

ARTICLE

# Cue-specificity of contrastive hyperarticulation: evidence from the voicing contrast in Japanese

Shinichiro Sano<sup>1</sup> and Céleste Guillemot<sup>2</sup>

<sup>1</sup>Faculty of Business and Commerce, Keio University, Yokohama, Japan

<sup>2</sup>Faculty of Sciences and Engineering, Hosei University, Tokyo, Japan

**Corresponding author:** Shinichiro Sano; Email: [shinichirosano@gmail.com](mailto:shinichirosano@gmail.com)

**Received:** 28 February 2022; **Revised:** 1 October 2023; **Accepted:** 18 June 2024

**Keywords:** hyperarticulation; contrast; minimal pair; VOT; Corpus of Spontaneous Japanese

## Abstract

In natural speech, phonetic cues that distinguish lexical items can be hyperarticulated when there is a minimal pair competitor, a process known as contrastive hyperarticulation. Building upon prior work, this article examines the cue-specific nature of contrastive hyperarticulation in Japanese, focussing on stop voice onset time (VOT) using a speech corpus. We confirmed that the existence of a voicing minimal pair competitor in the lexicon triggers hyperarticulation of VOT duration on the target segment (shorter for voiced stops and longer for voiceless stops), while other contrasts (singleton vs. geminate) did not. The results also suggest that contrastive hyperarticulation (a) is more compatible with casual speech than slow/clear speech, (b) is sensitive to position in a word (greater in word-initial position than in non-initial position) and (c) applies to a greater degree in Japanese than in English due to properties of stops. This provides further evidence that the phonetic specificity of contrastive hyperarticulation is cross-linguistically relevant.

## Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Phonetic properties of Japanese	4
1.2. Goals and research agenda	5
<b>2. Methods</b>	<b>6</b>
2.1. The corpus and data collection	6
2.2. Factors in statistical analysis	6
2.3. Model building	8
<b>3. Results</b>	<b>9</b>
<b>4. Discussion</b>	<b>11</b>
4.1. Degree of contrastive hyperarticulation	12
4.2. Other issues	13
<b>5. Conclusion</b>	<b>14</b>

## 1. Introduction

Mathematical theories such as information theory (Shannon 1948) and Bayesian inference (Bayes 1763; Laplace [1820] 1886) offer a means of quantifying the notion of communicative efficiency under the assumption that language is an effective system of message transfer. Recent linguistic research has

provided increasing evidence that information-theoretic measures such as predictability, surprisal and informativity/entropy play a significant role in accounting for patterns in linguistic behaviour and the course of linguistic change. One of the underlying ideas of this research program is that linguistic signals differ in amount and quality, which causes a tendency for contextually predictable linguistic units to be reduced. Conversely, signals that increase the probability of successfully conveying the intended message may be enhanced in order to achieve robust information transmission. Prior research has investigated the role of information-theoretical and probabilistic measures in linguistic behaviour (see Jaeger & Buz 2017 for a comprehensive review). For example, findings establish links with morphological contraction (Frank & Jaeger 2008; Bresnan & Spencer 2012), omission of morphemes (e.g., case marker omission; Fry 2001; Lee 2006; Kurumada & Jaeger 2015; Norcliffe & Jaeger 2015), function words (e.g., *that*; Jaeger 2010; Jaeger & Grimshaw 2013) or referring expressions (e.g., pronouns; Tily & Piantadosi 2009).

This kind of approach is especially flourishing in phonology in the framework of Message-Oriented Phonology (Hall *et al.* 2016, 2018; henceforth MOP), which applies the basic concepts of information theory and Bayesian inference to phonological research. The idea that constitutes the foundation of MOP is that context-specific differences in the amount of information that phonological units (segments, syllables, etc.) carry about messages influence the extent to which general articulatory and perceptual biases affect their realisation (Hall *et al.* 2016). In this way, more informative pieces of the signal will be enhanced, while less informative pieces will be reduced. When Hall *et al.* (2016) first introduced MOP, they demonstrated the concept based on final obstruent devoicing, a cross-linguistically common phenomenon, focussing on the patterns in English and German. They claimed that there is a direct correlation between the rate of devoicing in word-final obstruents and the number of minimal pairs (lexical competitors). Under the assumptions of MOP, final voiced obstruents are prone to reduction in languages such as German, where only a few minimal pairs exist, because word-final positions are ‘weak’ in terms of the amount of information they can convey, and maintaining voicing word-finally is physiologically costly. On the other hand, when the voicing distinction in word-final position allows distinguishing many lexical competitors, as in English, voicing will be preserved to maintain the contrast.

Probability effects at different levels provide supporting evidence for MOP: segments (Seyfarth 2014; Cohen Priva 2015; Turnbull 2018; with a focus on consonants in Kirov & Wilson 2012; Schertz 2013; Seyfarth *et al.* 2016; Nelson & Wedel 2017; Chodroff & Wilson 2018; Sano 2018a; Wedel *et al.* 2018; and on vowels in Aylett & Turk 2004; Hume & Bromberg 2005; Shaw & Kawahara 2017; Wedel *et al.* 2018), phonological patterns and processes (Hall *et al.* 2018; Kawahara & Lee 2018; Wedel *et al.* 2019), variation and sound change (Wedel *et al.* 2013a,b; Bowerman & Babinsky 2018) or morpheme/word duration (Bell *et al.* 2003, 2009; Hashimoto 2021), among others.

Within MOP-based research, close attention has been paid to the linguistic phenomenon referred to as contrastive hyperarticulation (Wedel *et al.* 2018). When a phonetic cue contributes to phonetically distinguishing a word from its lexical competitors, that cue tends to be hyperarticulated. That is, the distinctive feature of a sublexical unit that is relevant to a specific contrast is enhanced (e.g., longer duration, greater distance between vowels; Wedel *et al.* 2018) to make the distance in phonetic properties greater. This process is active not only within a single conversation, but also based on the existence of competitors in the lexicon as a whole (as also shown in Baese-Berk & Goldrick 2009).

Contrastive hyperarticulation must be distinguished from slow/clear-speech hyperarticulation. Although the two types seem to have similar purposes (i.e., better comprehension/identification of lexical items; Payton *et al.* 1994; Bradlow *et al.* 1996), they can have opposite effects on phonetic implementation. If we take the example of voice onset time (VOT; Lisker & Abramson 1964), the interval of time between the release of the stop closure and the onset of voicing of the following vowel, the logical consequence of slow speech hyperarticulation is that, together with the enhancement of segment duration, VOT for voiceless stops should be longer. Contrastive hyperarticulation has a similar effect. For voiced stops, however, enhancement due to contrastive hyperarticulation should result in a shorter VOT, while we expect that in clear speech it would become longer in most cases

(Smiljanić & Bradlow 2008). (In contrast, Kang & Guion 2008 report a shortening of VOT in Korean lenis stops for clear speech hyperarticulation.)

Previous studies provide evidence that context-based contrastive hyperarticulation is active cross-linguistically, for example, duration of /p/ aspiration, fricative voicing contrasts, vowel length and quality contrasts and stop VOT duration in English (Baese-Berk & Goldrick 2009; Kirov & Wilson 2012; Schertz 2013; Seyfarth *et al.* 2016; Nelson & Wedel 2017; Wedel *et al.* 2018); vowels in Korean (Kang *et al.* 2019); and closure duration in singleton–geminate contrasts in Japanese (Sano 2018a).

These studies suggest, in part, that hyperarticulation is cue-specific. That is, the only phonetic cues that participate in contrastive hyperarticulation are those that specifically contribute to the maintenance of the phonetic distance between the target and competitors for a given contrast. For example, hyperarticulation of VOT can be triggered by the presence of ‘voicing’ minimal pairs (e.g., /pit/ vs. /bit/), but not by other lexical neighbours of the target word (e.g., /kit/, /sit/, ...).

