# Consequences of high vowel deletion for syllabification in Japanese

Annual Meeting of Phonology 2017
@NYU

*Jason Shaw & Shigeto Kawahara*
*Yale University & Keio University*

Yale University

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

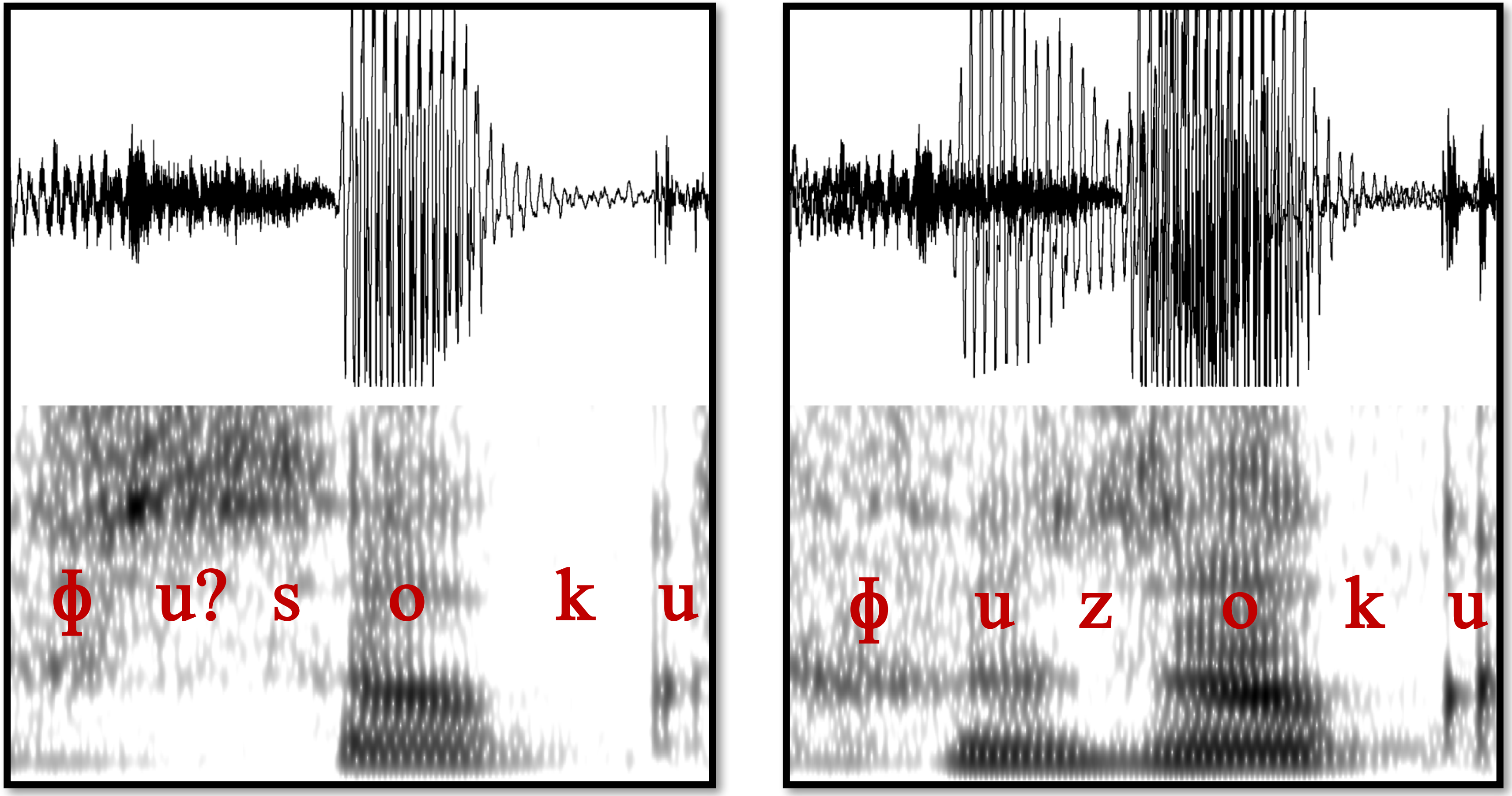
# Introduction

# Introduction

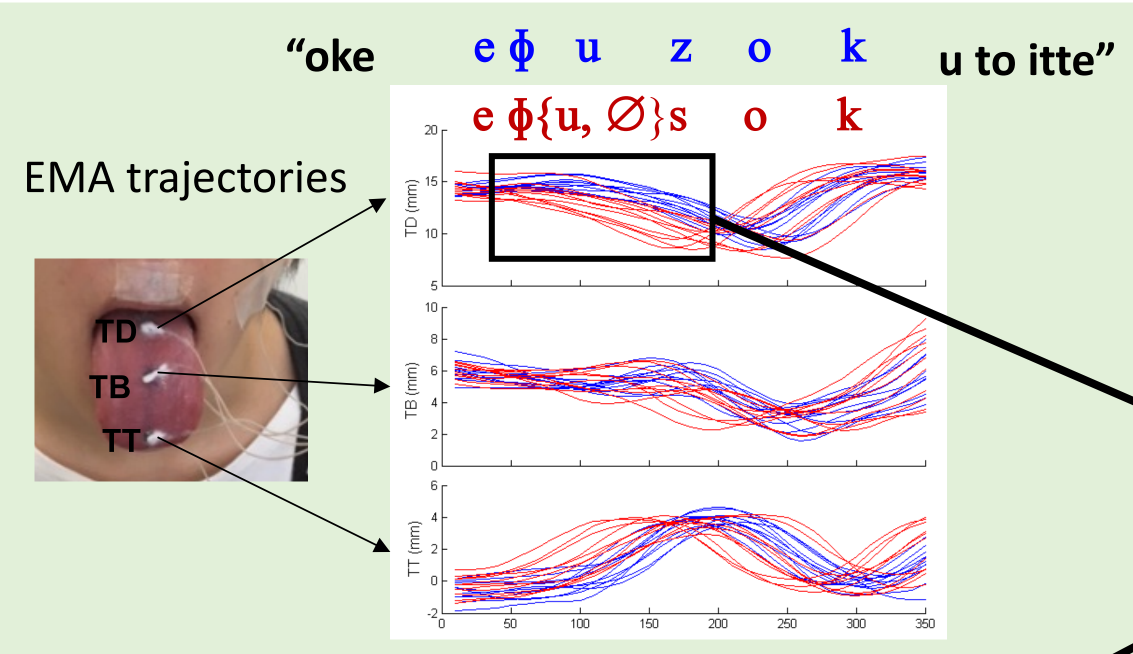
- Japanese is known to be a language without consonant clusters.
  - ➢ Epenthesis  breaks up consonant clusters in loans:
  “strike” => [sutoraiku]; Wurmbrand => [urumuburando]
  - ➢ “Perceptual epenthesis” in French/Portuguese clusters, e.g., *ebzo* (Dupoux et al., 1999; 2011)

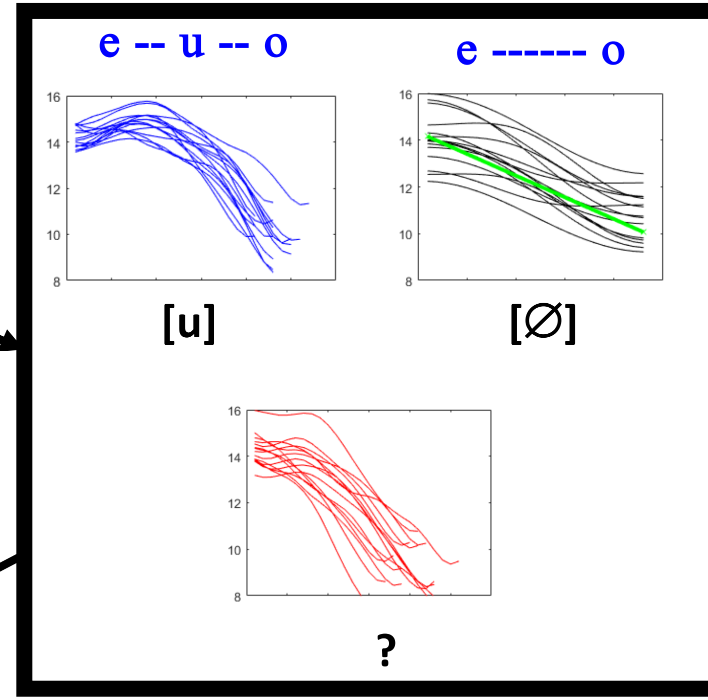
- However, Japanese is also known to devoice high vowels, which result in apparent consonant clusters (Beckman 1982; Beckman & Shoji 1984; Kawakami 1977; Kondo 1997; Matsui 2017; Whang 2014)

# *Acoustically*, there is **no vowel**.

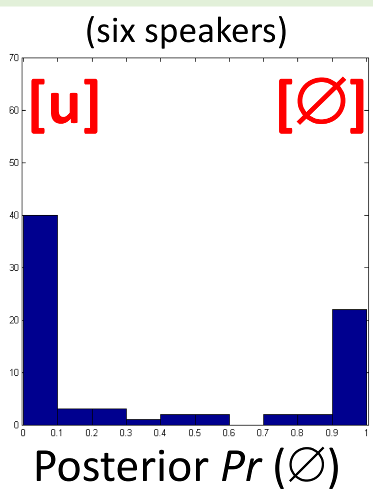


# *Articulatorily*, there are **vowel ~ ∅** alternations

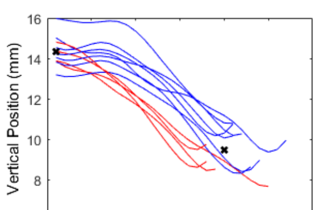



Bayesian classification of devoiced trajectories




Largely categorical results—most tokens either [u] or [∅]




Shaw, J. A., & Kawahara, S. (not yet rejected). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics*, 62 pgs.

