Interaction of high vowel devoicing and syllabification

NINJAL, 13 October, 2016
Syllables and Prosody Workshop

Shigeto Kawahara & Jason Shaw
Keio University and Yale University

慶應義塾大学 言語文化研究所 The Keio Institute of Cultural and Linguistic Studies

Yale University

Introduction

What we would like to do

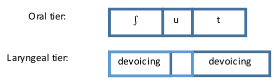
- Japanese high vowels are "devoiced" in some environments.
- Q1: Are these vowels deleted? Or are they devoiced?
- Q2: If they are deleted, then what's the consequence for moraification/syllabification?
- (Q3: if Japanese speakers have CVC syllables due to high vowel deletion, are they sensitive to syllable-related phonotactic restrictions, possibly provided by UG: cf. Berent's recent work?)
- We offer new articulatory data on the table to bear on these questions.

Japanese high vowel devoicing

- The standard description: High vowels are devoiced between two voiceless consonants.
- An example, [ʃ**u**taisei] "willingness".
- "Devoicing" is the traditional term used. However, whether these vowels are simply devoiced or deleted still remain debated.
  - Devoiced=oral gestures remain intact or reduced.
  - Deleted=oral gestures deleted.

Devoicing vs. Deletion?

- A phonetic devoicing view:
  - Laryngeal gestures overlap to passively devoice vowels (e.g., Jun & Beckman 1993).



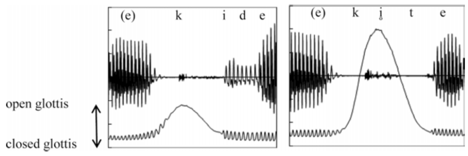
- A phonological deletion view:
  - V → ∅ / [-voice] _____ [-voice] (Kondo 2000)
  - Or a phonological process induced by a set of markedness constraints (Tsuchida 1997).

Or could it be both?

- Kawakami (1971) lists environments where the target vowels are deleted and where they are devoiced.
- But Kawakami gives no instrumental evidence.
- Whang (2014, 2016) argues that vowels are deleted when their identity can be recovered from the preceding consonants.
- For example, the devoiced vowel is predictably [u] after [ɸ], but it is not after [k]. Only in the former environment can the vowel be deleted.

Japanese laryngeal gestures

- In devoiced vowel contexts, e.g., /**kit**e/, there is a single laryngeal gesture of **greater magnitude** than a single consonant gesture, c.f /**k**ide/ (Fujimoto et al. 2002)



How about lingual gestures?

- However, the lingual gesture of devoiced vowels is understudied.
- Funatsu & Fujimoto (2011), as far as we know, is the only exception. They used EMMA, EGG, and PGG.
- Their EMMA results show only small differences between voiced [i] and devoiced [i] (implying "devoicing", not "deletion").
- But they studied only one speaker. And only one pair of stimuli ([kite] vs. [kide]), one of whose member is a real word. 4 repetitions and no quantitative analyses.

Summary of past work

- The laryngeal data indicate that devoicing in Japanese is actively controlled.
- Acoustic evidence for presence/absence of a lingual vowel articulation has been largely equivocal, with some studies claiming the vowel has been deleted, while others claiming that it is present.
- Deletion: Beckman 1982; Beckman & Shoji 1984; Kondo 1997, 2000.
- Devoicing: Jun & Beckman 1993; Faber & Vance 2001; McCawley 1968.
- Both possible: Kawakami 1971; Tsuchida 1997; Whang 2014

Our contribution to the debate

- Electromagnetic Articulography (EMA) data on the **lingual articulation** of vowels in voiced and devoiced contexts.
- A **computational method** for evaluating vowel deletion on the basis of movement kinematics (Shaw & Kawahara, submitted).
- Showing that some cases do involve cases of actual **deletion**.
- Some implications for prosodic organization of Japanese (still in progress).