The studies mentioned above offer a picture of the cue-specific nature of contrastive hyperarticulation in English. Experimental findings in Kirov & Wilson (2012), for example, indicate that the VOT of word-initial voiceless stops correlates with the existence of a competitor for voicing and place of articulation, while no such correlation is observed in other positions. In Fricke (2013), rather than the number of neighbours of a given word (neighbourhood density), the number of minimal pairs targeting the initial stop was the better predictor of VOT duration of voiceless stops in the Buckeye Corpus. In another experimental study exploring voiced and voiceless stops, Schertz (2013) report that hyperarticulation of VOT was triggered by the existence of a voicing competitor, rather than a place/manner competitor. Seyfarth *et al.* (2016) focus on word-final /s/–/z/ voicing contrast and experimentally demonstrate that speakers hyperarticulate when it is contextually relevant. Specifically, they observed that the signal was enhanced (i.e., shorter vowels before /s/ and longer voicing for /z/) when it resulted in increasing a relevant contrast, namely when there is a lexical competitor. Nelson & Wedel (2017), using a speech corpus, found that the presence of a VOT-specific minimal pair was a better predictor of contrastive hyperarticulation than the number of lexical neighbours differing in the initial segment. They also found a more robust effect for voiced stops (decreased duration when there is a competitor) than for voiceless stops. Lastly, Wedel *et al.* (2018), using the same corpus, investigated both VOT of word-initial stops in voicing contrasts and F1–F2 Euclidean distance for vowel contrasts. For both types, they report that the existence of a cue-specific minimal pair competitor is a trigger for hyperarticulation: a greater phonetic distance with the competitor was observed. Conversely, neighbourhood density was shown to be irrelevant.

This article offers a case study of the cue-specificity of contrastive hyperarticulation, focussing on voicing and length contrasts in Japanese using a speech corpus. Based on the assumption that lexical competition induces synchronic, phonetically specific enhancement of phonemic contrasts (Aylett & Turk 2004, *inter alia*), this study analyses the patterns in sub-lexical hyperarticulation and examines the hypothesis that a specific cue (viz., VOT) that allows distinguishing a word from its minimal pair competitor is hyperarticulated to provide more information and maintain the contrast. Note that corpus studies appear to be especially relevant to the study of contrastive hyperarticulation: Nelson & Wedel (2017) point out that experimental studies, which often involve the use of speech paradigms, do not seem to provide enough motivation to trigger hyperarticulation in speakers. They suggest that speech in an experimental setting may induce so much clear-speech hyperarticulation that contrastive hyperarticulation is obscured or does not occur, as further hyperarticulation may not seem necessary to speakers.

Japanese differs from languages that are already well-studied in the MOP framework (English and other typologically related Indo-European languages) in rhythmic/prosodic properties, phonological characteristics and grammatical structure (such as word order or morphological system). Thus, investigating Japanese provides evidence to the question of whether the cue-specificity of contrastive hyperarticulation holds across language types.

In terms of rhythmic/prosodic properties, Japanese is a mora-timed language (Han 1962, 1994; Port *et al.* 1987), in which moras tend to be of similar duration (see Port *et al.* 1980; Homma 1981; Warner

& Arai 2001a,b for a critical review of mora isochrony and mora-timing compensation), which may affect the realisation of hyperarticulation unlike in other languages. That is, we expect hyperarticulation targeting durational cues to be constrained to maintain mora-based durational contrasts: variations affecting duration should be restricted to a threshold of categorial perception in order to avoid the violation of language-specific timing properties that might hinder communication (e.g., by causing a short segment to be identified as long). Conversely, in a stress-timed language like English, no such restriction should be active, as there are no contrasts based solely on durational cues, and so the degree of hyperarticulation of durational cues should be greater in English than in Japanese. It is therefore meaningful to examine whether there is contrastive hyperarticulation despite the timing restrictions, or whether the phonetic implementation is rigid.

Furthermore, Japanese has two different phonological contrasts that target stops: voicing (see §1.1.1) and length (§1.1.2), which means that the same segment can have both a voicing and a geminate minimal pair (e.g., /kaki/ ‘persimmon’ vs. /kagi/ ‘key’ and /kakki/ ‘energy’). This allows us to test the cue-specificity of contrastive hyperarticulation by comparing the effects that contrasts have on VOT. If contrastive hyperarticulation is cue-specific, then we expect it to target only the specific cue that is relevant to a contrast. However, if it is not cue-specific, then we expect it to target the segment as a whole (as in clear speech hyperarticulation). For example, VOT duration should vary as well in the case of a geminate counterpart (and not only closure duration as reported in previous studies; Sano 2019). Additionally, the literature reports an ongoing change in the VOT of Japanese voiced stops (from long lead to short lag; Riney *et al.* 2007; Gao *et al.* 2019; see §1.1.1), and it seems of interest to investigate how this sound change may affect patterns of hyperarticulation.

Lastly, we are not aware of any study examining multiple contrasts targeting the same segment in the same language. Because Japanese offers a variety of minimal pairs (voicing and length, both contrasts having a high functional load) for stop consonants, it provides a good opportunity to explore the phonetic specificity of contrastive hyperarticulation.

## 1.1. *Phonetic properties of Japanese*

*1.1.1. Voicing contrast in Japanese.* Japanese has a two-way contrast in voicing: voiceless stops /p/, /t/, /k/ contrast with their voiced counterparts /b/, /d/, /g/ (e.g., /kin/ ‘gold’ vs. /gin/ ‘silver’), in both word-initial and word-medial positions (Vance 1987). Prior phonetic research investigating Japanese stops has identified the presence/absence of laryngeal activity and VOT as the main acoustic cues that contribute to the distinction between the two categories (Shimizu 1977 *et seq.*). Thus, the traditional description of VOT opposes two types of plosives, ‘prevoiced’ (negative VOT; Shimizu 1996, 1999) and ‘voiceless unaspirated’ (positive VOT; Homma 1980; Shimizu 1996, 1999), with variations depending on speech rate and position in the word.

However, the nature of VOT in Japanese is more complex than the dichotomy presented above. Recent studies indicate that the realisation of voicing is variable and undergoing a synchronic change: younger speakers tend to devoice stops in word-initial position (while older speakers tend to retain prevoicing), which suggests that VOT might be progressively losing its status as the primary cue for voicing (Takada *et al.* 2015; Gao & Arai 2018; Gao *et al.* 2019). Takada *et al.* (2015) found that the degree of voicing appears to be shifting, as they observe extreme-lead, short-lead and short-lag VOT in speakers from oldest to youngest. Additionally, the VOT of voiceless consonants in Japanese does not match straightforwardly the traditional short lag vs. long lag dichotomy (Lisker & Abramson 1967). Instead, voiceless stops are ‘moderately’ aspirated (Riney *et al.* 2007), with a VOT that is shorter than long-lag VOT but longer than short-lag VOT.

The consequence of the above is that the VOT of voiced and voiceless consonants in initial position may overlap, suggesting that VOT alone might not be enough to maintain the contrast. However, the investigation of other possible secondary cues (f<sub>0</sub>, voice quality, following vowel) has produced limited results, and VOT still remains the primary and necessary cue for the voicing contrast (Riney *et al.* 2007;

Takada *et al.* 2015; Byun 2021). In the case of English, although both voiced and voiceless stops may have a positive VOT (Lisker & Abramson 1964), there is no confusion between categories, because they fall into the short-lag and long-lag categories. In Japanese, however, the distinction may be difficult in initial position, because voiceless stop VOT values are between the short- and long-lag categories. In medial position, however, the contrast is retained.

