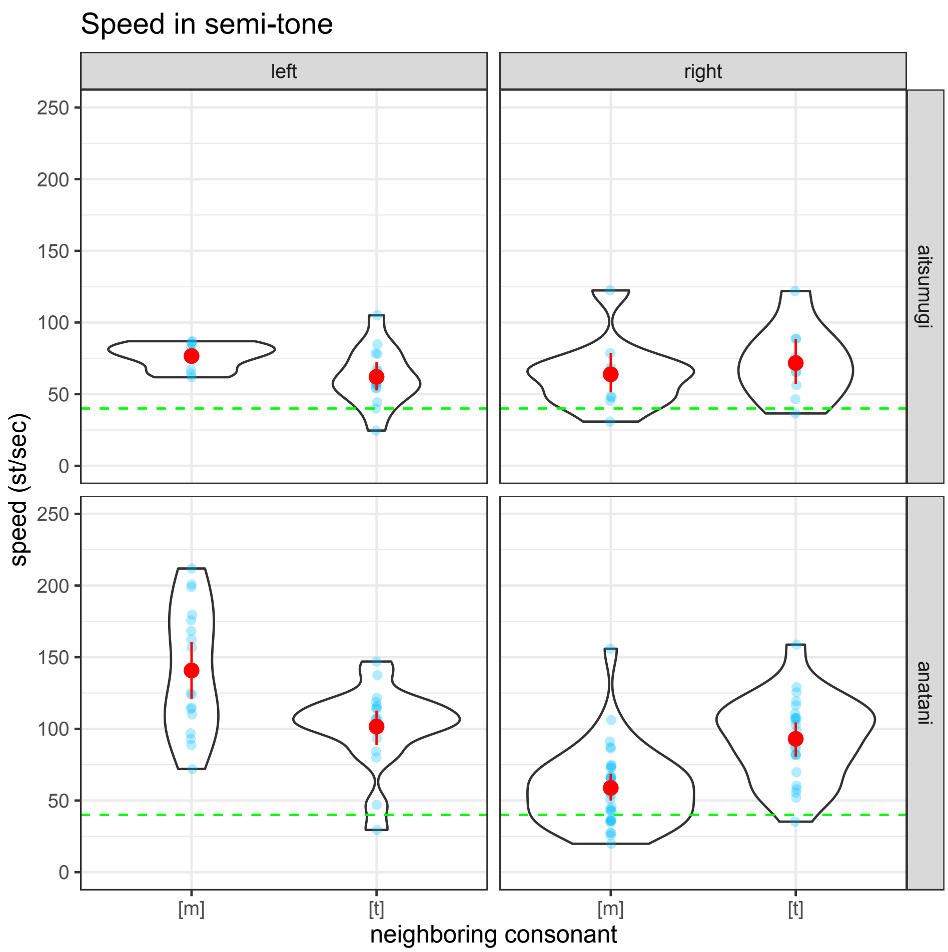

Figure 6: The speed of the kobushi rise in semi-tone per second, shown by (1) the preceding consonant type, (2) the kobushi types and (3) the two songs. A horizontal line is drawn at 40 st/sec in each facet, which is around the average maximum speed of pitch changes observed in the experiment by Xu & Sun (2004).

Now moving on to the interaction between the kobushi speed and the preceding consonant type, we generally observe patterns that are similar to the kobushi rise magnitude (Figure 5). We acknowledge at this point that since speed is defined as rise magnitude divided by rise time, there is a correlation between rise magnitude and rise speed in our dataset ($r = 0.67$), and thus, an argument based on speed should not be considered to be entirely independent of the argument based on rise magnitude. With this said, left-aligned kobushi, which phonologically disfavors a preceding [t], is slower after [t] than after [m] (84.8 st/sec vs. 116.0 st/sec). Right-aligned kobushi, which arguably makes use of F0 perturbation by voiceless obstruents, is faster before [t] than before [m] (87 st/sec vs. 60 st/sec). These observations are again compatible with the idea that left-aligned kobushi is marked with a preceding obstruent, while right-aligned kobushi can make active use of F0 perturbation by voiceless consonants.