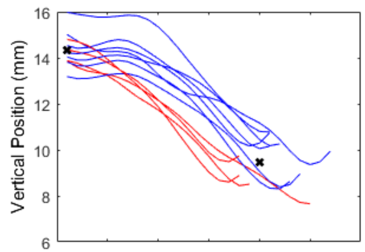
# Main question

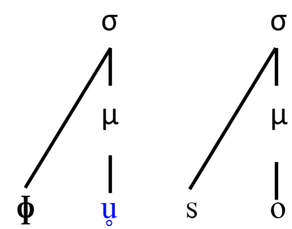
- What are the consequences of **high vowel deletion** for **syllabification** in Japanese?

- Two lines of evidence:
  1. Phonological processes sensitive to syllable structure: prosodic truncation, pitch accent placement (e.g., Ito 1990; Kubozono 2011)
  2. Patterns of temporal stability in speech production (e.g., Browman and Goldstein 1988; Shaw et al. 2009)

# Assumptions and hypotheses

- We assume full vowel targets are parsed as Cu̥.CV

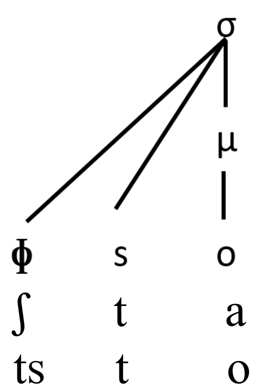



- Two hypotheses for vowel deletion cases:

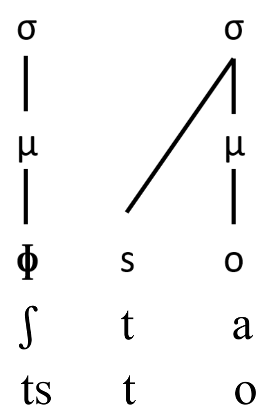
**$H_1$ Re-syllabification**

(Kondo 1997)




**$H_2$ Syllabic consonant**

(Matsui 2017)



# Phonological considerations

# Evidence that the mora remains

- For the purposes of bimoraic truncation (Poser 1980 *et seq*.) "devoiced" vowels count (Kawahara 2015; Tsuchida 1997).

  E.g. [su̥to] < [su̥torɑiki] (loanword truncation)

  E.g. [tʃi̥kɑ(-tʃɑɴ)] < [tʃi̥kɑko] (hypocoristic)

  E.g. [ɸu̥kɑ-ɸu̥kɑ] (mimetics)

- Devoiced vowels also count in *haiku* (Hirayama 2009)

# Syllable remains too: truncation

Bimoraic truncation patterns (Ito 1990; Kawahara 2016)

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

PrWd must branch

(Ito and Mester 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 a syllable

a. [mai.ku.ro.ɸo.oɴ.] → [mai.ku], *[mai] 'microphone'

b. [am.pu.ri.ɸai.aa] → [am.pu], *[aɴ] 'amplifier'

# Syllable remains too: Accent

- Japanese default accent pattern is Latin Stress Rule (Kubozono 2011).
  - Place accent on the penultimate syllable if it is heavy: [fu-re’n-do] “friend”.
  - Otherwise place accent on the antepenultimate syllable: [re’-ba-non] “Lebanon”.

- Devoiced syllables do not disrupt the Latin Stress Rule:

  [bu.ra’n.ku̥] “blank” cf. [fu-re’n-do]
  [pu.ro’.se.su̥] “process” cf. [re’-ba-non]

Kubozono, H. (2011). Japanese pitch accent. *The Blackwell companion to phonology*, *5*, 2879-2907.

# Phonological evidence favors $H_2$

- Both moras and syllables remain for devoiced high vowels; **evidence favors $H_2$**

$H_1$ Re-syllabification
(Kondo 1997)

$H_2$ Syllabic consonant
(Matsui 2017)

*Higher level moraic and syllabic structure appear unperturbed by vowel devoicing/deletion*

# Temporal stability analysis

# Temporal stability analysis

- Cross-linguistic work on the **articulatory timing of consonant clusters** has shown that timing differences sometimes correlate with **syllable structure**.

- These differences are reflected in patterns of temporal stability across CVX and CCVX sequences (Browman and Goldstein 2007; Hermes et al. 2013, 2017; Marin and Pouplier 2010; Marin 2012; ; Shaw et al. 2009; Shaw and Gafos 2015).

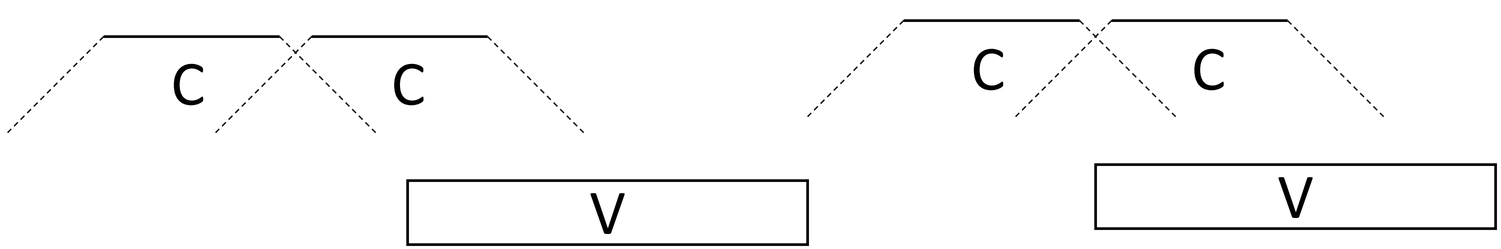
# Patterns of temporal alignment

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

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

Coordination Topology

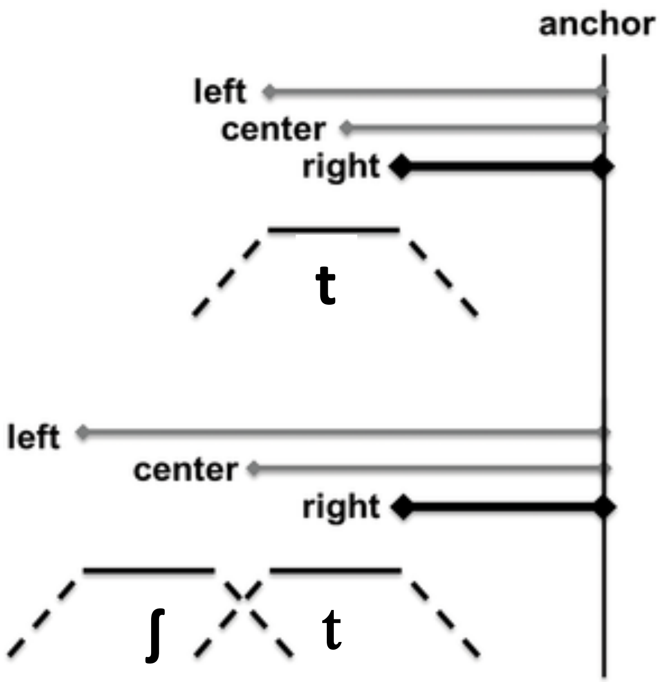

Surface Pattern



# Temporal stability metrics

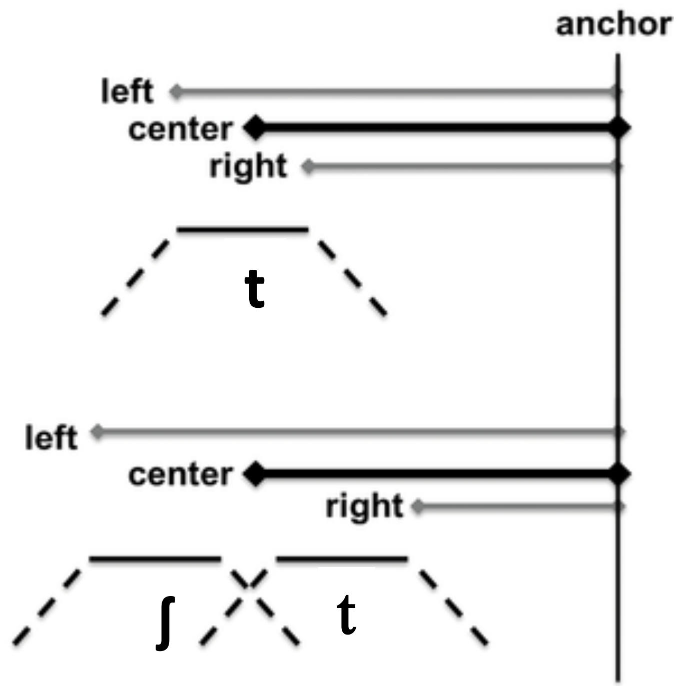
Following: Browman CP, Goldstein L (1988) Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45: 140–155. PMID: 3255974

## [C.CV]




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

## [CCV]




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

# Experimental stimuli

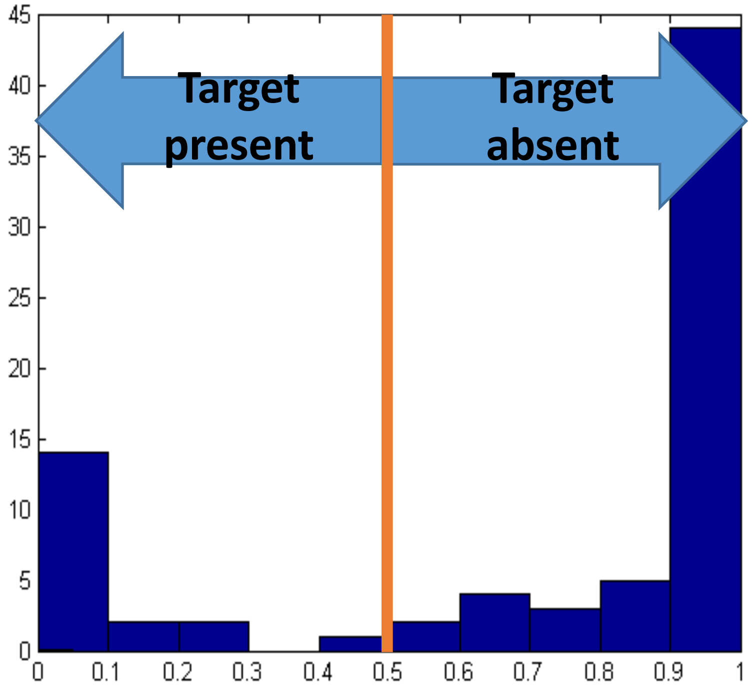
| Voiced V | Deleted V | Control |
|---|---|---|
| [masuda] | [masta:] | [bata:] |
| [yakuzai] | [haksai] | [dasai] |
| [ʃudaika] | [ʃtaise:] | [taise:] |
| [ɸuzoku] | [ɸsoku] | [kasoku] |
| [katsudo:] | [katstoki] | [mirutoki] |