*1.1.2. Consonantal length contrast in Japanese.* Segmental length plays an important role in Modern Japanese. Consonants in intervocalic position can be short (singleton) or long (geminate), and the distinction between them carries lexical contrasts in a variety of minimal pairs, such as /kako/ ‘past’ vs. /kakko/ ‘parenthesis’ or /hato/ ‘dove’ vs. /hatto/ ‘hat’. A large body of research has been conducted on the phonetics of singletons and geminates in Japanese. These studies have contributed to identifying the acoustic correlates involved in this distinction. The primary acoustic correlate of the singleton–geminate contrast in Japanese is closure duration: constriction for geminates is two to three times longer than for their singleton counterparts, with duration varying according to place and voicing (Han 1962, 1994; Homma 1981; Port *et al.* 1987; Kawahara 2015). Previous studies have also identified factors such as the duration of preceding or following vowels and non-durational acoustic correlates like intensity,  $f_0$  and  $F_1$  (Port *et al.* 1987; Han 1994; Hirata 2007; Kawahara 2015). On the other hand, other phonetic cues such as VOT have been shown to be unrelated to the phonetic implementation of the contrast (Homma 1981; Hirata & Whiton 2005; Sano 2019).

## 1.2. Goals and research agenda

In examining voicing and consonantal length contrasts in Japanese, we aim to provide additional evidence for the cue-specific nature of contrastive hyperarticulation, as MOP predicts that the same phonological unit can be affected in different ways by contrastive hyperarticulation when it contributes to identifying specific words relative to competitors. That is, depending on the competitors in the lexicon (e.g., the presence or absence of minimal pair), specific phonetic cues (e.g., VOT) are enhanced, while others remain unchanged.

Previous studies on VOT in the MOP framework suggest that, in English, the existence of a voiced stop competitor for a word-initial voiceless stop tends to induce contrastive hyperarticulation of VOT. That is, in an experimental setting, contrastive hyperarticulation of a voiceless stop leads to a longer VOT for target words with a minimal pair competitor when compared to those without (Baese-Berk & Goldrick 2009; Peramunage *et al.* 2011). Other studies have also found that hyperarticulation of the voicing contrast could be enhanced by visual stimuli (Kirov & Wilson 2012; Buz *et al.* 2016; Seyfarth *et al.* 2016) and specifically targets VOT, the primary acoustic cue, instead of triggering a lengthening at the word level (Buz *et al.* 2014). In the case of Japanese, it appears reasonable to expect that the presence of lexical competitors should also trigger hyperarticulation of VOT duration: a shorter VOT for voiced consonants and a longer one for voiceless consonants. This is supported by Sano (2018a,b), who found in a corpus study that hyperarticulation can also be observed in Japanese consonantal length contrasts: when a lexical item has lexical competitors, the closure duration is longer for geminates and shorter for singletons.

In the current study, we specifically investigate the prediction that a distinctive phonetic cue (VOT) will undergo hyperarticulation when involved in a lexical contrast in which it maintains the phonetic distance between the target and the competitor. We show that in Japanese the existence of a voicing minimal pair competitor in the lexicon affects (i.e., enhances) the VOT duration of the target segment (shorter for voiced stops, longer for voiceless stops), while no such effect is induced by the existence of other types of contrasts, here consonantal length, in which VOT is not a relevant phonetic cue. This provides further evidence that the phonetic specificity of contrastive hyperarticulation is not limited to English and related languages, but also holds in typologically different languages like Japanese.

The remainder of this article is structured as follows: in §2, we introduce the corpus used for the analysis and the variables included in the statistical model. §3 summarises the results obtained in the statistical analysis. Based on the results, §4 discusses the nature of cue-specificity and other issues related to hyperarticulation with reference to the previous literature. §5 concludes this study.

## 2. Methods

### 2.1. The corpus and data collection

The analysis of the present study is based on the *Corpus of Spontaneous Japanese Relational Database* (NINJAL 2012; henceforth CSJ-RDB). The CSJ-RDB is a subset of the *Corpus of Spontaneous Japanese* (CSJ), one of the largest annotated corpora of spoken Japanese. The CSJ is abundantly annotated with linguistic and non-linguistic information that is suitable for detailed analysis. The target data were retrieved from the CSJ-RDB by focussing on a selection of 12 speech samples that is balanced in speech style (the *CSJ-mini* provided by Hanae Koiso of the National Institute for Japanese Language and Linguistics, which consists of the following speech samples: A01F0055, A02M0098, A05F0039, A11M0846, D01F0023, D01M0009, D04F0022, D04M0010, S00F0014, S01M0005, S02M0043, S03F0108). The breakdown of the *CSJ-mini* is as follows: monologue (eight speech samples) and dialogue (four speech samples), amounting to about 34,000 words produced by 11 speakers (five males and six females; age: 20s (3), 30s (5), 40s (1) and 60s (2)). Using the SQLite database language (<https://jp.navicat.com/>), this study employed the phonetic/phonological and morphological information annotated in the CSJ-RDB.

The target segments were stops (/p/, /t/, /k/, /b/, /d/, /g/). In addition to the target segments, we also retrieved information from the *CSJ-mini* about (a) segments immediately preceding/following the target segments, (b) syllables immediately preceding/following the syllables that contain the target segments and (c) words and phrases that contain the target segments. Tokens were excluded from the dataset, however, if the targeted segments occurred in filled pauses or word fragments.

After exclusion, the remaining target segments in the dataset were categorised as voiceless or voiced, based on the phonetic annotation provided in the CSJ-RDB. Other segmental properties, such as place, position in a word, height of the following vowel and existence of a minimal pair (see §2.2.2) were manually annotated for each target segment.

Segmental intervals in the CSJ-RDB (onset time and offset time) are annotated for each linguistic unit, such as segment, syllable, word and phrase. For stops, separate annotation is provided for the closure (with the label '<cl>') and the burst (labelled with the segment) portions. The duration of VOT was obtained by subtracting the onset time of the target segments from their offset time based on the annotation in the CSJ-RDB. Note that the nature of the labels did not allow us to take into account the presence of laryngeal activity during the closure portion, thus VOT was treated as positive in our analysis, as in English (which has an aspiration contrast). However, as mentioned in §1.1.1., the VOT of voiced plosives in Japanese is undergoing a change, and prevoicing is often absent in younger generations. Given that speakers in the *CSJ-mini* mainly belong to younger generations, we expect to observe few occurrences of prevoicing, and so this treatment of VOT seems suitable. The duration of other units was calculated in the same manner. An exhaustive search and filtering of the data from the *CSJ-mini* resulted in a dataset of 4,448 tokens, of which 1,222 (27.5%) were voiced stops and 3,226 (72.5%) were voiceless stops.

### 2.2. Factors in statistical analysis

**2.2.1. Response variable.** For the purpose of examining the relationship between speech rate and VOT duration, we calculated speech rate by dividing the number of moras in a word containing the target segment by the duration of the word in seconds. As previously reported, VOT varied considerably

depending on speech rate in such a way that they inversely correlated with each other (voiced:  $r = 0.189$ ,  $t(1, 220) = -6.724$ ,  $p < 0.01$ ; voiceless:  $r = 0.123$ ,  $t(3224) = -7.015$ ,  $p < 0.01$ ). In other words, VOT tends to be shorter as the speech rate increases and longer as it decreases. Furthermore, phonetic enhancement resulting from contrastive hyperarticulation should be distinguished from enhancement due to slow/clear-speech hyperarticulation (see, e.g., Wedel *et al.* 2018). For these reasons, raw VOT values observed in the corpus were normalised by speech rate using a measure-internal method (see Wedel *et al.* 2018 and the references cited therein): namely, we multiplied raw VOT by speech rate. Following the previous literature, the speech rate-normalised VOT was then log-transformed (Bell *et al.* 2009; Seyfarth 2014).