We also found that these differences are more prominent in "Anata-ni aete yokatta" than in "Aitsumugi." We show this in a table format in Table 1. Moreover, from the inspection of the length of the error bars as well as the shapes of violins in Figure 6, it looks to be the case that "Anata-ni aete yokatta" shows higher variability than "Aitsumugi": the difference in standard deviation is 42.7 vs. 20.8. We have no clear explanations regarding where this difference comes from, except for noting that "Aitsumugi" is Ms. Kizuki's debut song; therefore she may be more used to singing this song than "Anata-ni aete yokatta"[5] and thus articulation of kobushi in this song may be more robust against potential segmental effects.

[5] She told us that she can sing "Aitsumugi" even when she is sick.

Table 1: The differences in the speed of kobushi in detail.

	kind	seg	speed (st/sec)	difference
Aitsumugi	left	[m]	76.6	14.5
	left	[t]	62.1	
	right	[m]	63.8	
	right	[t]	71.7	7.9
Anata-ni	left	[m]	141.0	39
	left	[t]	102.0	
	right	[m]	58.8	
	right	[t]	93.0	34.2

3.3 *The distance between the consonant offset and the kobushi onset* A picture that is emerging from the preceding discussion, which was also identified in the previous impressionistic study (Kawahara & Furusawa, 2023), is that there is a sense in which left-aligned kobushi is incompatible with a preceding voiceless obstruent. As we examined the a cappella sounds sung by Ms. Kizuki, we have noticed that for the [ma]-rendition of the songs, there are tokens in which the onset of the kobushi rise starts almost at the same time as—or even earlier than—the [m]'s offset. No such tokens were observed for the [ta]-rendition of the songs. This difference is illustrated in Figure 7.

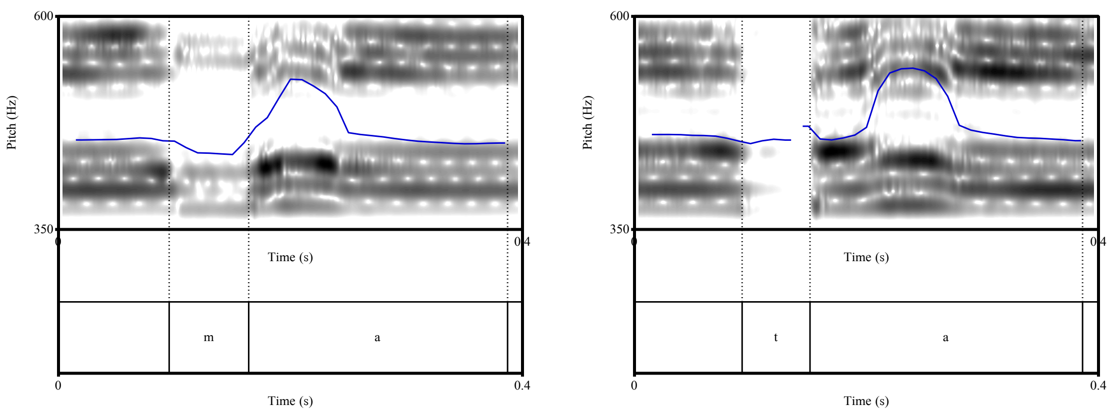

Figure 7: Different coordination patterns between the preceding consonantal intervals and the left-aligned kobushi gestures.

To quantitatively examine this observation, we calculated the duration of the interval between the consonant offset and the left-aligned kobushi onset, the results of which appear in Figure 8.[6] We can see that the intervals are generally longer after [t] than after [m] (for "Aitsumugi": 48.5 ms vs. 32.2 ms; for "Anata-ni aete yokatta": 91.7 ms vs. 72 ms). While there is a substantial overlap between the two segmental conditions, the tendency is in the expected direction. This observation accords well with the postulate that there is a sense in which left-aligned kobushi with a preceding obstruent is marked. We finally note that the intervals are generally longer for "Anata-ni aete yokatta" than for "Aitsumugi," although we do not have a good explanation for this difference.