EMA Experiment

EMA experiment

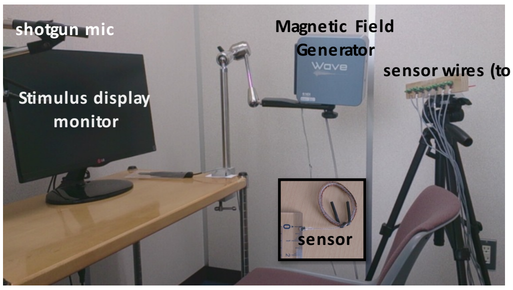
- Record the articulatory dynamics of voiced vowels and devoiced vowel counterparts.
- Apply Discrete Cosine Transform to fit the observed data.
- Use a Bayesian classifier to evaluate the likelihood that devoiced tokens have a lingual articulation.

Materials (target words)

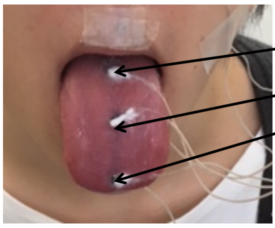
	Devoicing/deletion	Voiced vowel
Entropy = 1.99 (W,K)	masutaa マスター	masuda 益田
Entropy = 1.89 (W)	hakusai 白菜	yakuzai 薬剤
Entropy = 1.46	ʃutaisei 主体性	ʃudaika 主題歌
Entropy = 1.08 (W,K)	ɸusoku 不足	ɸuzoku 付属
Entropy = 0.09 (W,K)	katsutoki 勝つ時	katsudou 活動

10-15 repetitions of the target words in the carrier phrase: *okee* _____ *to itte* ‘Ok, say _____’. Participants were instructed to speak as if they were making a request of a friend. Interspersed with 10 words lacking /u/

Apparatus



Sensor placement



Lingual sensors

Tongue Dorsum (TD) as far back as comfortable for participant

Tongue Blade (TB) half-way between the TD and TT sensors

Tongue Tip (TT) 1 cm behind the tip

Jaw sensor

Additional sensors:
Upper lip (UL)
Lower Lip (LL)
Jaw
Nasian
Left/Right mastoids

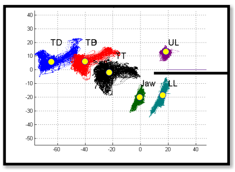
Procedure

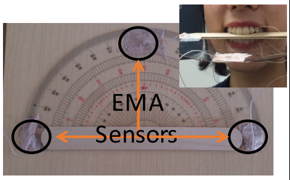
- Stimuli were presented on a monitor in random order.
- To promote fluent reading, target words were previewed before displayed in the carrier sentence.
- A native speaker of Japanese monitored pronunciation manually advancing trials after accepting or rejecting each token.

Target preview (500ms)	Target sentence
主体性	オーケー主体性と言って。

Post-processing

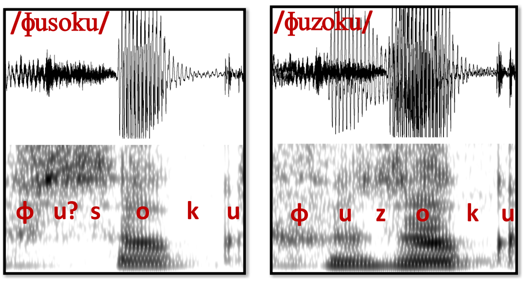
- Head movements corrected computationally
- Data rotated to the occlusal (bite) plane
- Robust smoothing (Garcia 2010)





Acoustic results: vowel devoicing

In line with current descriptions of Tokyo Japanese, /u/ was **devoiced** between voiceless consonants and voiced otherwise.