**2.2.2. Factors of interest.** As the working hypothesis of this study is to examine if the minimal-pair-driven contrastive hyperarticulation in Japanese is cue-specific, the factors of our primary interest are the presence or absence of minimal pairs for (1) the voicing contrast and (2) the singleton–geminate contrast. Labels regarding the presence/absence of minimal pairs were coded item-by-item, using a lexicon based on three Japanese dictionaries (*Kōjien*, Shimmura 2018; *Sanseidō Kokugo Jiten*, Kembo *et al.* 2013; and *Goo Kokugo Jisyo*, Matsumura 2024). For each word token, from the corresponding phonemic representation of the lemma annotated in the CSJ-RDB, we identified a potential minimal-pair competitor that contrasts with the lemma by substituting the distinctive feature (voicing or length) of the stop in question and checked if the competitor was present as a dictionary entry. If the potential competitor was present, the value of the minimal pair existence was coded as TRUE; otherwise it was coded as FALSE. Lexical accent was not taken into account. In this process, if a member of a pair was a personal name, jargon, or an archaic or dialectal form, the pair was not regarded as a minimal pair, as it is not likely to be shared by the majority of Japanese speakers. This process resulted in the distribution of minimal pairs summarised in Tables 1 (for the voicing contrast) and 2 (for the length contrast).

**2.2.3. Control variables.** Factors that may have an effect on VOT were also included in the model:

**Place of articulation (labial, coronal, dorsal):** VOT differs across places of articulation, as in labial < coronal < dorsal (Lisker & Abramson 1967).

**Position in word (initial/non-initial):** Previous studies on VOT hyperarticulation in English focussed on word-initial stops (Wedel *et al.* 2018), but did not explore stops in non-initial position. However, following previous research on VOT in Japanese, this study explored both word-initial and non-initial positions, taking the effect of position on VOT into account.

**Table 1.** Distribution of minimal pair existence by segment and position for the voicing contrast ( $x/y$ : for each cell,  $x$  is the number of types and  $y$  the number of tokens).

Stop	Word-initial position		Word-medial position	
	Presence	Absence	Presence	Absence
p	0/0	1/1	1/1	5/9
t	5/32	11/238	15/91	115/734
k	6/257	18/281	43/250	335/1332
b	2/4	3/8	4/6	69/198
d	5/38	4/326	8/32	39/267
g	6/29	6/63	21/72	66/179

**Table 2.** Distribution of minimal pair existence by segment and position for the singleton–geminate contrast ( $x/y$ : for each cell,  $x$  is the number of types and  $y$  the number of tokens).

Stop	Word-initial position		Word-medial position	
	Presence	Absence	Presence	Absence
p	0/0	1/1	1/1	5/9
t	0/0	15/270	23/233	103/592
k	0/0	24/538	43/297	334/1285
b	0/0	5/12	2/2	71/202
d	0/0	7/364	0	47/299
g	0/0	12/92	4/15	83/236

**Word frequency:** We counted and log-transformed the number of occurrences of each word in the complete CSJ ( $N(word_x)$ ). Word frequency is known to affect word duration: frequent words are shorter than less frequent ones (Zipf 1935; Wright 1979). Therefore, in the present study, word frequency may affect VOT duration.

**Contextual predictability/local phonotactic probability (backward and forward):** The average probability of each two-phoneme sequence was calculated by dividing the probability (number of occurrences in the corpus) of the target segment and the preceding/following vowel by the unconditional probability of the target segment (backward:  $p(phoneme_x \mid phoneme_{x-1})$ ; forward:  $p(phoneme_x \mid phoneme_{x+1})$ ). The values were log-transformed. Contextual predictability has been shown to be a predictor of duration in a corpus study of English (Seyfarth 2014).

**Following vowel height (high/non-high):** VOT is affected by the height of the following vowel (e.g., Klatt 1975). /i, u/ were coded as high and /a, e, o/ as non-high.

**Following vowel duration:** VOT of stops may be affected by the duration of the following vowel, based on the principle of mora-timing compensation (Port *et al.* 1980; Homma 1981). We can expect that if vowel duration is longer, VOT is shorter and *vice versa*. The values were normalised by speech rate.

**Word length:** the number of moras in a word containing the target segment. The values were log-transformed. Turk & Shattuck-Hufnagel (2000) propose that mean syllable duration may decrease with the number of syllables in a word, a phenomenon that they call ‘polysyllabic shortening.’ If we postulate that the same is true with moras, then the more moras a word contains, the shorter a mora will be. Segments should be affected accordingly at the level of the phonetic cue.

2.3. Model building

Following prior work (e.g., Wedel *et al.* 2018), we fit separate models for voiced and voiceless stops, since (a) VOT significantly differs depending on voicing (voiced: mean = 16.15 ms, SD = 7.17; voiceless: mean = 21.84 ms, SD = 10.05;  $t(3068.8) = -21.008, p < 0.01$ ); (b) contrastive hyperarticulation is expected to affect VOT in opposite directions: voiced stops become shorter and voiceless stops become longer (if there are minimal-pair competitors); and (c) other control variables may not affect voiced and voiceless stops the same way. In particular, we fitted linear mixed-effects models to our data using *lmer* of the *lmerTest* package (Kuznetsova *et al.* 2017) in R (R Core Team 2019). Variables included in the full model were the normalised VOT as the dependent variable; the fixed effects were presence/absence of minimal pairs (for voicing and singleton–geminate contrasts), place of articulation, position in word, word frequency, contextual predictability (backward and

forward), following vowel height, following vowel duration and word length. Random intercepts for speaker and item (lemma) and by-speaker random slopes (consisting of position, voicing minimal pair, singleton-geminate minimal pair and an interaction term of position and voicing minimal pair) were also included in the model. The model was structured to include the greatest number of theoretically relevant factors, rather than focussing on the most complex random effects structure (see Tang & Bennett 2018, following Baayen *et al.* 2017; Matuschek *et al.* 2017). The model fit was assessed by referring to AIC, BIC and log-likelihood.

### 3. Results

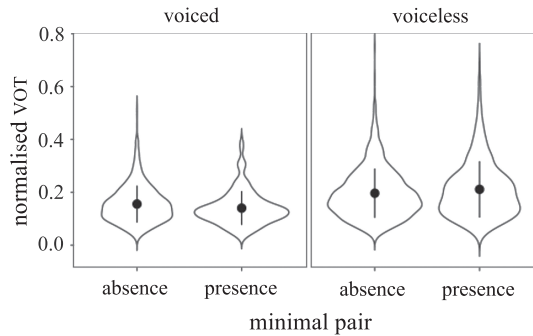
We now turn to the description of the final models. Results are presented separately for voiced and voiceless stops. Table 3 presents the summary of fixed factors for the voiced-stop model. For the control predictors, place of articulation, position in word, word frequency, contextual predictability (backward and forward), following vowel height, following vowel duration and word length were all retained in the final model, which indicates that they make significant contributions to model fit. Although word frequency did not reach significance, other factors had highly significant positive or negative effects on VOT duration.

The factors of interest regarding the two minimal pair contrasts were retained in the final model. Figure 1 illustrates the distribution of speech rate normalised VOT for voiced and voiceless stops, by presence/absence of a minimal pair competitor.

Figure 1 shows that the mean VOT value for voiced stops is lower when a lexical competitor (voiceless counterpart) exists than when there is no such competitor. This is also confirmed in Table 3. The presence of a minimal-pair competitor in a voicing contrast was shown to significantly predict a shorter VOT ( $p < 0.01$ ) for voiced stops, as indicated by the negative coefficient for this variable. The average difference in raw VOT between voiced and voiceless stops when there is a minimal pair competitor was 8.45 ms, and it was 5.08 ms with no such competitor (see §4.1 for a comparison with English). On the other hand, the presence of a lexical competitor for the singleton–geminate contrast (i.e., a geminate counterpart) was not significantly predictive ( $p = 0.38304$ ). This is consistent with the hypothesis that contrastive hyperarticulation of a given cue is triggered by the existence of a minimal pair distinguished by that cue. Additionally, the shorter VOT in voiced stops caused by the existence of voicing minimal pairs suggests that what is observed in this model is contrastive hyperarticulation, rather than slow/clear-speech hyperarticulation. Next, let us turn our attention to the summary of fixed factors for the voiceless stop model in Table 4.