[6] There was one token for the [t]-condition, in which the interval had a very large negative value, i.e., a token in which the kobushi rise starts well before the offset of [t]. For this particular token, it is likely that Ms. Kizuki "mispronounced" the left-aligned kobushi as a right-aligned one, as the kobushi rise started in the preceding vowel. Since this was a clear outlier, we excluded that token from the analysis.

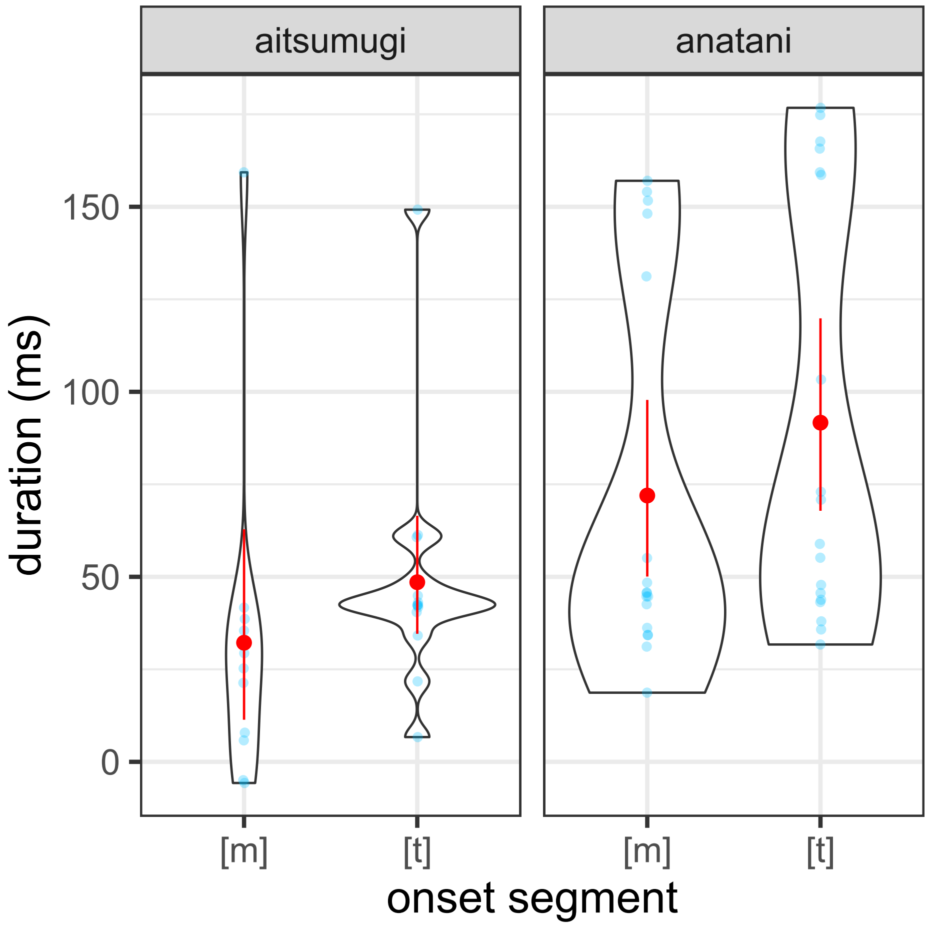

Figure 8: The duration of the intervals between the consonantal offset and the left-aligned kobushi onset.

3.4 *The rise time* Next we explored the temporal dimension—how quickly the kobushi rises are implemented in ms, the results of which appear in Figure 9. Overall, kobushi is implemented on average within 32 ms (s.d. = 10.5), and there are tokens that are completed within less than 20 ms. These are surprisingly quick, given that the excursion time to implement rise in normal speech is reported to be about 130 ms in elicited speech (Xu & Sun, 2004) (again bearing in mind that the two sets of data are not directly comparable).