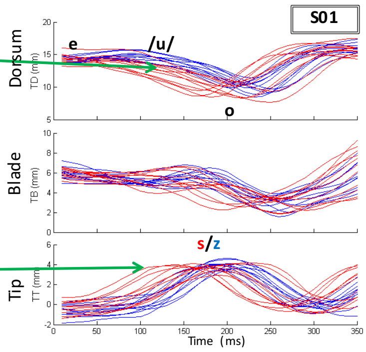


Results: ɸusoku~ɸuzoku

TD is lower for devoiced vowel

Red=devoiced
Blue=voiced

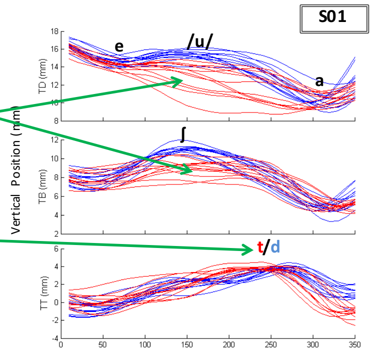
TT rise is earlier for /s/ than /z/



Results: ʃutaisei~ʃudaika

Lower TD and TB for devoiecd /u/

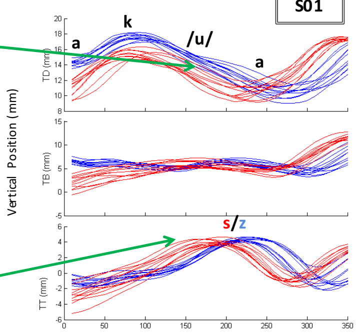
Earlier TT gesture for /t/ than for /d/



Results: hakusai~yakuzai

TD is lower for devoiced /u/

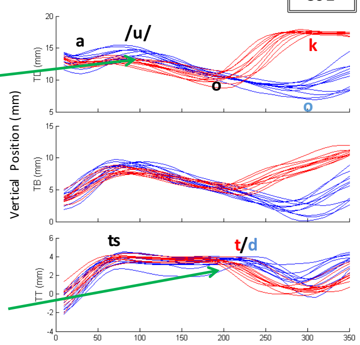
TT rise is earlier after the devoiced /u/



Results: katsutoki~katsudou

TD is lower for devoiced vowel

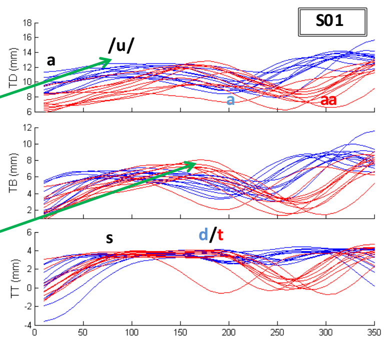
TT release is earlier after the devoiced vowel



Results: masutaa~masuda

Temporal difference in TD rise; devoiced vowel is later

And TB rise; devoiced vowel is later and higher



Computational analysis

Analysis

- **Voicing effect**: do voiced and devoiced vowel trajectories differ?
- **Phonological deletion**: is the vowel /u/ absent in any of these words?
  - Discrete Cosine Transform (DCT) on TD trajectories
  - Simulate a "targetless" trajectory
  - Compare DCT coefficients of "targetless" trajectory to data, c.f., t-test against zero.

Discrete Cosine Transform (DCT)

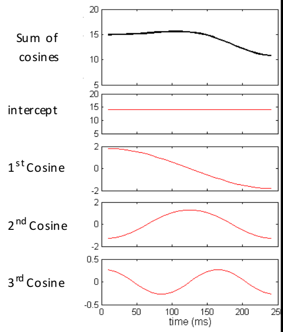
Complex curve represented as the sum of Cosines:

$C(m)\cos(\theta)$

where m is the nth cosine and θ is a function of length

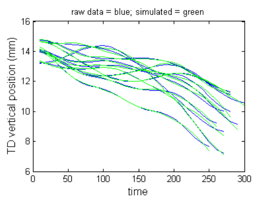
$$C(m) = \frac{2}{N} k_m \sum_{n=0}^{N-1} x(n)\cos\left(\frac{(2n+1)(m-1)\pi}{2N}\right)$$

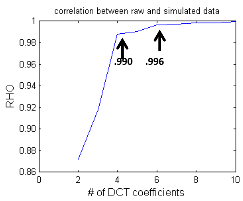
Where N is the number of data samples;
$M = 1,\dots,N$; $k_m = \frac{1}{\sqrt{2}}$ when m = 1, else $k_m = 1$;
$x(n)$ is the intercept;



How many DCT coefficients are needed?