**Procedure**: six native speakers of Tokyo Japanese (3 male) read items in a carrier phrase "okee___to itte"; items were randomized in a block; 10-15 blocks were recorded

# Analysis

**(1) Classify trajectories as [u] or ∅**

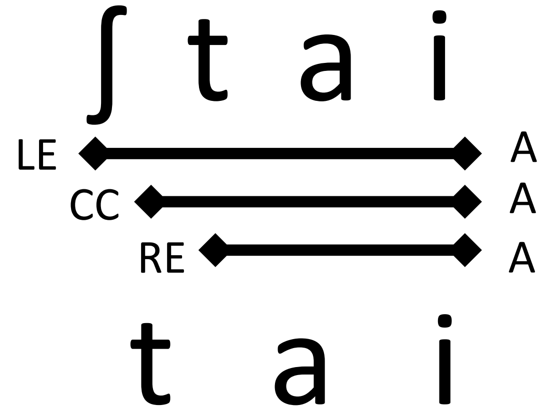
Bayesian decision rule applied to posterior probabilities



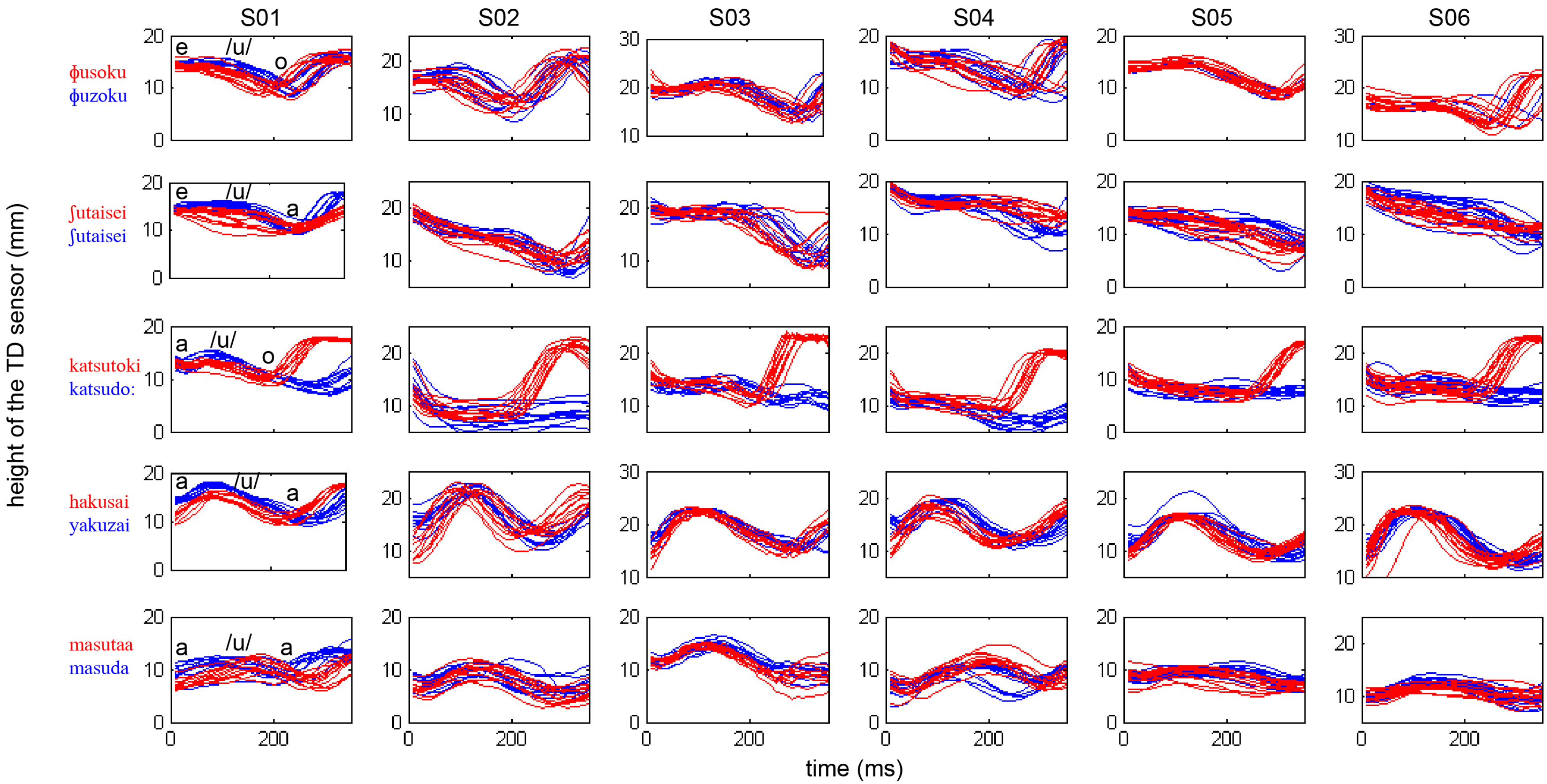

**Posterior *Pr* (∅)**

**(2) Stability analysis of ∅ tokens**

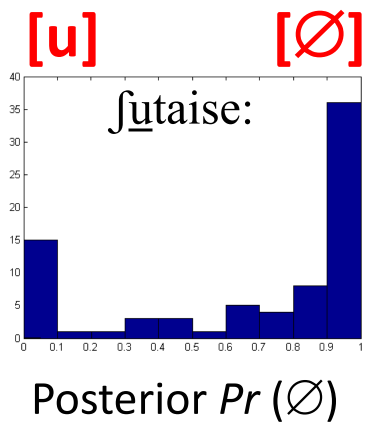
**Target absent** tokens (n = 138) were compared to singleton controls (n = 138)

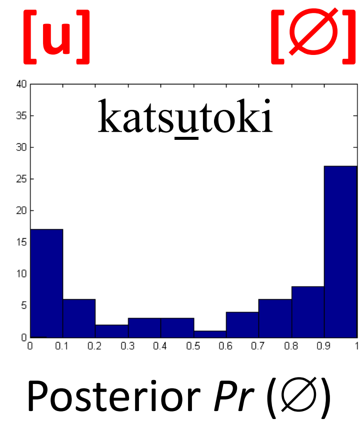


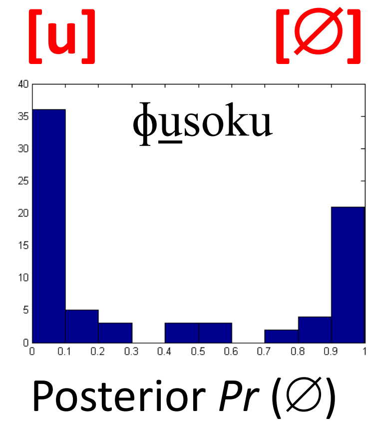
# Raw data: TD trajectories by subject and item