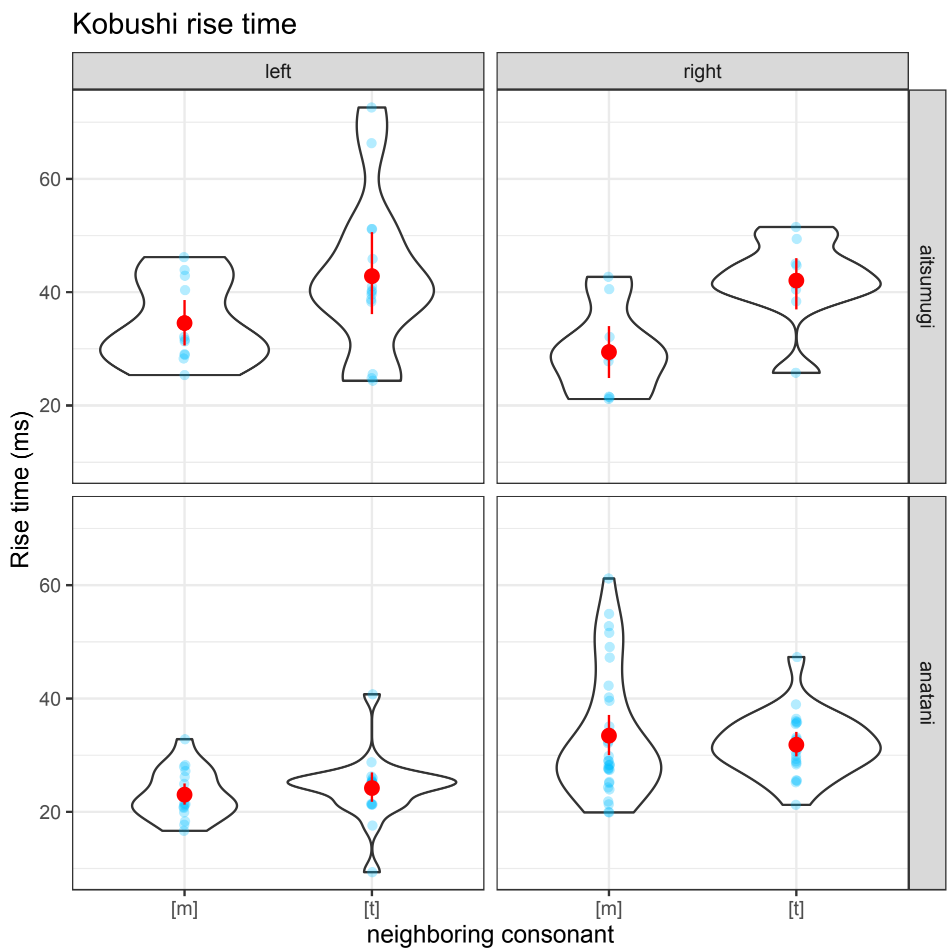

Figure 9: The rise time of kobushi in ms.

From the discussion so far, one might expect that left-aligned kobushi, which disfavors a preceding [t], takes longer after [t] than after [m]. However, this is true only for "Aitsumugi" (42.8 ms vs. 34.6 ms) but not for "Anata-ni aete yokatta" (24.2 ms vs. 23.0 ms). One may also expect that right-aligned kobushi, which arguably depends on [t]'s F0 perturbation, takes shorter time before [t] than before [m], but this was not the case, at least not so clearly ("Aitsumugi": 42.1 ms vs. 29.4 ms; "Anata-ni aete yokatta": 31.8 ms vs. 33.4 ms). It thus seems to be the case that the interaction between kobushi and the neighboring consonant does not manifest itself in the temporal domain.

3.5 *Does kobushi occur in high musical notes?* The preceding discussion revolved around one theoretical issue—given that there seem to be interesting co-occurrence restrictions between the two types of kobushi and preceding consonant types, how do they manifest themselves in the phonetic implementation of kobushi? In addition to this issue, we also addressed another question—does kubushi occur in relatively high musical notes? Kawahara & Furusawa (2023), based on their impressionistic description, noted that kobushi may appear in musical notes that are relatively high, and in our conversation, Ms. Kizuki agreed that that is probably the case, suggesting that she feels more comfortable implementing kobushi when the musical tones are already high.