**Table 3.** Fixed effects summary for the voiced-stop model.

	Estimate	Standard error	<i>t</i> -Value	<i>p</i> -Value
(Intercept)	0.132	0.0327	4.038	<0.001
Minimal pair (voicing) TRUE	−0.0285	0.0102	−2.787	<0.01
Minimal pair (sing-gem) TRUE	0.0224	0.0245	0.914	0.38304
Place dorsal	0.0189	0.0086	2.195	<0.05
Place labial	0.0298	0.0104	2.854	<0.01
Position noninitial	−0.0258	0.008	−3.23	<0.01
Word frequency	0.0022	0.0028	0.77	0.44161
Backward contextual pred.	0.0251	0.0091	2.76	<0.01
Forward contextual pred.	−0.0482	0.022	−2.192	<0.05
Following V non-high	0.0709	0.0282	2.518	<0.05
Following V duration	−0.0574	0.0105	−5.456	<0.001
Word length	0.0486	0.0215	2.261	<0.05



**Figure 1.** Distribution of speech rate-normalised VOT values by presence/absence of a minimal-pair competitor for voiced and voiceless stops. Solid circles represent mean values and vertical lines interquartile ranges.

**Table 4.** Fixed effects summary for the voiceless stop model.

	Estimate	Standard error	<i>t</i> -Value	<i>p</i> -Value
(Intercept)	0.0941	0.0251	3.748	<0.001
Minimal pair (voicing) TRUE	0.0253	0.0098	2.572	<0.05
Minimal pair (sing-gem) TRUE	−0.0109	0.0079	−1.387	0.16767
Place dorsal	0.0616	0.0068	8.988	<0.001
Place labial	−0.0536	0.0411	−1.303	0.19292
Position noninitial	−0.0054	0.0104	−0.522	0.60556
Word frequency	0.0101	0.0029	3.447	<0.001
Backward contextual pred.	−0.0104	0.0117	−0.888	0.37454
Forward contextual pred.	−0.0233	0.0121	−1.928	0.05404
Following V non-high	−0.0272	0.0069	−3.963	<0.001
Following V duration	−0.0238	0.0095	−2.508	<0.05
Word length	0.1229	0.0213	5.779	<0.001

For the control predictors, place of articulation, position in word, word frequency, contextual predictability (backward and forward), following vowel height, following vowel duration and word length were all retained in the final model, suggesting that these predictors significantly contribute to model fit. Unlike in the voiced model, place of articulation (labial), position in word and backward contextual predictability were not significant, while word frequency was highly significant. Forward contextual predictability was close to significance at the five percent level.

For the factors of interest, both retained in the final model, the presence of a minimal pair competitor in a voicing contrast (voiced counterpart) was found to significantly affect VOT in the expected direction: longer VOT ( $p < 0.05$ ), as illustrated in Figure 1, while the presence of a minimal-pair competitor in the singleton–geminate contrast did not reach significance ( $p = 0.16767$ ). The result in the voiceless stop model is again consistent with the hypothesis that contrastive hyperarticulation is triggered by the existence of cue-specific minimal pairs, but not by non-cue-specific minimal pairs. The direction of the effect of cue-specific lexical competition in the voiceless-stop model was reversed from what was found in the voiced-stop model, in support of the hypothesis that contrastive hyperarticulation is realised in such a way that the phonetic correlate of a distinctive feature moves away from the competitor.

However, care should be taken in interpreting these results, since the distribution of geminates is biased with respect to position and voicing. As Table 2 shows, there is no singleton–geminate contrast

**Table 5.** Fixed effects summary for the word-medial voiceless stop model.

	Estimate	Standard error	<i>t</i> -Value	<i>p</i> -Value
(Intercept)	0.1272	0.0209	6.096	<0.001
Minimal pair (voicing) TRUE	0.0218	0.0087	2.515	<0.05
Minimal pair (sing-gem) TRUE	−0.0075	0.0048	−1.564	0.12024
Place dorsal	0.0424	0.0053	8.066	<0.001
Place labial	0.0127	0.035	0.363	0.7164
Word frequency	0.0091	0.0021	4.291	<0.001
Backward contextual pred.	0.0208	0.0117	1.772	0.07647
Forward contextual pred.	−0.0507	0.0201	−2.519	<0.05
Following V non-high	−0.0278	0.0056	−4.974	<0.001
Following V duration	0.0232	0.0156	1.484	0.13791
Word length	0.0703	0.0169	4.164	<0.001
Forward contextual pred.: Following V duration	0.1008	0.0374	2.693	<0.01

word-initially in Japanese; consonantal length is contrastive only word-medially. Additionally, there are very few voiced geminates. Considering the possible effects of these biases, we ran a third model consisting only of the word-medial voiceless stop fraction of the data for triangulation purposes. For consistency, position in word, as a control predictor and as a term in the by-speaker random slope, was taken out of this model. The results are presented in Table 5.

In the word-medial-voiceless-stop model, the pattern for the control predictors is mostly similar to the voiceless-stop model in Table 4, except for the following vowel duration, which did not reach significance. Most importantly, the pattern of the factors of interest in this model is consistent with the voiceless-stop model in that the presence of a minimal-pair competitor in a voicing contrast (i.e., a voiced counterpart) was found to significantly affect VOT, making it longer ( $p < 0.05$ ), while the presence of a minimal-pair competitor in the singleton–geminate contrast did not reach significance ( $p = 0.12024$ ). Thus, we can safely reject the possibility that the lack of significant contrastive hyperarticulation effect of singleton–geminate minimal pairs is due to an unbalanced distribution of geminates regarding position and voicing.

#### 4. Discussion

We confirmed in the previous section that (a) the presence of a voicing minimal pair competitor induces hyperarticulation of VOT in such a way that voiceless stops become longer, while voiced stops become shorter; and (b) the presence of a singleton–geminate minimal pair competitor does not induce hyperarticulation of VOT. The VOT of stop consonants in Japanese thus provides additional support for the cue-specific nature of contrastive hyperarticulation. This is the first study that attests the existence of minimal pair-driven contrastive hyperarticulation of VOT in Japanese, which is typologically different from better-studied languages in its grammatical structure and rhythmic/prosodic properties. This provides additional evidence that the cue-specificity of contrastive hyperarticulation holds across language types.

As reviewed above, VOT functions as the main acoustic cue for the distinction between voiced and voiceless stops in Japanese (Shimizu 1977 *et seq.*). For the singleton–geminate contrast, closure duration has been shown to be the primary acoustic correlate in natural speech (Sano 2019). Furthermore, the perceptual distance between singletons and geminates is found to be enhanced by contrastive hyperarticulation (Sano 2018a,b). Building upon prior research, the present study provides

further evidence for the existence of contrastive hyperarticulation in VOT in Japanese, but its effect is induced only by the existence of voiced/voiceless minimal-pair competitors, and it is insensitive to lexical competition with singleton–geminate minimal pairs. This is consistent with previous findings that contrastive hyperarticulation occurs only at the level of the phonetic cue; hence, it is cue-specific.

#### 4.1. Degree of contrastive hyperarticulation

The results of this study are consistent with recent seminal work on the effect of contrastive hyperarticulation on VOT in English (Wedel *et al.* 2018). We found that a greater difference was observed between voiceless and voiced stops' VOT when a lexical competitor exists in the lexicon than when there is no such competitor. In this study, the average difference in raw VOT between voiced and voiceless stops with a minimal-pair competitor (8.45 ms) is about 66% greater than without such a competitor (5.08 ms).