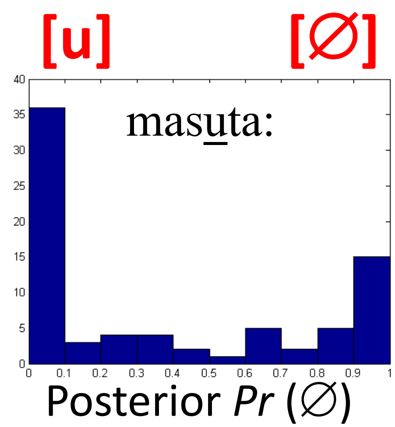


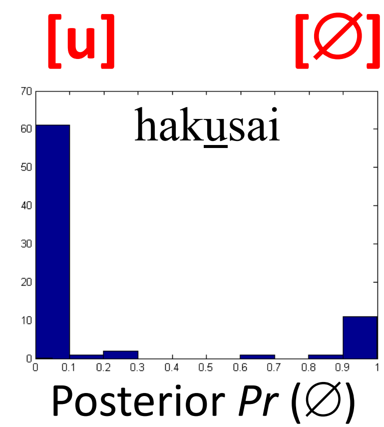
# Classification results



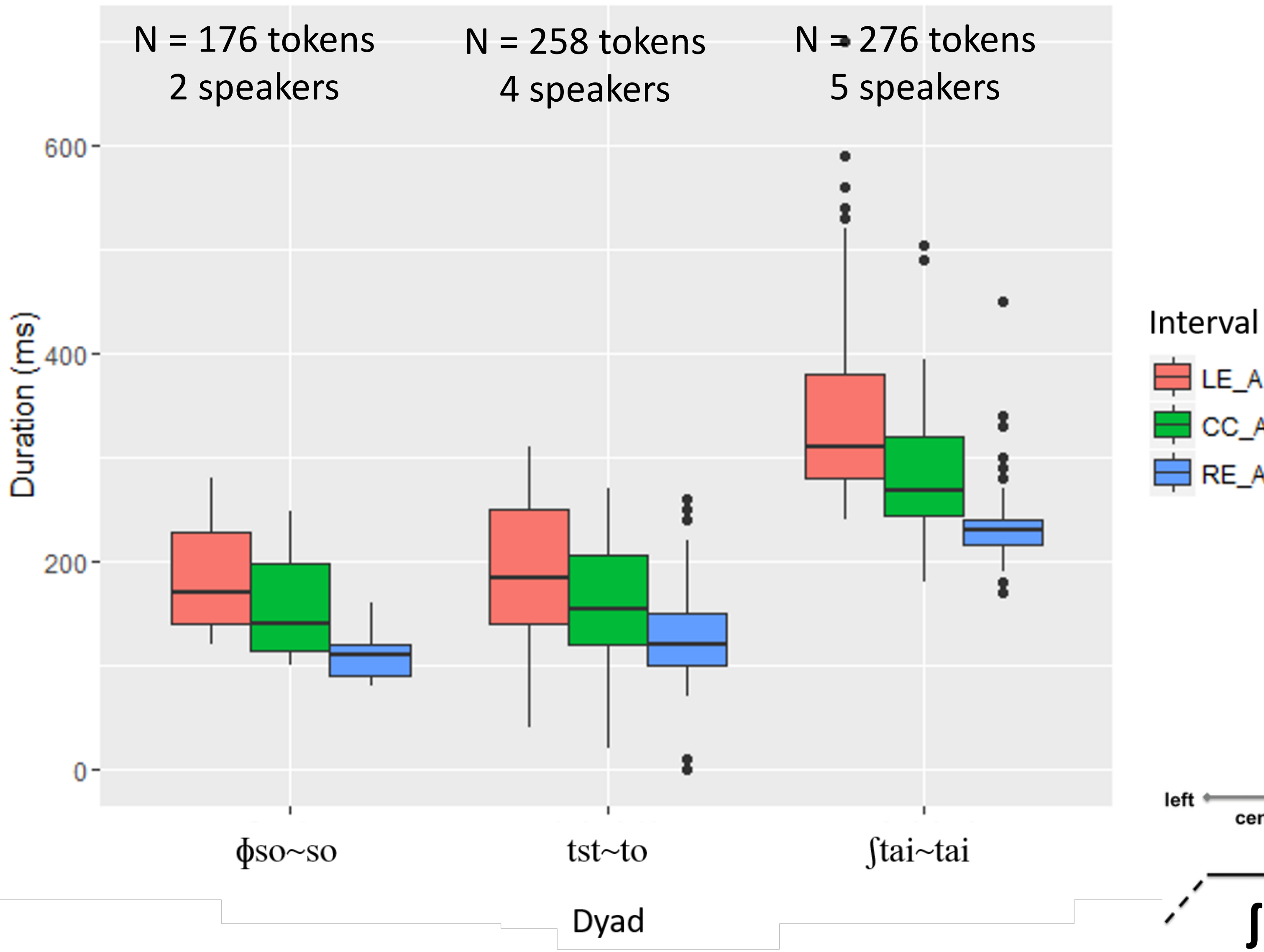



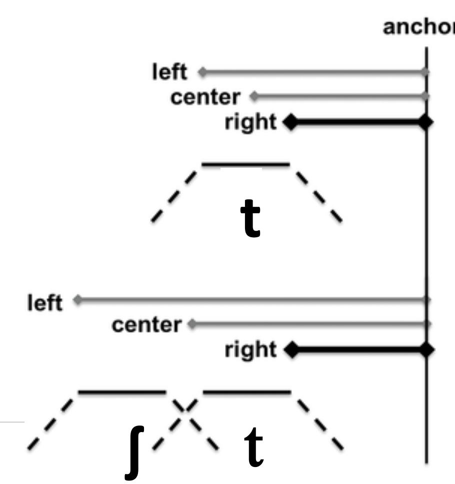



# Syllable-referential intervals



# Stability analysis

| target dyad | Relative Standard Deviation (RSD) | | |
|---|---|---|---|
| | LE_A | CC_A | RE_A |
| [ɸso]~[so] | 0.32 | 0.34 | **0.24** |
| [tsto]~[to] | 0.25 | 0.23 | **0.20** |
| [ʃtai]~[tai] | 0.23 | 0.28 | **0.11** |

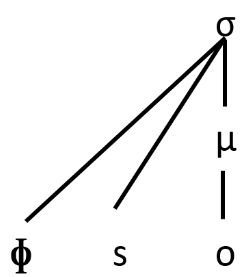
The right-edge to anchor (RE_A) Interval is the most stable, an indication of simplex onsets

# Phonetic evidence favors $H_2$

- The right edge to anchor interval is more stable than the center-to-anchor interval (no c-center effect) and the left edge-to-anchor interval.

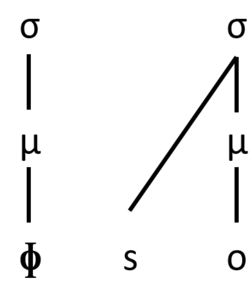
$H_1$ Re-syllabification

(Kondo 1997)




$H_2$ Syllabic consonant

(Matsui 2017)



# Discussion

# Summary

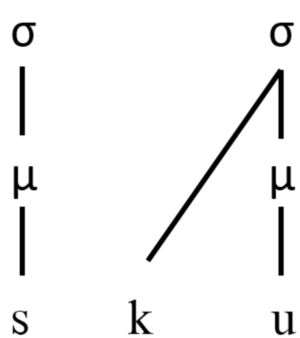
- Japanese /u/ **optionally deletes** in devoicing environments, yielding consonant clusters.
- This fact has not been known until now because devoicing obscures the acoustic consequences of the lingual gesture and articulatory data haven't been available (*modulo* Funatsu & Fujimoto 2011)
- The evidence reviewed (some phonological, some phonetic) is consistent with a **heterosyllabic parse of clusters** resulting from /u/ deletion.

# Caveats

- /u/ deletion is variable but corresponding variability in relevant bimoraic truncation has not been reported…
  - we don’t know whether, e.g., there is always a lingual gesture in [su̥to] < [su̥torɑiki]
  - Deletion may be blocked by prosodic requirements
- The status of devoiced vowels between consonants may be different from those in word-final position (Kilbourn-Ceron and Sondreggor 2017)

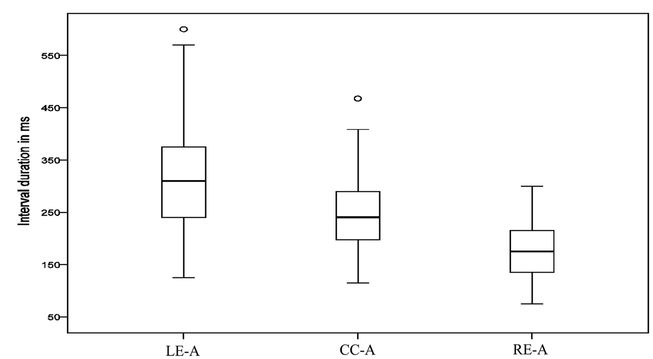
# Moroccan Arabic and Tashlhiyt Berber

- To the extent that the syllabic consonant analysis of Japanese is supported, it resembles synchronic analyses of Moroccan Arabic (and Tashlhiyt Berber).
- MA is particularly relevant as word-initial clusters arose from the loss of a short vowel (e.g., Benhallam 1980)



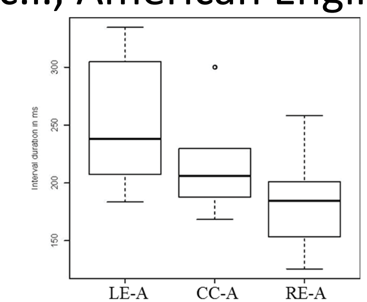

Moroccan Arabic

(Shaw and Gafos 2015)




Evidence from phonotactic patterns, vowel-glide alternations, and prosodic templates in oral verse (Dell and Elmedlaoui 2002)

c.f., American English




Also Berber vs. Polish: Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics*.

# Broader issues and future work

- Factors conditioning optional deletion
  - ➢ frequency of /u/ deletion varies across items: **ʃt** > tst > **ɸs** > st > ks
  - ➢ **How is this learned?** Auditory cues required to learn probabilities from the input are unavailable.
  - ➢ Grammatical factors (**Emergence of the unmarked**)
  - ➢ Phonetic factors: /u/ may be squeezed out by the shared laryngeal gesture & laryngeal-oral timing requirements of flanking consonants (“articulatory binding” Kingston 1990).

- Our data present a case (like compensatory lengthening) in which **prosodic and temporal stability** are maintained despite **segmental variability** (deletion).
  - ➢ Independent representations of timing and articulation, which may have a neural basis (Long et al., 2016)
  - ➢ C.f., evidence for close interaction between prosodic rhythm and segmental articulation (Tilsen 2009).

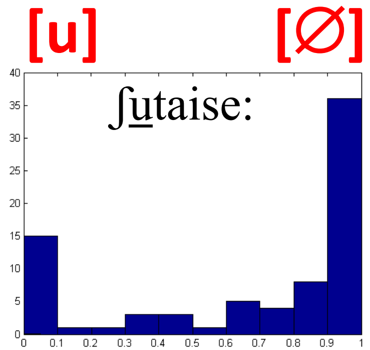
Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., . . . Greenlee, J. D. (2016). Functional segregation of cortical regions underlying speech timing and articulation. *Neuron, 89*(6), 1187-1193.

# Acknowledgements

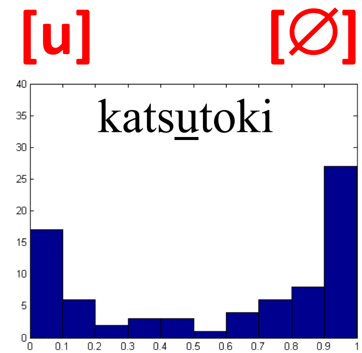

- Research supported by the **JSPS** grants (#26770147 and #26284059, and especially #15F15715).

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

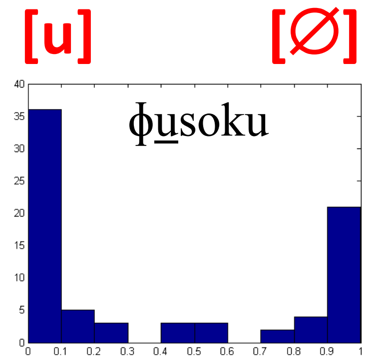
# Questions?