Figure 10 is a histogram showing the distribution of the F0 values at the onset of kobushi, which represent musical heights at points where kobushi appears. We observe that for both types of kobushi and for both songs, most kobushi appears above 400 Hz (G and G-sharp), and at least around 350 Hz (around A), which confirms our previous observation that kobushi appears on relatively high musical notes. Zen Nagatsuka (p.c. 2023/01/10) pointed out that these values—350 Hz and 400 Hz—may have to do with a “breakpoint”, a point above which it is easier to implement the F0 targets with thin vocal fold contact, or a falsetto voice.

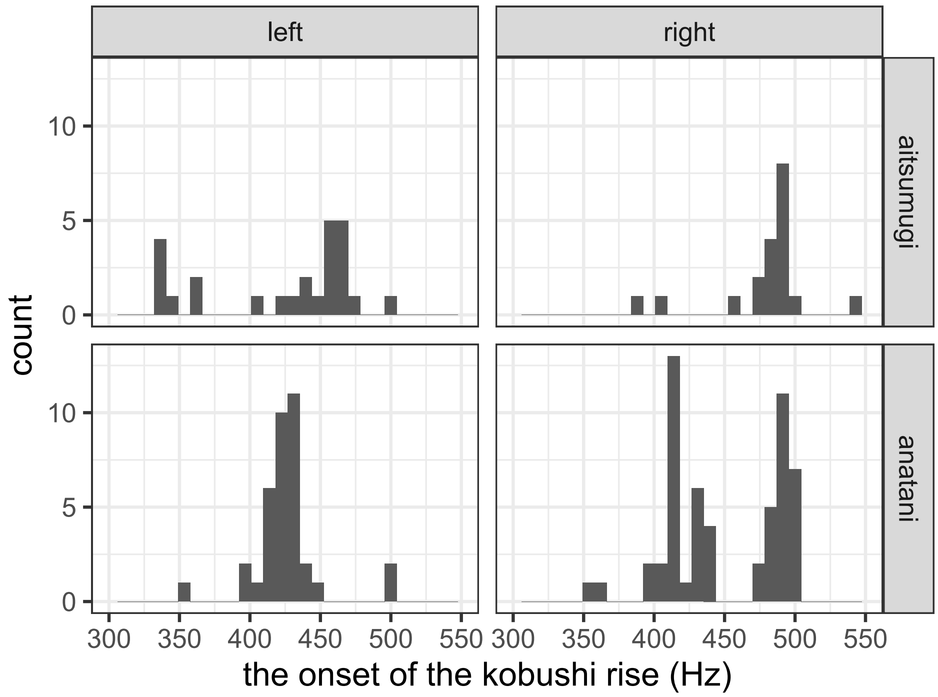

Figure 10: The distribution of the F0 values at the onset of kobushi, representing the musical heights at points where kobushi appears.

3.6 *Does kobushi involve any systematic changes in intensity?* So far our discussion has more or less focused on one acoustic dimension—F0—because from our experiences with kobushi so far, it seems to have been the case that the best definition of kobushi can be made in terms of F0 (Kawahara & Furusawa, 2023). One question that naturally arises, however, is whether other acoustic dimensions systematically vary with kobushi. As a step toward addressing this question, we examined how intensity changes with kobushi. The impressionistic observation by Kawahara & Furusawa (2023) suggested that some tokens of kobushi involve an intensity rise whereas others involve an intensity fall, although this observation was based on a few specific tokens (Figure 11).

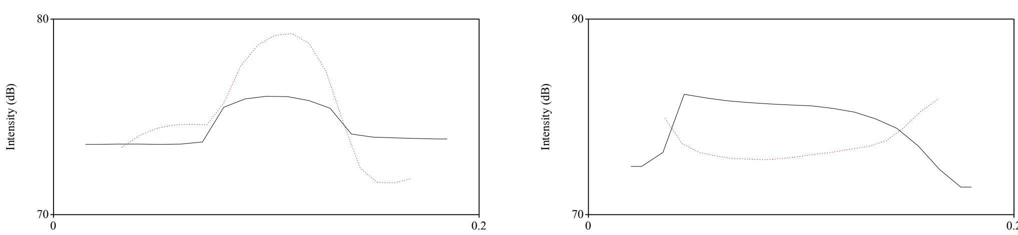

Figure 11: A token in which intensity rises together with F0 (left) and a token in which intensity falls during kobushi (right). Intensity contours are shown with a dotted line; the F0 movement with a thick line. The token on the right is a rendition of kobushi that Ms. Kizuki produced with some exaggeration.