This difference between VOT of voiced *vs.* voiceless stops in the presence/absence of a minimal pair is below the threshold of the Weber–Fechner law of the just-noticeable difference – that is, the minimal difference between two stimuli that leads to a change in experience (Treutwein 1995). Under the assumptions of the just-noticeable difference, a 3 ms difference in duration seems indeed too small for listeners to make a conscious distinction between the two competitors (see Lehiste 1976, *inter alia*; a 10 ms difference is necessary for hearers to distinguish two sounds). However, contrastive hyperarticulation is assumed to be an unconscious mechanism of spontaneous speech (in contrast with the conscious mechanism involved in clear/elicited speech): it is speakers' implicit knowledge of the existence of a minimal pair in the lexicon that leads them to enhance specific cues in order to better keep lexical items apart. We can thus postulate that the same is true for perception: both speakers and listeners should be able to unconsciously use sub-lexical information at a higher level for a more efficient communication process. Findings from a study on VOT in English by McMurray *et al.* (2002) suggest that this inference is correct. Their results in a perceptual experiment where VOT varied on a continuum indicate that fine-grained acoustic differences at the level of the phonemic cue that have minimal effects on phoneme identification play, in fact, a crucial role in lexical access. Drawing a parallel with the results of the present study, we propose that a minimal change in VOT, which should not be relevant when examining the two segments in isolation, provides listeners with additional cues for lexical identification.

Let us now turn our attention to the reasons behind the small size of this difference. One possibility is that, since the just-noticeable difference is defined as being proportional to the size of the original unit, the increase should not be considered in terms of raw duration, but as a proportion. In this case, the small size of the difference in VOT between voiced and voiceless stops without a minimal-pair competitor (5.08 ms) might justify the small size of the difference between stops with competitors (8.45 ms), as a 66% increase between the two is observed. Note that contrastive hyperarticulation of the closure duration in the singleton–geminate contrast (Sano 2018a) represents an increase of 59% of the difference between singleton and geminate closure (81.7 ms with a minimal-pair competitor *vs.* 51.2 ms without).

We can find further food for thought in the nature of VOT in Japanese. As previously mentioned in §1.1.1, recent studies report that VOT in Japanese is undergoing a change and tends to deviate from its original description, especially in younger speakers: devoicing is often observed in initial position, and VOT seems to be shifting from a long lead to a short lag (Riney *et al.* 2007; Takada *et al.* 2015; Gao & Arai 2018; Gao *et al.* 2019). What these studies suggest is that VOT seems to be losing its role as a primary cue for the voicing contrast in Japanese. Thus, one might postulate that cues other than VOT might also be hyperarticulated to enhance the contrast. In this regard, our models seem to indicate that the following vowel plays a role in the contrast, as we found a correlation with the following vowel duration in both the voiced ( $p < 0.001$ ) and the voiceless ( $p < 0.05$ ) models. On the other hand, as indicated in Gao *et al.* (2019), although the status of VOT as a phonetic cue for voicing appears to be

shifting, there is a lag between production and perception, and they find that in terms of perception, listeners still heavily rely on VOT. This suggests that even a small difference in VOT duration should be meaningful.

If we put the present results into perspective with those of Wedel *et al.* (2018), it is interesting to note that the degree of contrastive hyperarticulation was found to be greater in Japanese than in English. In Wedel *et al.* (2018), the average difference in raw VOT between voiced and voiceless stops in words that have a voicing minimal-pair competitor (63 ms) is approximately 20% greater than those in words that do not have such a competitor (53 ms), while it was 66% in this study.

One clue to understanding this difference between English and Japanese may be the role of aspiration. Wedel *et al.* (2018) and most previous studies focussed on word-initial stops, where voiceless segments are aspirated, producing longer VOT durations (Lisker & Abramson 1964). For this reason, the effect of contrastive hyperarticulation on VOT in English may be inhibited; that is, it takes more effort to produce contrastive articulation at longer VOTs. The present study, on the other hand, targeted both word-initial and non-initial stops, because Japanese does not have positional aspiration, and VOT is thus not expected to differ greatly between these two positions. Compared to English, the shorter VOT in Japanese may allow extra room for the effect of contrastive hyperarticulation (i.e., lengthening). Furthermore, for a shorter VOT to be salient enough to convey information about lexical contrast, the duration should be enhanced to a greater extent. In other words, voiced and voiceless stops in Japanese are phonetically more similar, with closer VOT durations; the stops of interest therefore are subject to a stronger effect of contrastive hyperarticulation (even 66% of enhancement results in no more than 3.37 ms of difference). This is not the case for English, where the difference in VOT between voiced and voiceless stops is greater than in Japanese (20% of enhancement results in about 10 ms of difference).

## 4.2. Other issues

**4.2.1. Position in word.** It has been shown that language users pay more attention to word-initial positions (Bruner & O'Dowd 1958), which contribute more information to lexical identification (van Son & Pols 2003; Wedel *et al.* 2019). Beckman (1998) classifies initial syllables as 'psycholinguistically prominent', and therefore initial positions favour phonological processes that enhance the realisation of lexical contrasts. Other previous research has shown that cross-linguistically, strengthening phonological processes are more likely to target word beginnings, while neutralisation processes prefer word ends (Barnes 2006; Wedel *et al.* 2019). In terms of prosody as well, domain-initial positions are stronger (Keating 2003 and references cited therein). In our dataset, the difference in raw VOT between voiced and voiceless stops with and without a minimal-pair competitor was slightly greater in initial positions (3.82 ms) than in non-initial positions (3.58 ms), although the predictor, position in word, was significant in the voiced-stop model ( $p < 0.01$ ) but not in the voiceless-stop model. From this, a possibility arises that the degree of contrastive hyperarticulation differs depending on the position of the segment in the word. If the positional difference is confirmed in future work, it can provide additional evidence for the cross-linguistic precedence of (word) initial positions over non-initial positions (e.g., Wedel *et al.* 2019).

**4.2.2. Slow/clear speech vs. casual speech.** In §3, our results confirmed that contrastive hyperarticulation tends to be incompatible with slow/clear speech. This supports previous findings in Wedel *et al.* (2018) that, assuming that the purpose of contrastive hyperarticulation is to avoid perceptually confusable productions near the category boundary, its effect may be less robust in slow/clear speech, where phonemic categories are less likely to overlap. Conversely, the context where contrastive hyperarticulation can be observed more robustly and is more likely to occur would be casual speech, which includes more perceptually confusable productions due, for example, to reduction processes (Wedel *et al.* 2018).

We examined the potential connection between speech style and the effect of contrastive hyperarticulation by running additional models with style in interaction with the existence of a voicing minimal pair. The distinction in the CSJ-RDB (monologue *vs.* dialogue) corresponds to the distinction between slow/clear speech and casual speech, as monologues represent slow/clear speech and dialogues represent casual speech (Maekawa 2003). The gap in raw VOT duration due to the existence of a minimal-pair competitor was greater in dialogues than in monologues both in voiced and voiceless stops (voiced: monologues = 0.47 ms, dialogues = 3.05 ms; voiceless: monologues = 2.21 ms, dialogues = 3.11 ms), although the predictor was not significant in the voiceless-stop model, and the voiced-stop model only showed a tendency to significance ( $p = 0.09068$ ). If it is confirmed in future work that the degree of contrastive hyperarticulation is greater in casual speech than in slow/clear speech, it will reinforce our understanding of the distinction between slow/clear-speech hyperarticulation and contrastive hyperarticulation, and suggest that contrastive hyperarticulation is induced only when necessary.