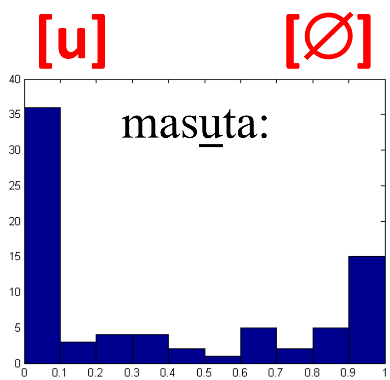

Posterior *Pr* (∅)



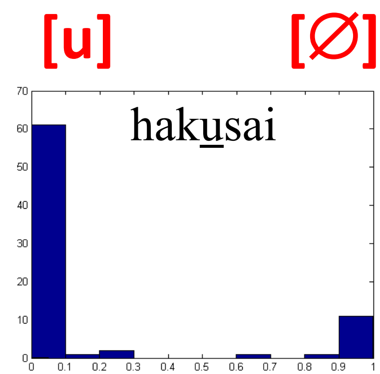

Posterior *Pr* (∅)




Posterior *Pr* (∅)




Posterior *Pr* (∅)




Posterior *Pr* (∅)

# Experimental stimuli: accent

| Voiced V | Deleted V | Control |
|---|---|---|
| [ʃuda’ika] | [ʃtaise:] | [taise:] |
| [ɸuzoku] | [ɸsoku] | [kasoku] |
| [katsudo:] | [ka’tstoki] | [mi’rutoki] |
| [masuda] | [ma’sta:] | [ba’ta:] |
| [yaku’zai] | [hak’sai] | [dasa’i] |

# Matsui’s (2017) observation

Groove for [s] extends throughout the syllable.

EPG data from Matsui (2017), Journal of the Phonetic Society of Japan.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 39 | 22 | 0 | 0 | 0 | 47 | |
| 100 | 100 | 69 | 0 | 0 | 0 | 23 | 80 |
| 100 | 61 | 26 | 0 | 0 | 0 | 0 | 100 |
| 100 | 52 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 19 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 19 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 13 | 0 | 0 | 0 | 0 | 0 | 93 |
| 100 | 93 | 0 | 0 | 0 | 0 | 0 | 31 |

/s/

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 31 | 14 | 0 | 0 | 0 | 47 | |
| 100 | 100 | 58 | 0 | 0 | 0 | 19 | 76 |
| 100 | 58 | 21 | 0 | 0 | 0 | 0 | 100 |
| 100 | 48 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 14 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 11 | 0 | 0 | 0 | 0 | 0 | 100 |
| 100 | 6 | 0 | 0 | 0 | 0 | 0 | 90 |
| 100 | 86 | 0 | 0 | 0 | 0 | 0 | 22 |

/ɯ̥/

# Shared laryngeal gesture

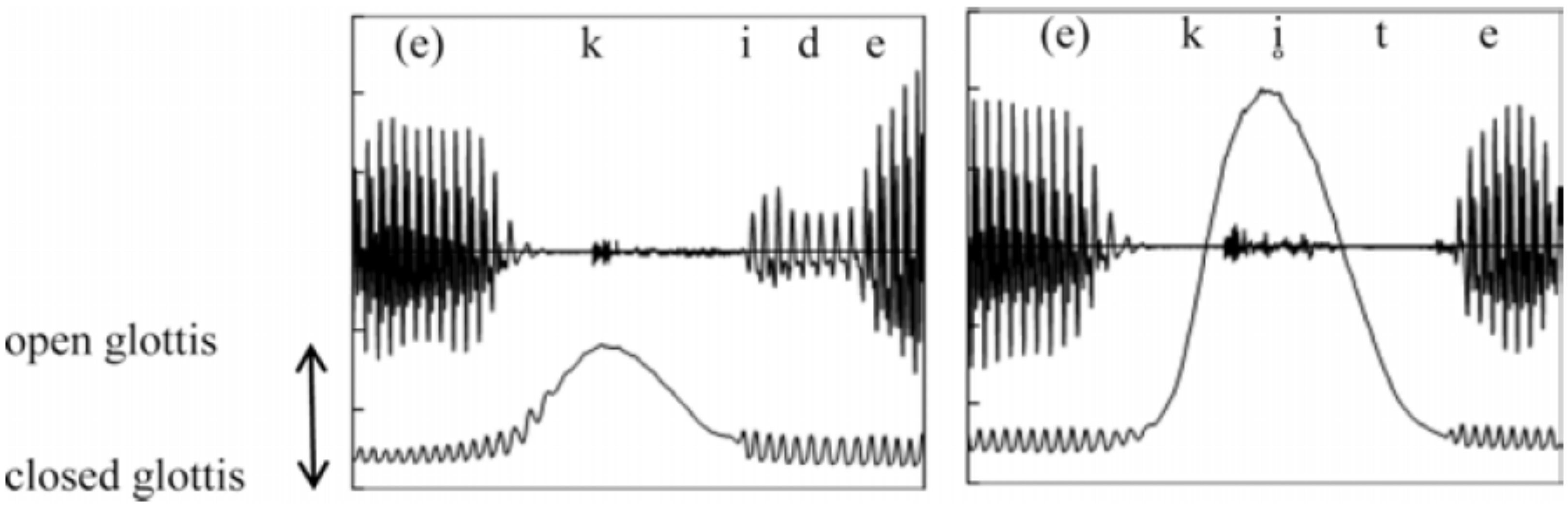
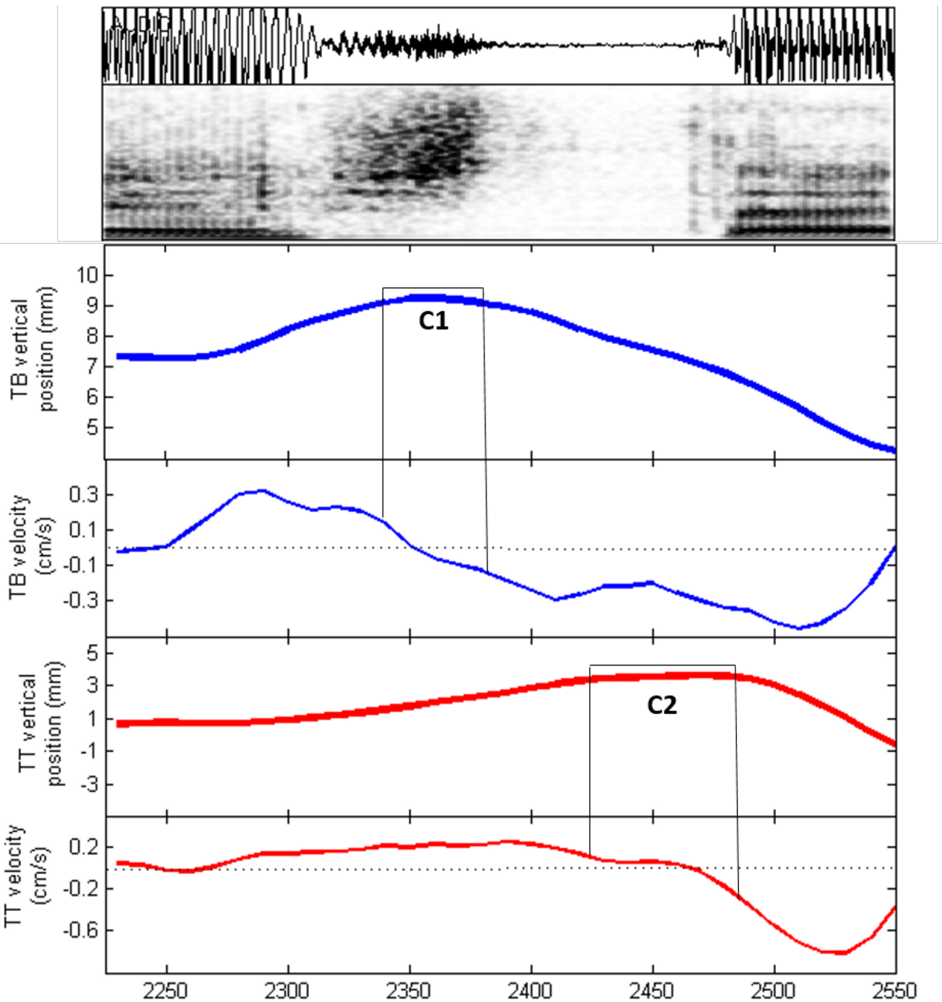


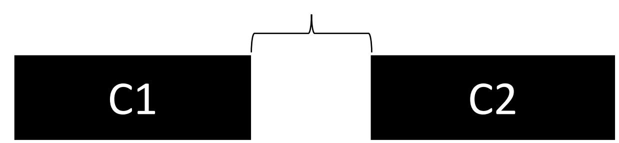

**Figure 1: The degrees of glottal abduction in Japanese. The left panel: a voiceless stop followed by a voiced stop, which has a single abduction gesture for /k/. The right panel: a voiceless stop /k/ followed by a voiceless vowel and another voiceless stop /t/, which also has a single abduction gesture. The magnitude of the abduction gesture in the right panel is larger than twice the size of the abduction gesture in the left panel. Taken from Fujimoto et al. (2002), cited and discussed in Fujimoto (2015).**