- Nearly lossless compression (99.6%) with 6 coefficients.
- We used 4 DCT coefficients (99.0%)





Each cosine components

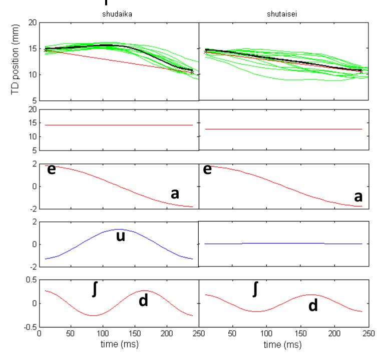
Raw data (green)
Mean DCT (black)
[e]-to-[a] line (red)

1st DCT Coefficient
→ Average TD height

2nd DCT Coefficient
→ V-to-V trajectory

3rd DCT Coefficient
→ Intervening vowel

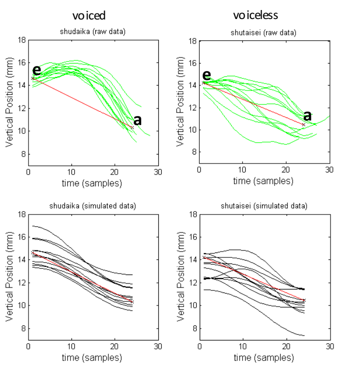
4th DCT Coefficient
→ Consonantal effects



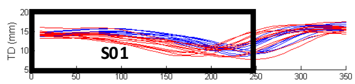
Vowel deletion ("noisy null") trajectories

raw data (green lines)
direct e-to-a trajectory (red line)

Noisy null trajectories (black lines) generated from stochastic sampling of Gaussian distributions defined by mean DCT coefficients fit to the direct e-to-a trajectory (red line) and the standard deviation of DCT coefficients fit to the raw data (green lines).

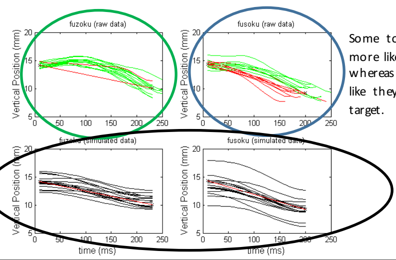


Within-speaker, within-word variation



Phonetic reduction
Or variable deletion?

Almost all tokens look pretty different from noisy null



Some tokens look more like noisy null, whereas others look like they have a clear target.

Noisy null

Token-by-token evaluation

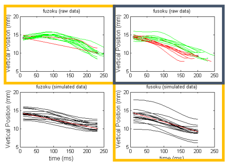
Fit a naïve Bayes classifier to the data and used it to generate (posterior) deletion probabilities

$$p(D|c_1,\ldots,c_d) = \frac{p(D)\ p(c_1,\ \ldots, c_d|D)}{p\ (c_1,\ldots,\ c_d)}$$

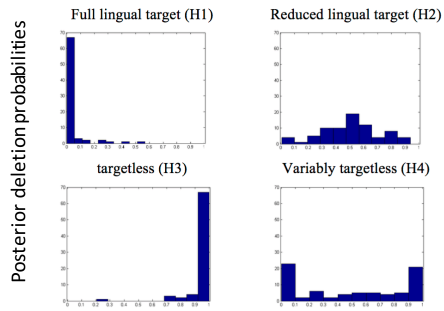
Training data = voiced tokens (full target) & noisy null (no target)
Test data = voiceless tokens

where...

D = {deletion, full vowel}
c_1 = 1^{st} *DCT Coefficient*
c_2 = 2^{nd} *DCT Coefficient*
c_3 = 3^{rd} *DCT Coefficient*
c_4 = 4^{th} *DCT Coefficient*



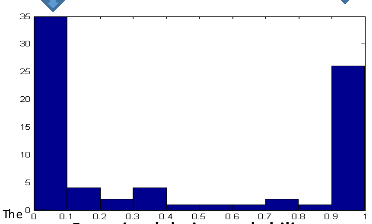
Simulated predictions



Deletion probabilities by token: /ɸusoku/ (all speakers)

Less than .1 chance of deletion

Greater than .9 chance of deletion

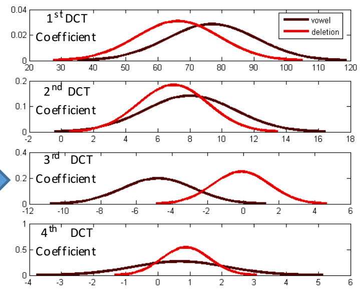


Note: DCT coefficients are speaker-specific. The results are tallied.