## 5. Conclusion

This study sheds light on hitherto unexplored aspects of contrastive hyperarticulation. Building upon prior work, this study provided an additional test case that supports cue-specificity. As mentioned above, in Japanese two kinds of minimal pairs are distinguished by duration-based cues coexisting within a single stop consonant: VOT for the voicing contrast and closure duration for the singleton–geminate contrast. By taking advantage of the variety of minimal pairs distinguished based solely on durational considerations, we examined how information about lexical competition is reflected at the level of the phonetic cue. The results showed that what matters in contrastive hyperarticulation of VOT is the existence of voicing minimal pairs, providing further support for the cue-specificity of contrastive hyperarticulation. Our results also offer additional support for MOP’s perspective that language is an effective tool for communication, and that speakers phonetically enhance cues for accurate message transmission: when there is a lexical competitor, VOT is hyperarticulated and the distance is increased, but speakers are less likely to hyperarticulate when little benefit is expected.

**Acknowledgements.** We would like to thank Andrew Wedel for his valuable feedback and support on earlier versions of this article. We are also deeply grateful to the editors and anonymous reviewers for their constructive suggestions. Any remaining errors are solely our responsibility.

**Funding statement.** This study is supported by the Japan Society for the Promotion of Science KAKENHI Grant No. 19K00558.

**Competing interests.** The authors declare no competing interests.

## References

- Aylett, Mathew & Alice Turk (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* **47**, 31–56.
- Baayen, Harald R., Shravan Vasishth, Reinhold Kliegl & Douglas Bates (2017). The cave of shadows: addressing the human factor with generalized additive mixed models. *Journal of Memory and Language* **94**, 206–234.
- Baese-Berk, Melissa & Matthew Goldrick (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes* **24**, 527–554.
- Barnes, Jonathan (2006). *Strength and weakness at the interface: positional neutralization in phonetics and phonology*. Berlin: Mouton de Gruyter.
- Bayes, Thomas (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**, 370–418.
- Beckman, Jill N. (1998). *Positional faithfulness*. PhD dissertation, University of Massachusetts, Amherst.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* **60**, 92–111.
- Bell, Alan, Dan Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *JASA* **113**, 1001–1024.

- Bowern, Claire & Sarah Babinsky (2018). Redundancy, and category variation. *Topics in Cognitive Science* 8, 503–513.
- Bradlow, Ann R., Gina M. Torretta & David B. Pisoni (1996). Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255–272.
- Bresnan, Joan & Jessica Spencer (2012). Frequency effects in spoken syntax: Have and be contraction. Presented at the symposium on New Ways of Analyzing Syntactic Variation, Radboud University, Nijmegen, November 2012.
- Bruner, Jerome S. & Donald O'Dowd (1958). A note on the informativeness of parts of words. *Language and Speech* 1, 98–101.
- Buz, Esteban, Florian T. Jaeger & Michael K. Tanenhaus (2014). Contextual confusability leads to targeted hyperarticulation. In Paul Bello, Marcello Guarini, Marjorie McShane & Brian Scassellati (eds.) *Proceedings of the 36th annual meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 1970–1975.
- Buz, Esteban, Michael K. Tanenhaus & Florian T. Jaeger (2016). Dynamically adapted context-specific hyper-articulation: feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language* 89, 68–86.
- Byun, Hi-Gyung (2021). Acoustic characteristics for Japanese stops in word-initial position: VOT and post-stop f0. *Journal of the Phonetic Society of Japan* 23, 174–197.
- Chodroff, Eleanor R. & Colin Wilson (2018). Predictability of stop consonant phonetics across talkers: between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard* 4, article no. 20170047 (11 pp.).
- Cohen Priva, Uriel (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6, 243–278.
- Frank, Austin F. & Florian T. Jaeger (2008). Speaking rationally: uniform information density as an optimal strategy for language production. In Bradley Love, Ken McRae & Vladimir Sloutsky (eds.) *Proceedings of the 30th annual meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 939–944.
- Fricke, Melinda D. (2013). *Phonological encoding and phonetic duration*. PhD dissertation, University of California, Berkeley.
- Fry, John S. (2001). *Ellipsis and wa-marking in Japanese conversation*. PhD dissertation, Stanford University.
- Gao, Jiayin & Takayuki Arai (2018). F0 perturbation in a “pitch accent” language. In *Proceedings of the 6th International Symposium on Tonal Aspects of Languages (TAL 2018)*. International Speech Communication Association, 56–60. Published online at <https://doi.org/10.21437/TAL.2018-12>.
- Gao, Jiayin, Jihyeon Yun & Takayuki Arai (2019). VOT-f0 coarticulation in Japanese: production-biased or misparsing? In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*. Australasian Speech Science and Technology Association, Inc., and International Phonetic Association, 210–214.
- Hall, Kathleen Currie, Elizabeth Hume, Florian T. Jaeger & Andrew Wedel (2016). The message shapes phonology. Ms, University of British Columbia, University of Canterbury, University of Rochester and University of Arizona. Available at [https://www.researchgate.net/publication/309033386\\_The\\_Message\\_Shapes\\_Phonology](https://www.researchgate.net/publication/309033386_The_Message_Shapes_Phonology).
- Hall, Kathleen Currie, Elizabeth Hume, Florian T. Jaeger & Andrew Wedel (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4, article no. 20170027 (15 pp.).
- Han, Miekō (1962). The feature of duration in Japanese. *Onsei no Kenkyuu* 10, 65–80.
- Han, Miekō (1994). Acoustic manifestations of mora timing in Japanese. *JASA* 96, 73–82.
- Hashimoto, Daiki (2021). Probabilistic reduction and mental accumulation in Japanese: frequency, contextual probability, and average predictability. *JPh* 87, 1–17.
- Hirata, Yukari (2007). Durational variability and invariance in Japanese stop quantity distinction: roles of adjacent vowels. *Journal of the Phonetic Society of Japan* 11, 9–22.
- Hirata, Yukari & Jacob Whiton (2005). Effects of speaking rate on the singleton/geminate stop distinction in Japanese. *JASA* 118, 1647–1660.
- Homma, Yayoi (1980). Voice onset time in Japanese stops. *Onsei Gakkai Kaiho* 163, 7–9.
- Homma, Yayoi (1981). Durational relationship between Japanese stops and vowels. *JPh* 9, 273–281.
- Hume, Elizabeth & Ilana Bromberg (2005). Predicting epenthesis: an information-theoretic account. Paper presented at the 7th Annual Meeting of the French Network of Phonology, Aix-en-Provence, January 2005.
- Jaeger, Florian T. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology* 61, 23–62.
- Jaeger, Florian T. & Esteban Buz (2017). Signal reduction and linguistic encoding. In Eva M. Fernandez & Helen Smith Cairns (eds.) *Handbook of psycholinguistics*. Hoboken, NJ: Wiley-Blackwell, 38–81.
- Jaeger, Florian T. & Jane Grimshaw (2013). Information density affects both production and grammatical constraints. Paper presented at Architectures and Mechanisms for Language Processing (AMLaP), Université Aix-Marseille, September 2013.
- Kang, Kyoung-Ho & Susan G. Guion (2008). Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *JASA* 124, 3909–3917.
- Kang, Yoonjung, Na-Young Ryu & Suyeon Yun (2019). Contrastive hyperarticulation of vowels in two dialects of Korean. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*. Australasian Speech Science and Technology Association, Inc., and International Phonetic Association, 43–47.
- Kawahara, Shigeto (2015). The phonetics of sokuon, or geminate obstruents. In Haruo Kubozono (ed.) *Handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton, 43–78.
- Kawahara, Shigeto & Seunghun J. Lee (2018). Truncation in message-oriented phonology: a case study using Korean vocative truncation. *Linguistics Vanguard* 4, article no. 20170016.