To quantitatively assess this observation, intensity measures in dB were extracted at the onset of kobushi and at the target of kobushi by creating a 10 ms analysis window. Average measures within in each window were calculated, and the intensity at the kobushi onset was subtracted from the intensity at the kobushi target. Positive values thus indicate that intensity rises together with kobushi, whereas negative values indicate that intensity falls with kobushi. The results are shown in the form of histograms in Figure 12.

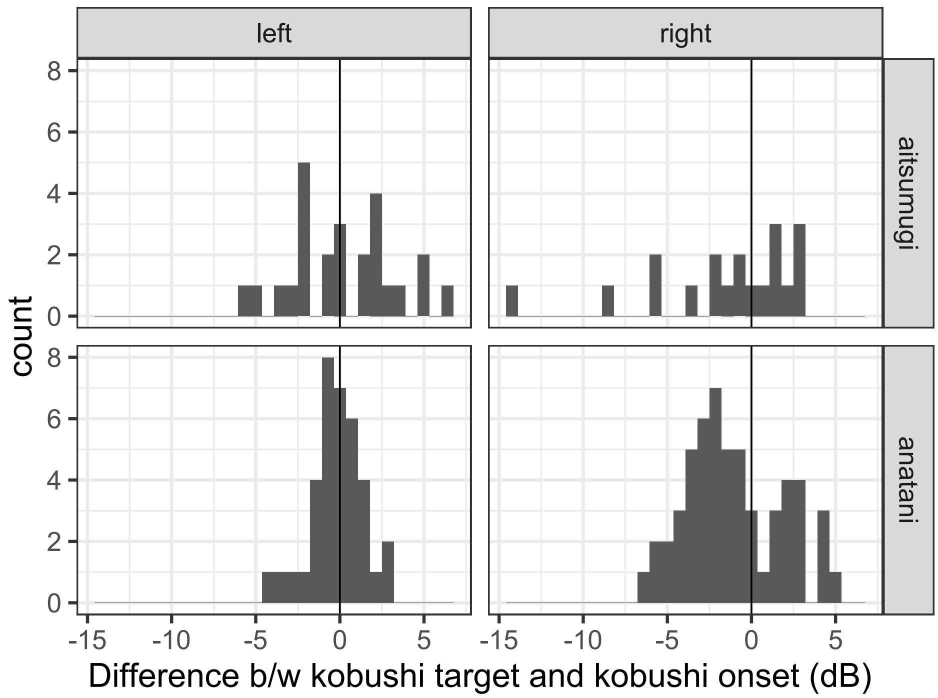

Figure 12: The difference in intensity between the kobushi target and the kobushi onset in dB. Positive values indicate an intensity rise whereas negative values indicate a fall. A vertical line is drawn at $x = 0$ in each facet.

We observe that the distributions are more or less random i.e., some tokens are positive whereas other tokens are negative, although there is a slight skew toward negative values for right-aligned kobushi in "Anata-ni aete yokatta." For completeness sake, Table 2 shows the average in each specific condition, which does not seem to show any systematic patterns. For now it seems safe to conclude that intensity is not a defining characteristic of kobushi.

Table 2: The differences in intensity in detail.

	kind	seg	difference (dB)
Aitsumugi	left	[m]	0.10
	left	[t]	0.20
	right	[m]	-0.71
	right	[t]	-2.56
Anata-ni	left	[m]	-0.36
	left	[t]	-0.08
	right	[m]	-1.92
	right	[t]	0.84

4 Discussion and conclusion

The current exploration has been more or less a descriptive one: we wanted to explore how kobushi is phonetically implemented in a quantitative fashion. The current study has shown that kobushi involves a rise that is about 70 Hz in magnitude, which is implemented within (or less than in) 30 ms. The speed in semi-tone is about 84 st/sec. These values are much higher than what is observed in normal speech (Xu & Sun, 2004), although we acknowledge that it is not too surprising that a professional singer can implement F0 changes that that are faster and larger in magnitude, compared to those observed in normal speech by non-singers (Sundberg, 1979). We note, however, that Xu & Sun (2004) refer to "physiological limitation" and "biological limits" in several places of their paper as they motivate their experiment and interpret their results. Our results generally show that Ms. Kizuki goes well above those purported physiological limits, although we fully admit that not every speaker can produce kobushi, i.e., Ms. Kizuki is special.