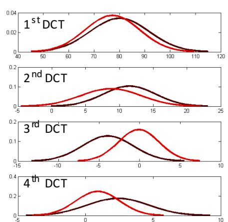
Classification parameters

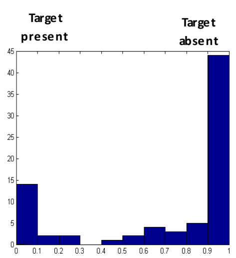
/ ɸusoku/

Greatest separation for 3^{rd} DCT coefficient



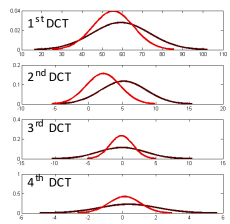
Parameters and probabilities: /ʃutaisei/ (all speakers)

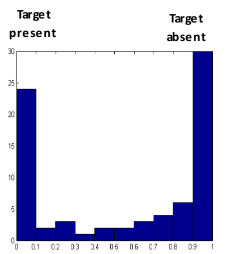




Posterior deletion probability

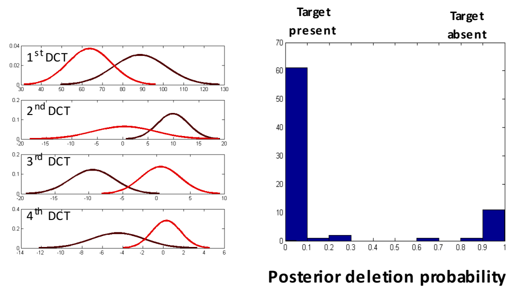
Parameters and probabilities: /katsutoki/ (all speakers)



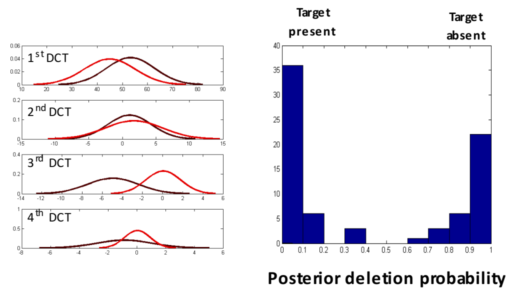


Posterior deletion probability

Parameters and probabilities: /hakusai/ (all speakers)



Parameters and probabilities: /masutaa/ (all speakers)



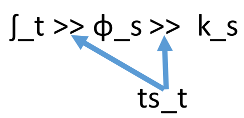
All the results

	S01	S02	S03	S04	S05	S06	average
ɸusoku	0.47	0.39	0.75	0.84	0.01	0.19	0.44
ʃutaisei	0.92	0.68	0.84	0.99	0.02	0.89	0.72
katsutoki	0.81	0.19	0.69	0.93	0.06	0.79	0.58
hakusai	0.00	0.00	0.51	0.50	0.00	0.07	0.18
masutaa	0.64	0.09	0.01	0.02	0.74	0.41	0.32
average	0.57	0.27	0.56	0.66	0.17	0.47	0.45

- /ʃutaisei/ shows the highest probability of deletion; /hakusai/ the lowest.
- S03 & S04 show high deletion probabilities; S05 barely showed deletion.

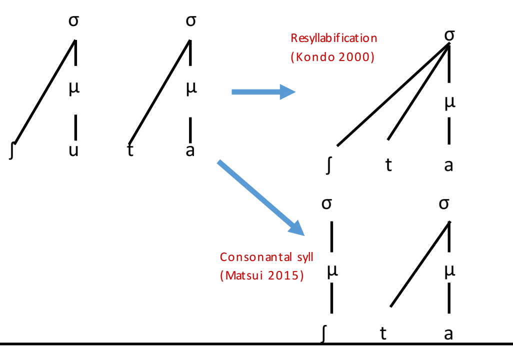
Effect of surface consonant cluster type?