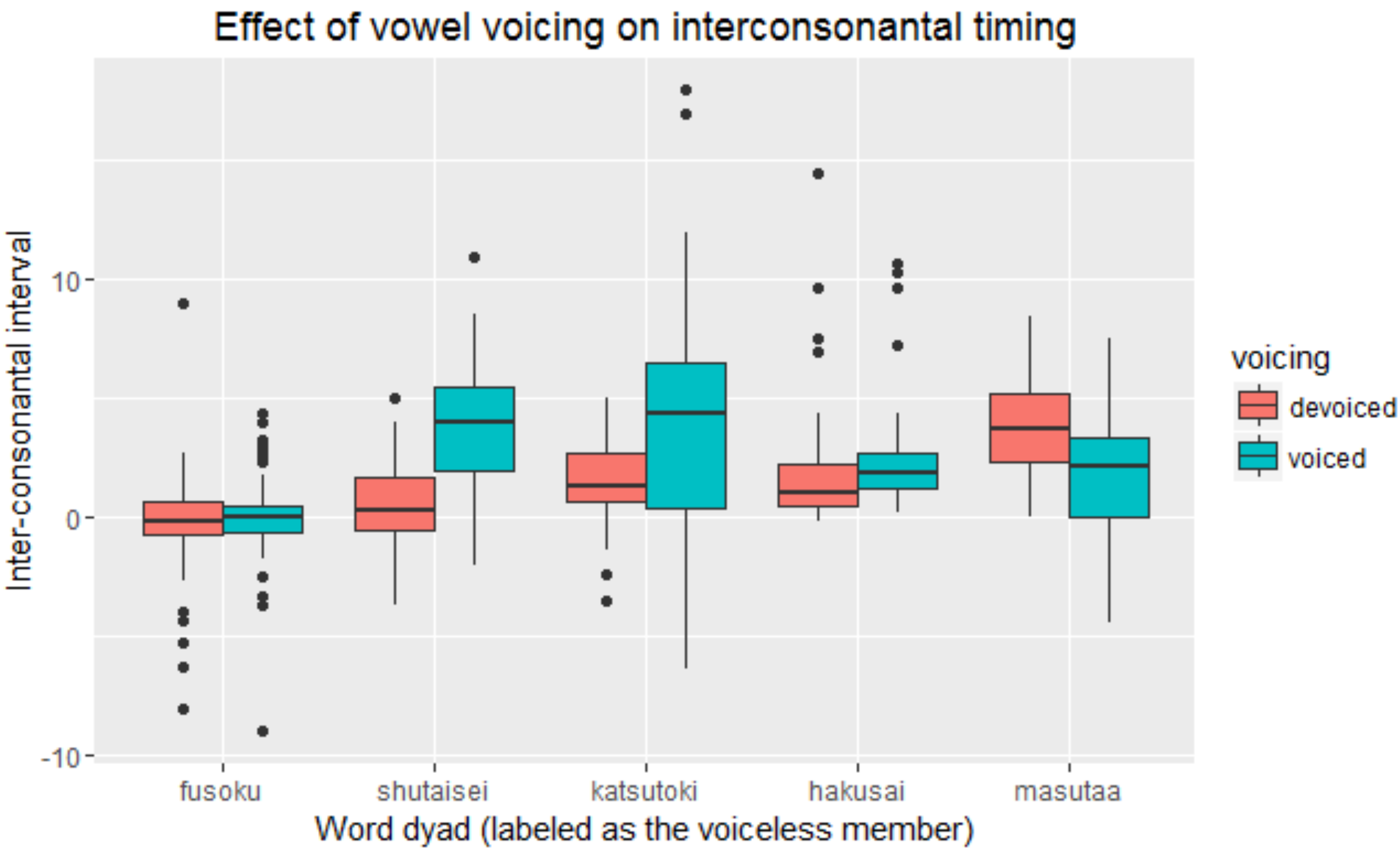
# Parsing gestures for stability analysis



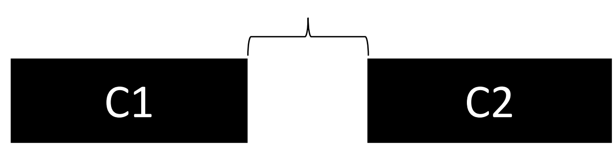

Inter-consonantal Interval (ICI) extends from release of C1 to the target of C2

Inter-consonantal interval



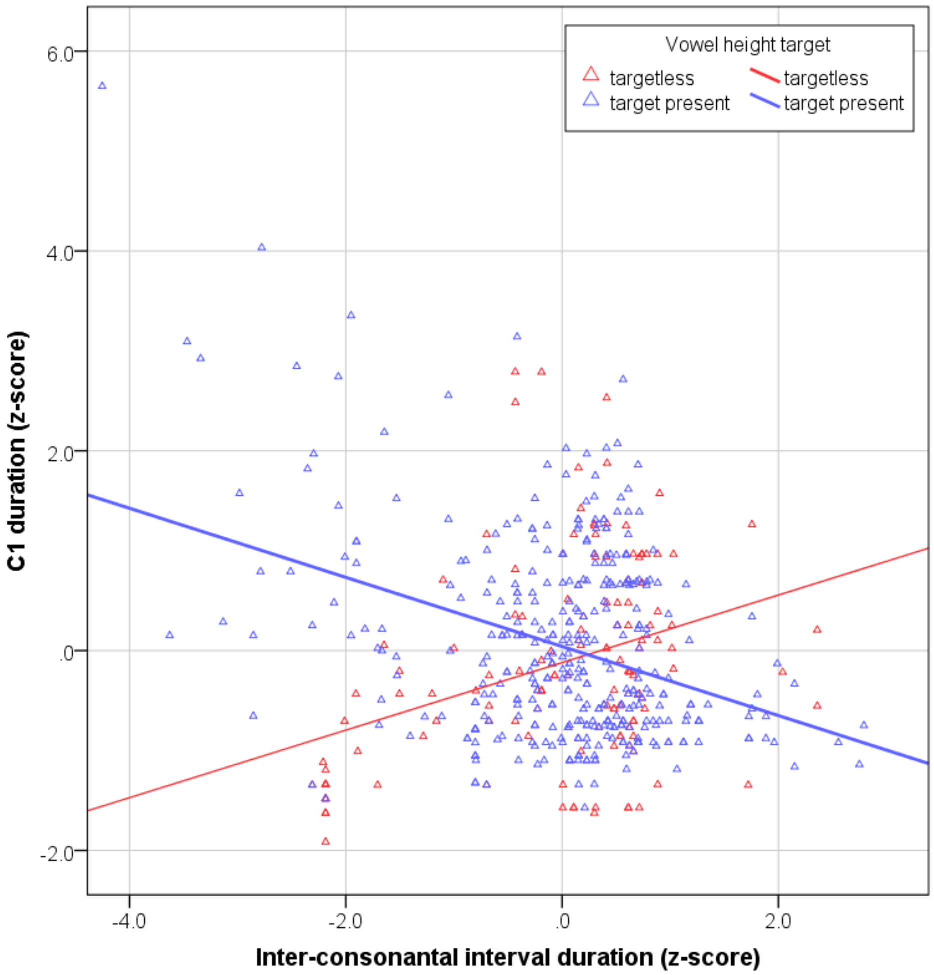



Inter-consonantal interval (ICI)

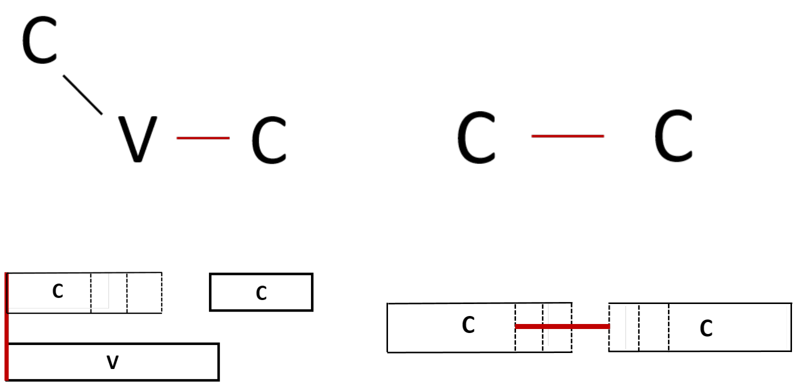



NB: */u/ deletion did not have a significant effect on ICI*

# C-V vs. C-C timing




*As C1 decreases, ICI increases, but only for tokens that contain a vowel target.*



# Discrete Cosine Transform (DCT)

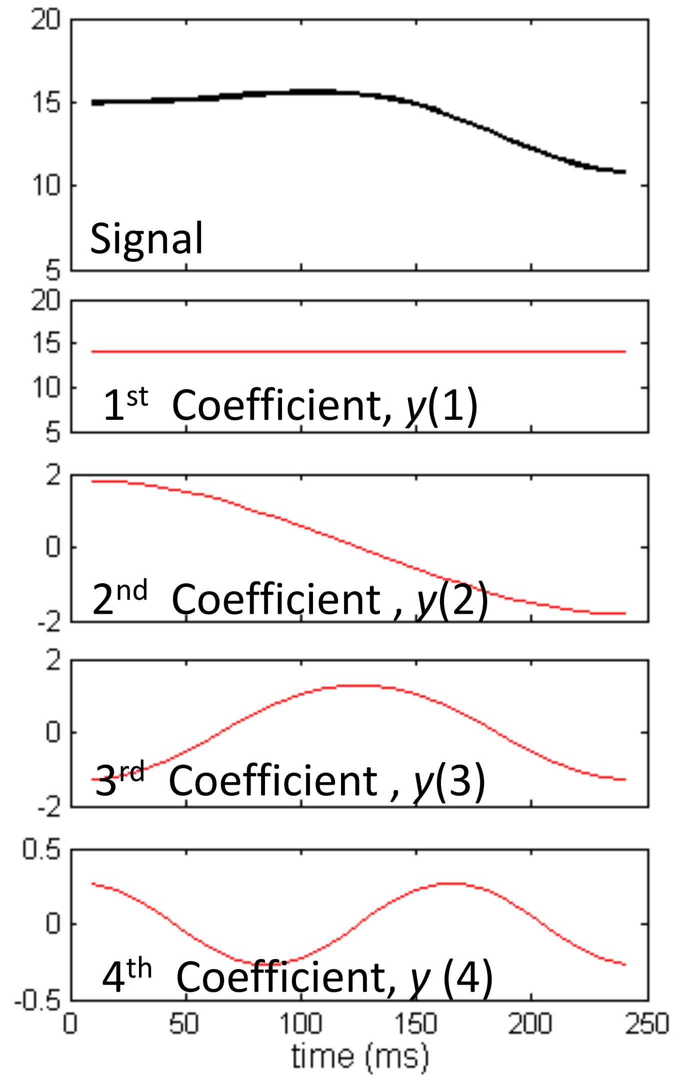
**Complex curve represented as the sum of Cosines:**

$$y(k) = w(k) \sum_{n=1}^{L} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)$$

$$k = 1,2, \ldots L$$

Where $L$ is the number of data samples and ***x(n)* is the trajectory to be modelled** and:

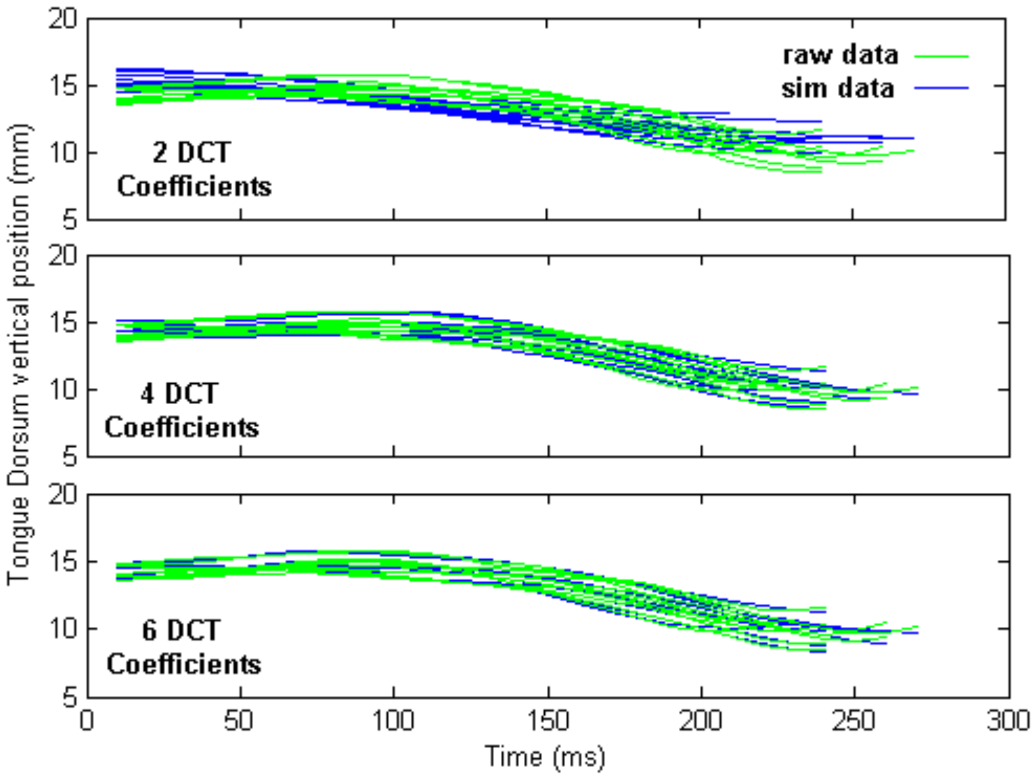
$$w(k) = \begin{cases} \frac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\frac{2}{L}} & 2 \le k \le L \end{cases}$$

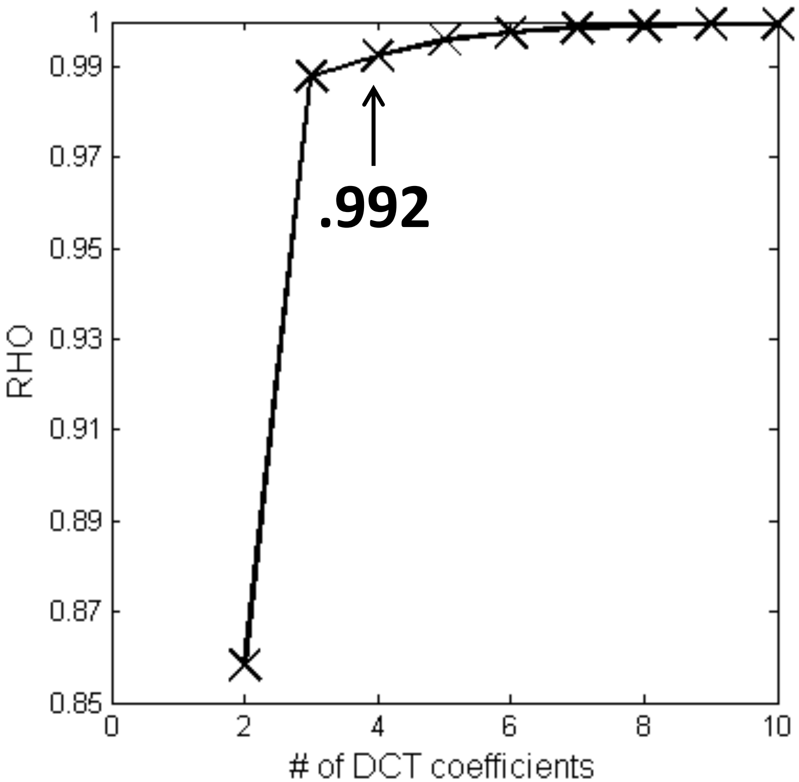



Rao, K. R., & Yip, P. (2014). *Discrete cosine transform: algorithms, advantages, applications*. Academic press.

# How many DCT coefficients?

- Real space signals can be represented to an arbitrary degree of precision;
- Nearly lossless compression (*r* = .992) with 4 coefficients.



# Interpretation of cosine components

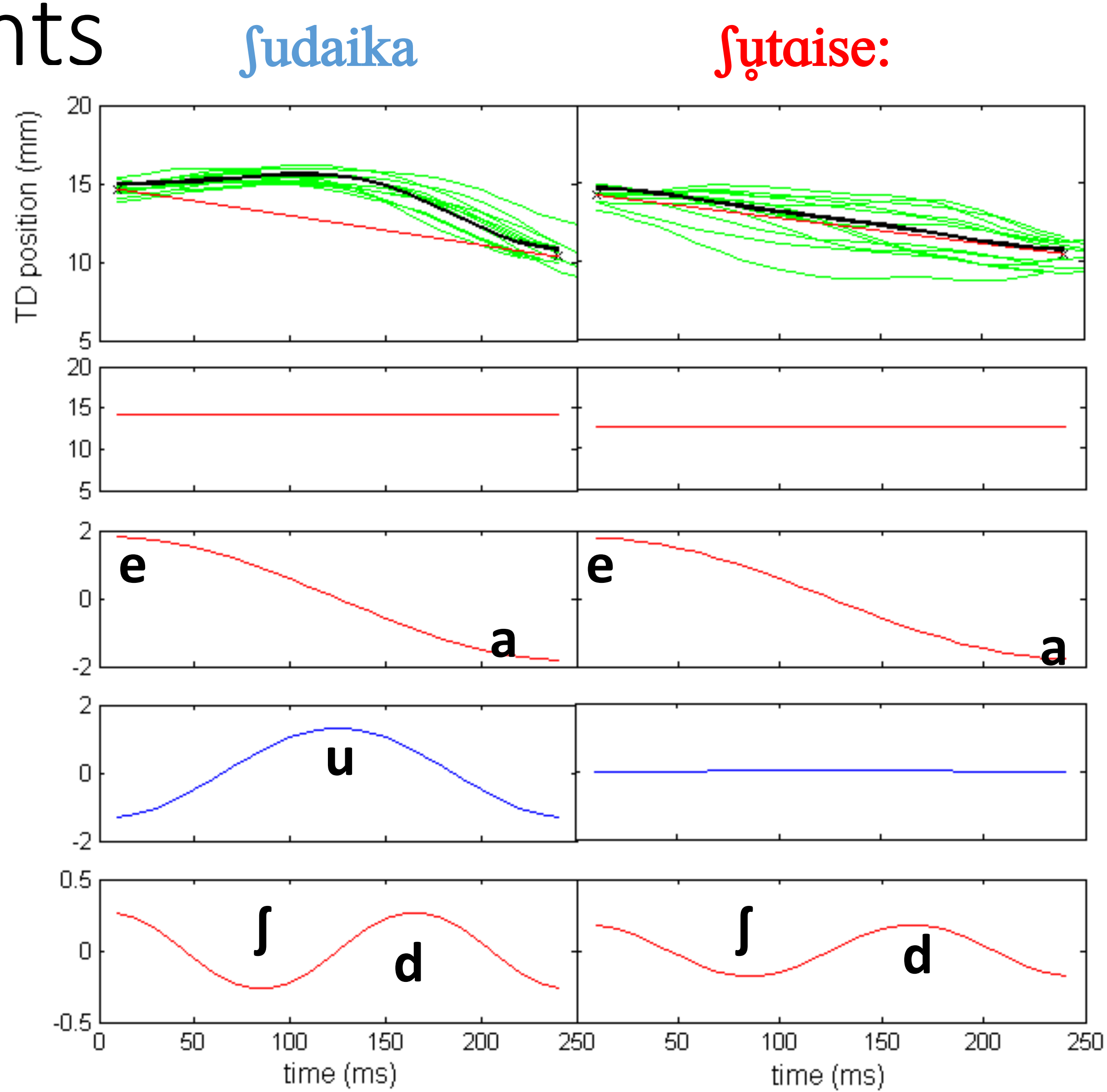
Raw data (green)
Mean DCT (**black**)
[e]-to-[a] line (**red**)