Another focus of interest was whether the phonetic implementation patterns of kobushi interact with a neighboring consonant. In particular, we have observed suggestive evidence that right-aligned kobushi seems to be easier to implement with [t], whereas left-aligned kobushi seems easier to implement with [m]. These observations were generally compatible with the distribution of kobushi in the actual songs, as identified by Kawahara & Furusawa (2023).

While we find the current observations to be interesting and have some descriptive values, one issue that remained entirely unresolved is the difference between the two songs. We were unable to offer any reasonable explanation, because we only have data from two songs. We hope to expand the scope of our analyses by examining kobushi tokens as they appear in other songs. Another clear limitation of this study is that we focused on F0 although we also briefly touched upon intensity—however, kobushi can be correlated with other phonetic parameters such as breathiness (Kawahara & Furusawa, 2023), which needs to be studied more systematically in future exploration. More generally speaking, we believe that the phonetics of singing voice (Sundberg, 1987) is an understudied topic, and we would like to continue our effort in further pursuing this research to unveil the phonetic nature of kobushi.

References

Boersma, Paul & David Weenink (1999–). Praat: Doing phonetics by computer. Software.

Browman, Catherine & Louis Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.

Gafos, Adamantios (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20, 269–337.

Gao, Jiayin & Takayuki Arai (2019). Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics* 77, p. 100932.

Garellek, Marc, Matthew Gordon, James Kirby, Wai-Sum Lee, Alexis Michaud, Christine Mooshammer, Oliver Niebuhr, Daniel Recasens, Timo B. Roettger, Adrian Simpson & Kristine Yu (2020). Toward open data policies in phonetics: What we can gain and how we can avoid pitfalls. *Journal of Speech Sciences* 9:1.

Karlin, R. (2022). Expanding the gestural model of lexical tone: Evidence from two dialects of Serbian. *Laboratory Phonology* 13:1, 1–43.

Kawahara, Shigeto & Rina Furusawa (2023). The phonetics and phonology of *kobushi* by Minami Kizuki: A first exploration [in Japanese]. *REPORTS of the Keio Institute of Cultural and Linguistic Studies* 54.

Kelso, J. A. Scott, Eric V. Vatikiotis-Bateson, Saltzman E.L. & B. Kay (1985). A qualitative dynamic analysis of reiterant speech production: phrase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America* 77, 266–280.

Kingston, John & Randy Diehl (1994). Phonetic knowledge. *Language* 70, 419–454.

Kirby, James (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics* 71, 326–354.

Kirby, James & D Robert Ladd (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *Journal of the Acoustical Society of America* 140:1, 2400–2411.

Kumagai, Gakuji (2022). What's in a Japanese *kawaii* 'cute' name? A linguistic perspective. *Frontiers in Psychology* `https://doi.org/10.3389/fpsyg.2022.104`.

Luo, Q. (2018). *Consonantal effect on F0 in tonal languages*. Doctoral dissertation, Michigan State University.

Oganian, Y & E. Chang (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances* 5, p. eaay6279.
R Development Core Team (1993–). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
Shaw, Jason & Adamantios Gafos (2015). Stochastic time models of syllable structure. *PLOS ONE* 10:5.
Sundberg, Johan (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71–79.
Sundberg, Johan (1987). *The science of the singing voice*. Northern Illinois University Press.
Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
Winter, Bodo (2019). *Statistics for Linguists*. Taylor & Francis Ltd, New York.
Xu, Y. & X. Sun (2004). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111, 1399–1413.