- All the histograms are bi-modal, **supporting the optional deletion hypothesis**.
- For all speakers we find:



Implications for phonological organization

Possible consequences of deletion



Mora remains

- A bimoraic truncation pattern (Poser 1980 *et seq.*) counts moras of a “devoiced” vowel as moraic (Kawahara 2015; Tsuchida 1997).
- E.g. [s<u>u</u>to] < [sutoraiki] (loanword truncation)
- E.g. [tʃ<u>i</u>ka(-tʃaɴ)] < [tʃikako] (hypocoristic)
- E.g. [ɸ<u>u</u>ka-ɸ<u>u</u>ka] (mimetics)
- Hirayama (2009) shows that devoiced vowels count as much as voiced vowels for *haiku*.

Syllable remains too (?)

Ito (1990), Kawahara (2016)

Bimoraic truncation patterns

a. [de.mon.su.to.ree.ʃoɴ] → [de.mo] ‘demonstration’
b. [ri.haa.sa.ru] → [ri.ha] ‘rehearsal’
c. [ro.kee.ʃoɴ] → [ro.ke] ‘location’

Monosyllabic outputs are not allowed: a light syllable is appended

= a word must be bisyllabic; Wd much branch (I&M 1992)

a. [dai.ja.mon.do] → [dai.ja], *[dai] ‘diamond’
b. [paa.ma.nen.to] → [paa.ma], *[paa] ‘permanent (hair style)’
c. [kom.bi.nee.ʃoɴ] → [kom.bi], *[koɴ] ‘combination’
d. [ʃim.po.ʤi.u.mu] → [ʃim.po], *[ʃiɴ] ‘Symposium’

Devoiced vowels count = a devoiced vowel projects its syllable

a. [mai.ku.ro.ɸo.oɴ.] → [mai.ku], *[mai] ‘microphone’
b. [am.pu.ri.ɸai.aa] → [am.pu], *[aɴ] ‘amplifier’

Syllable remains too (?)

- Predictions from the cross-linguistic perspectives:

Rising sonority is better; a.k.a. Sonority Sequencing Principle (SSR)

Falling sonority is better; a.k.a. Syllable Contact Law (SCL)

Effect of surface consonant cluster type?

- The hierarchy we found supports the second view; i.e. that two consonants are separated by a syllable boundary.

ʃ_t >> ɸ_s >> k_s

(t)s_t   t(s)_t

- Affricates can variably be treated as an ordered fricative-stop segmental complex (Sagey 1986) or strident stop (Clements 1999) (cf. Lombardi 1990).

A caveat

- Each phonological environment was tested with one item only.
- A follow-up experiment has been run with an additional 6 speakers producing each of the below dyads 10-15 times (but the data is yet to be analyzed).

FS	FF	SS	SF
ɸuton~ɸudou （布団—不動）	ɸusoku~ɸuzoku （不足—付属）	kutakuta~kudaranu くたくた-くだらぬ	kusami~kuzai 臭み—句材
ɸutan~ɸudan （負担—不断）	ɸusai~ɸuzai （負債—不在）	kutabaru~kudasaru くたばる くださる	kusari~kuzawa 鎖 久沢
ɸuta~ɸuda （ふた—札）	ɸusagaru~ɸuzakeru ふさがる ふざける	kutaniyaki~kudanshita 九谷焼 九段下	kusakari~kuzakicho 草刈り 久崎町
ʃutaisei~ʃudaika （主体性—主題歌）	ʃusai~ʃuzai （主催—取材）	***Entropy also varied across items***	
<u>ʃutou</u>~ʃudou （酒盗—手動）	ʃusa~ʃuzan （主查-珠算）		
ʃutokou~ʃudouken 首都高 主導権	ʃuso~ʃuzou 主訴 酒造		

Temporal stability analysis

- Cross-linguistic work on the articulatory timing of consonant clusters has identified timing differences correlated with syllable structure.
- These differences are reflected in patterns of temporal stability across CVX and CCVX sequences (Browman and Goldstein, 2007; Shaw et al., 2009; Marin, 2012; Hermes et al., 2013; Shaw and Gafos, 2012).