- Keating, Patricia (2003). Phonetic encoding of prosodic structure. In Sallyanne Palethorpe & Marija Tabain (eds.) *Proceedings of the 6th International Seminar on Speech Production*. Sydney: Macquarie Centre for Cognitive Science, 119–124.
- Kembo, Hidetoshi, Takashi Ichikawa, Yoshifumi Hida, Makoto Yamazaki, Hiroaki Iima & Takehiro Shioda (eds.) (2013). *Sanseidō kokugo jiten [Sanseidō Japanese dictionary]*. 7th edition. Tokyo: Sanseidō.
- Kirov, Christo & Colin Wilson (2012). The specificity of online variation in speech production. In Naomi Miyake, David Peebles & Richard P. Cooper (eds.) *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 587–592.
- Klatt, Dennis H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research* **18**, 686–706.
- Kurumada, Chigusa & Florian T. Jaeger (2015). Communicative efficiency in language production: optional case-marking in Japanese. *Journal of Memory and Language* **83**, 152–178.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* **82**, 1–26.
- Laplace, Pierre S. ([1820] 1886). *Théorie analytique des probabilités*. 3rd edition. Paris: Courcier. Reprinted in 1886 by Gauthier-Villars as volume 7 of *Œuvres complètes de Laplace*.
- Lee, Hanjung (2006). Parallel optimization in case systems: evidence from case ellipsis in Korean. *Journal of East Asian Linguistics* **15**, 69–96.
- Lehiste, Ilse (1976). Suprasegmental features in speech. In Norman Lass (ed.) *Contemporary issues in experimental phonetics*. New York: Academic Press, 225–239.
- Lisker, Leigh & Arthur S. Abramson (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384–422.
- Lisker, Leigh & Arthur S. Abramson (1967). Some effects of context on voice onset time in English stops. *Language and Speech* **10**, 1–28.
- Maekawa, Kikuo (2003). Corpus of spontaneous Japanese: its design and evaluation. In *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*. International Speech Communication Association, 7–12.
- Matsumura, Akira (ed.) (2024). *Goo kokugo jisho [goo Japanese dictionary]*. Tokyo: NTT. Available at <https://dictionary.goo.ne.jp/jn/>.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald R. Baayen & Douglas Bates (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* **94**, 305–315.
- McMurray, Bob, Michael K. Tanenhaus & Richard N. Aslin (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition* **86**, 33–42.
- Nelson, Noah & Andrew Wedel (2017). The phonetic specificity of competition: contrastive hyperarticulation of voice onset time in conversational English. *JPh* **64**, 51–70.
- NINJAL, National Institute for Japanese Language and Linguistics (2012). *Corpus of spontaneous Japanese relational database*. Available at <https://clrd.ninjal.ac.jp/csj/>.
- Norcliffe, Elizabeth J. & Florian T. Jaeger (2015). Predicting head-marking variability in yucatec maya relative clause production. *Language and Cognition* **8**, 1–39.
- Payton, Karen L., R.M. Uchanski & Louis D. Braid (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *JASA* **95**, 1581–1592.
- Peramunage, Dasun, Sheila E. Blumstein, Emily B. Myers, Mathew Goldrick & Melissa Baese-Berk (2011). Phonological neighbourhood effects in spoken word production: an fMRI study. *Journal of Cognitive Neuroscience* **23**, 593–603.
- Port, Robert F., Salman Al-Ani & Shosaku Maeda (1980). Temporal compensation and universal phonetics. *Phonetica* **37**, 235–252.
- Port, Robert F., Jonathan Dalby & Michael O'Dell (1987). Evidence for mora timing in Japanese. *JASA* **81**, 1574–1585.
- R Core Team (2019). R: a language and environment for statistical computing. Available at <https://www.r-project.org>.
- Riney, Timothy, Naoyuki Takagi, Kaori Ota & Yoko Uchida (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *JPh* **35**, 439–443.
- Sano, Shin-ichiro (2018a). Durational contrast in gemination and informativity. *Linguistics Vanguard* **4**, article no. 20170011.
- Sano, Shin-ichiro (2018b). Minimal pairs and hyperarticulation of singleton and geminate consonants as enhancement of lexical/pragmatic contrasts. *NELS* **48**, 53–66.
- Sano, Shin-ichiro (2019). The distribution of singleton/geminate consonants in spoken Japanese and its relation to preceding/following vowels. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*. Australasian Speech Science and Technology Association, Inc., and International Phonetic Association, 1833–1837.
- Schertz, Jessamyn (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *JPh* **41**, 249–263.
- Seyfarth, Scott (2014). Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. *Cognition* **133**, 140–155.
- Seyfarth, Scott, Esteban Buz & Florian T. Jaeger (2016). Dynamic hyperarticulation of coda voicing contrasts. *JASA* **139**, EL31–EL37.

- Shannon, Claude E. (1948). A mathematical theory of communication. *Mobile Computing and Communications Review* 5, 3–55.
- Shaw, Jason & Shigeto Kawahara (2017). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech* 62, 80–114.
- Shimizu, Katsumasa (1977). Voicing features in the perception and production of stop consonants by Japanese speakers. *Studia Phonologica* 11, 25–34.
- Shimizu, Katsumasa (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido.
- Shimizu, Katsumasa (1999). A study on phonetic characteristics of voicing of stop consonants in Japanese and English. *Journal of the Phonetic Society of Japan* 3, 4–10.
- Shimmura, Izuru (ed.) (2018). *Kōjien [Wide garden of words]*. 7th edition. Tokyo: Iwanami Shoten.
- Smiljanić, Rajka & Ann R. Bradlow (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *JPh* 36, 91–113.
- van Son, Robert J.J.H. & Louis C.W. Pols (2003). Information structure and efficiency in speech production. In Hervé Boulard (ed.) *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech)*. International Speech Communication Association, 769–772.
- Takada, Mieko, Eun Jong Kong, Kiyoko Yoneyama & Mary E. Beckman (2015). Loss of prevoicing in modern Japanese /g, d, b/. In *Proceedings of the 18th International Congress of Phonetic Sciences*. London: International Phonetic Association, 5 pp.
- Tang, Kevin & Ryan Bennett (2018). Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan). *JASA* 144, 997–1017.
- Tily, Harry & Steven T. Piantadosi (2009). Refer efficiently: use less informative expressions for more predictable meanings. In Kees van Deemter, Albert Gatt, Roger van Gompel & Emiel Krahmer (eds.) *Proceedings of the Workshop on the Production of Referring Expressions: bridging the gap between computational and empirical approaches to reference (PRE-CogSci 2009)*. Amsterdam: Cognitive Science Society, 8 pp.
- Treutwein, Bernhard (1995). Adaptive psychophysical procedures. *Vision Research* 35, 2503–2522.
- Turk, Alice & Stephanie Shattuck-Hufnagel (2000). Word-boundary-related duration patterns in English. *JPh* 28, 397–440.
- Turnbull, Rory (2018). Effects of lexical predictability on patterns of phoneme deletion/reduction in conversational speech in English and Japanese. *Linguistics Vanguard* 4, article no. 20170033.
- Vance, Timothy J. (1987). *An introduction to Japanese phonology*. Albany, NY: SUNY Press.
- Warner, Natasha & Takayuki Arai (2001a). Japanese mora-timing: a review. *Phonetica* 58, 1–25.
- Warner, Natasha & Takayuki Arai (2001b). The role of the mora in the timing of spontaneous Japanese speech. *JASA* 109, 1144–1156.
- Wedel, Andrew, Scott Jackson & Abby Kaplan (2013a). Functional load and the lexicon: evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech* 56, 395–417.
- Wedel, Andrew, Abby Kaplan & Scott Jackson (2013b). High functional load inhibits phonological contrast loss: a corpus study. *Cognition* 128, 179–186.
- Wedel, Andrew, Noah Nelson & Rebecca Sharp (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language* 100, 61–88.
- Wedel, Andrew, Adam Ussishkin & Adam King (2019). Crosslinguistic evidence for a strong statistical universal: phonological neutralization targets word-ends over beginnings. *Lg* 95, 428–446.
- Wright, Charles E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition* 7, 411–419.
- Zipf, George K. (1935). *The psycho-biology of language*. Boston, MA: Houghton Mifflin.