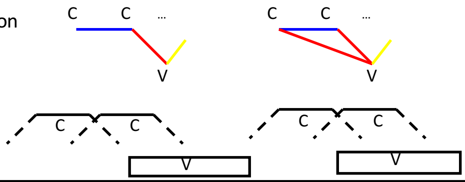
Patterns of temporal alignment

Syllable Parse	Heterosyllabic parse (simplex syllable onsets) [C.CV]	Tatuosyllabic parse (complex syllable onsets) [CCV]

On the hypothesis that the syllable nucleus is coordinated with the syllable onset… (Browman and Goldstein, 2000)

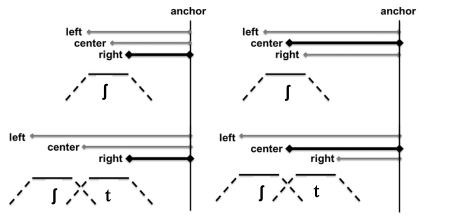
Coordination Topology

Surface Pattern



Temporal stability metrics

Following: Browman CP, Goldstein L (1988) Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45: 140–155. PMID: 3255974



Relative Standard Deviation (RSD)		
left	center	right
.12	.07	**.04**

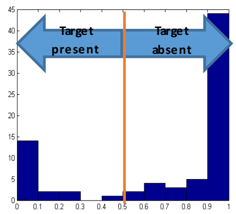
Relative Standard Deviation (RSD)		
left	center	right
.05	**.02**	.07

Singleton control words for stability analysis (already recorded)

	Consonant cluster	Singleton control
Entropy = 1.99 (W,K)	[mastaa] マスター	bataa バター
Entropy = 1.89 (W)	[haksai] 白菜	dasai ダサい
Entropy = 1.46	[ʃtaisei] 主体性	taisei 体制
Entropy = 1.08 (W,K)	[ɸsoku] 不足	kasoku 加速
Entropy = 0.09 (W,K)	[katstoki] 勝つ時	mirutoki 見る時

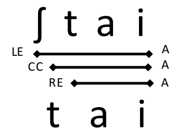
Methods

Bayesian decision rule applied to posterior probabilities

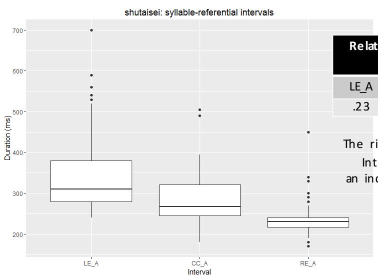


Posterior deletion probability

Target absent tokens (n = 138) were compared to singleton controls



Preliminary results: ʃutaisei



Relative Standard Deviation (RSD)		
LE_A	CC_A	RE_A
.23	.18	**.11**

The right-edge to anchor (RE_A) Interval is the most stable, an indication of simplex onsets

Data from a total of 276 tokens, 138 taisei and 138 shutaisei with <5 probability of vowel target; NB: data are from 5 speakers, as one speaker (S05) had no targetless /u/ tokens.

Conclusion

- The articulatory nature of devoiced vowels was barely known (*modulo* Funatsu & Fujimoto 2011).
- We provided first systematic analyses of the relevant articulatory data.

- Vowel is either deleted or retained, but never (or only very rarely) reduced.
- The likelihood of deletion varies across different consonantal environments, but systematically so.
- There is inter-speaker variation as well.
- Not entirely consistent with Kawakami's (1971) or Whang's (2014, 2016) predictions.
- Not clear effects of consonant-conditioned entropy either.

- We conjectured that our results support the heterosyllabic analysis of the "resulting clusters" (Matsui 2005).
  - Some phonological evidence
  - Syllable contact effect
  - Temporal stability patterns

- But more needs to be done, especially to confirm that Japanese follows Syllable Contact Law (the emergence of the unmarked?).

Acknowledgements

- Research supported by the **JSPS** grants (#26770147 and #26284059, and especially #15F15715, which supported our collaboration).

- Thanks to Jeff Moore and Chika Takahashi for their help with the experiment and EMA analysis.