1st DCT Coefficient
→ **TD height**

2nd DCT Coefficient
→ **V-to-V trajectory**

3rd DCT Coefficient
→ **Intervening vowel**

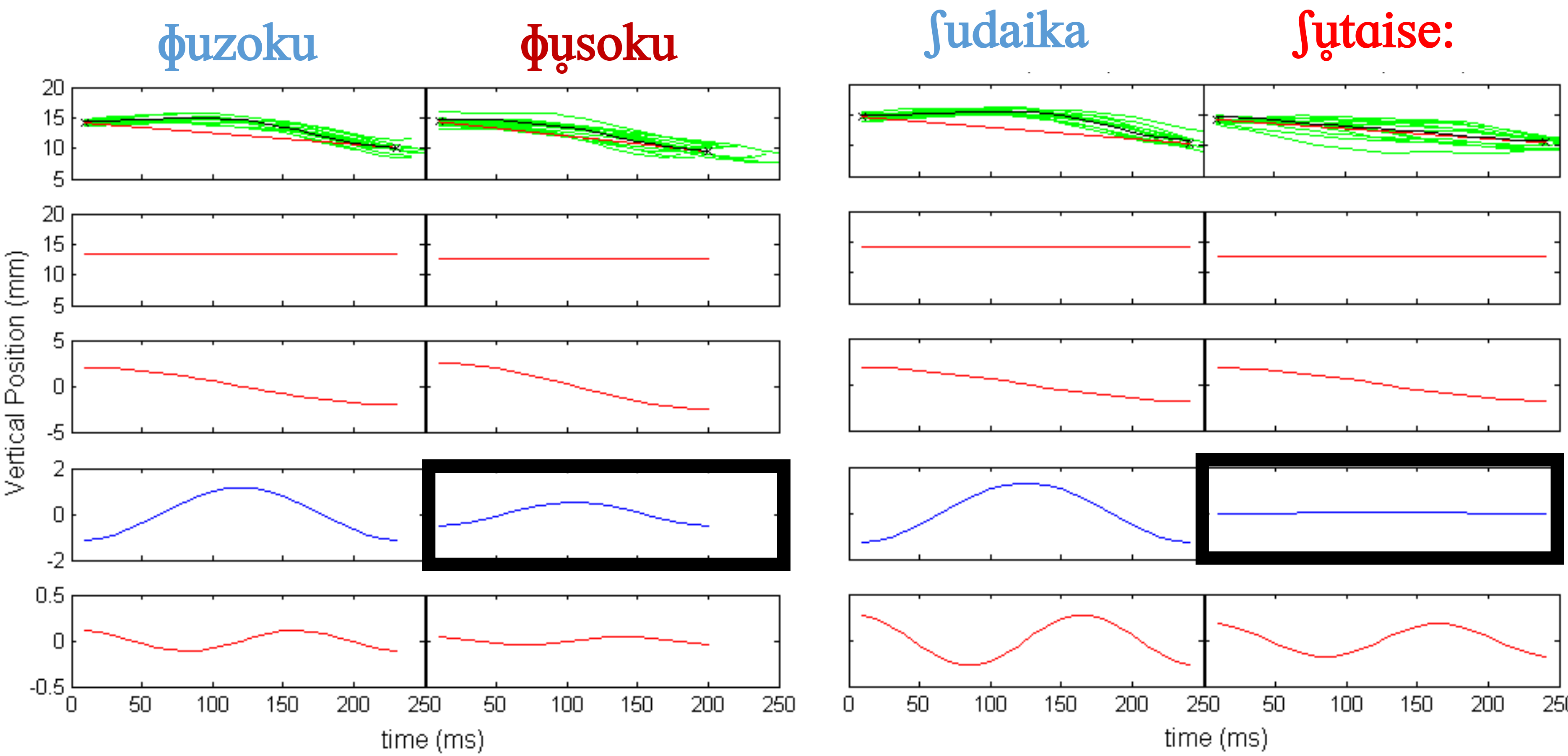
4th DCT Coefficient
→ **Coarticulation**



# Reduced or targetless: the view from DCT components

## Phonetic reduction?

## Targetless?

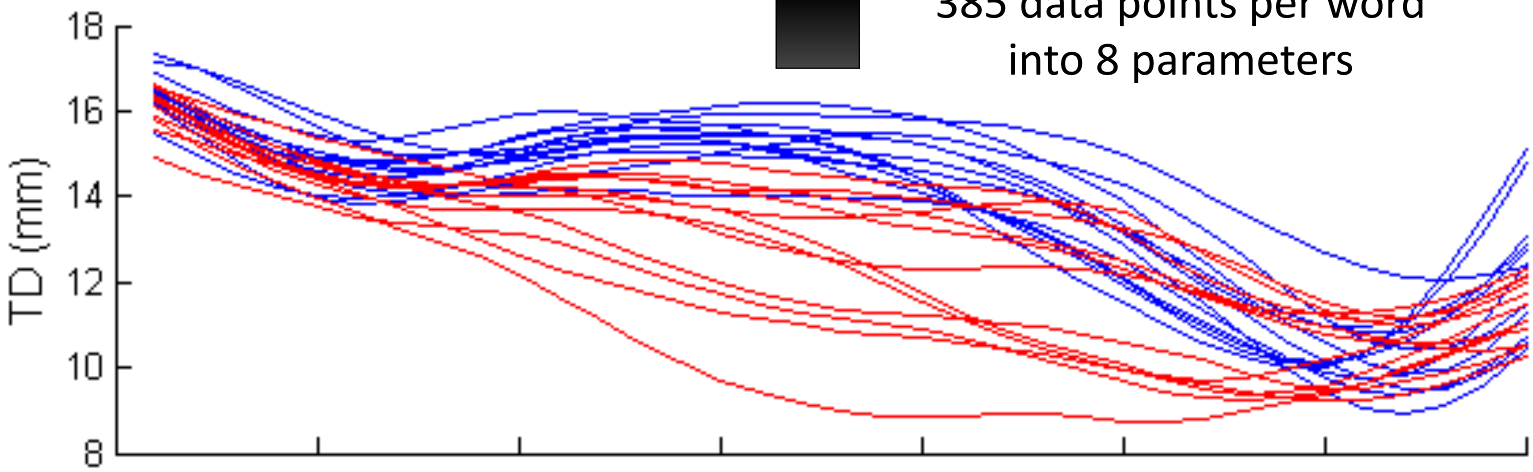


# Compact representations of tongue height trajectory over VCVCV

| S03 | Mean and standard deviation of DCT Coefficients | | | |
|---|---|---|---|---|
| | 1st Coeff | 2nd Coeff | 3rd Coeff | 4th Coeff |
| ʃudaika | **69.47(3.01)** | **6.31 (1.59)** | **-4.54(0.74)** | **0.94(0.48)** |
| ʃu̥taise: | **62.18 (6.34)** | **6.17 (1.83)** | **-0.04 ( 2.27)** | **0.63 (0.95)** |

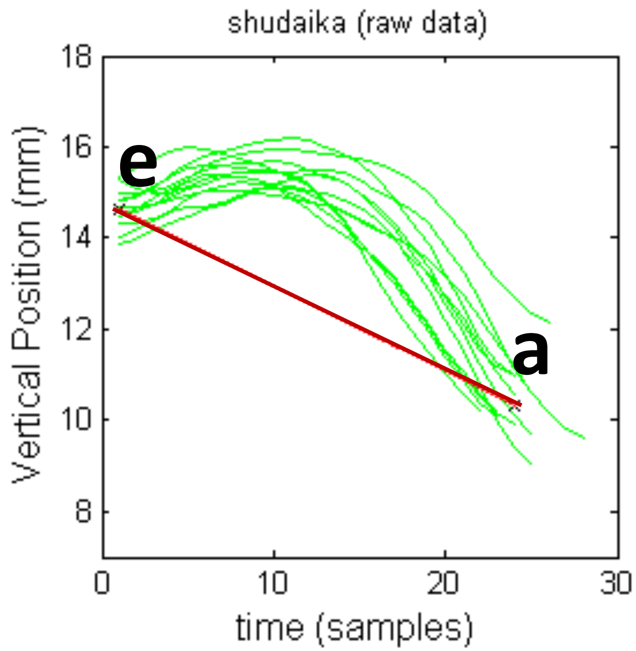
([$F(1,23)=23.30$, $p < .0001$***; Wilk's $\Lambda = 0.3209$])

385 data points per word into 8 parameters



# Defining the (noisy) targetless hypothesis in frequency space

| | 1st Coeff | 2nd Coeff | 3rd Coeff | 4th Coeff |
|---|---|---|---|---|
| ʃudaika | **69.47(3.01)** | **6.31 (1.59)** | **-4.54(0.74)** | **0.94(0.48)** |
| **targetless** | **60.49** (3.01) | **5.49**(1.59) | **0.00** (0.74) | **0.61** (0.48) |




Fit a line between vowel targets $V_1$(e) to $V_3$(a)

Transform the line into the same DCT space as data

Define targetless distribution using variance from the data

# Inverse Discrete Cosine Transform (iDCT)

**Simulate targetless trajectory from DCT coefficients:**

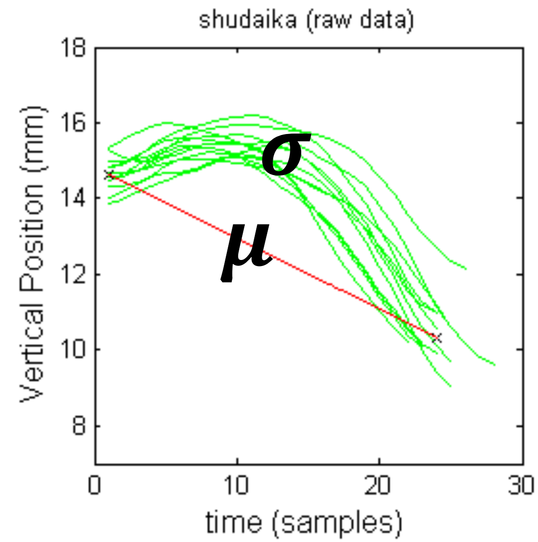
$$y(k) \sim \mathrm{N}(\mu(k), \sigma(k))$$

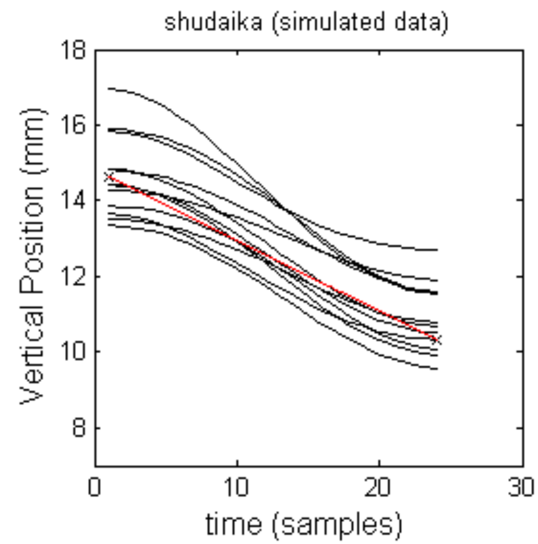
$$x(n) = \sum_{n=1}^{L} w(k) y(k) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)$$

$$n = 1,2, \dots L$$

Where *L* is the number of data samples and *x(n)* the trajectory to be simulated and:

$$w(k) = \begin{cases} \frac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\frac{2}{L}} & 2 \le k \le L \end{cases}$$



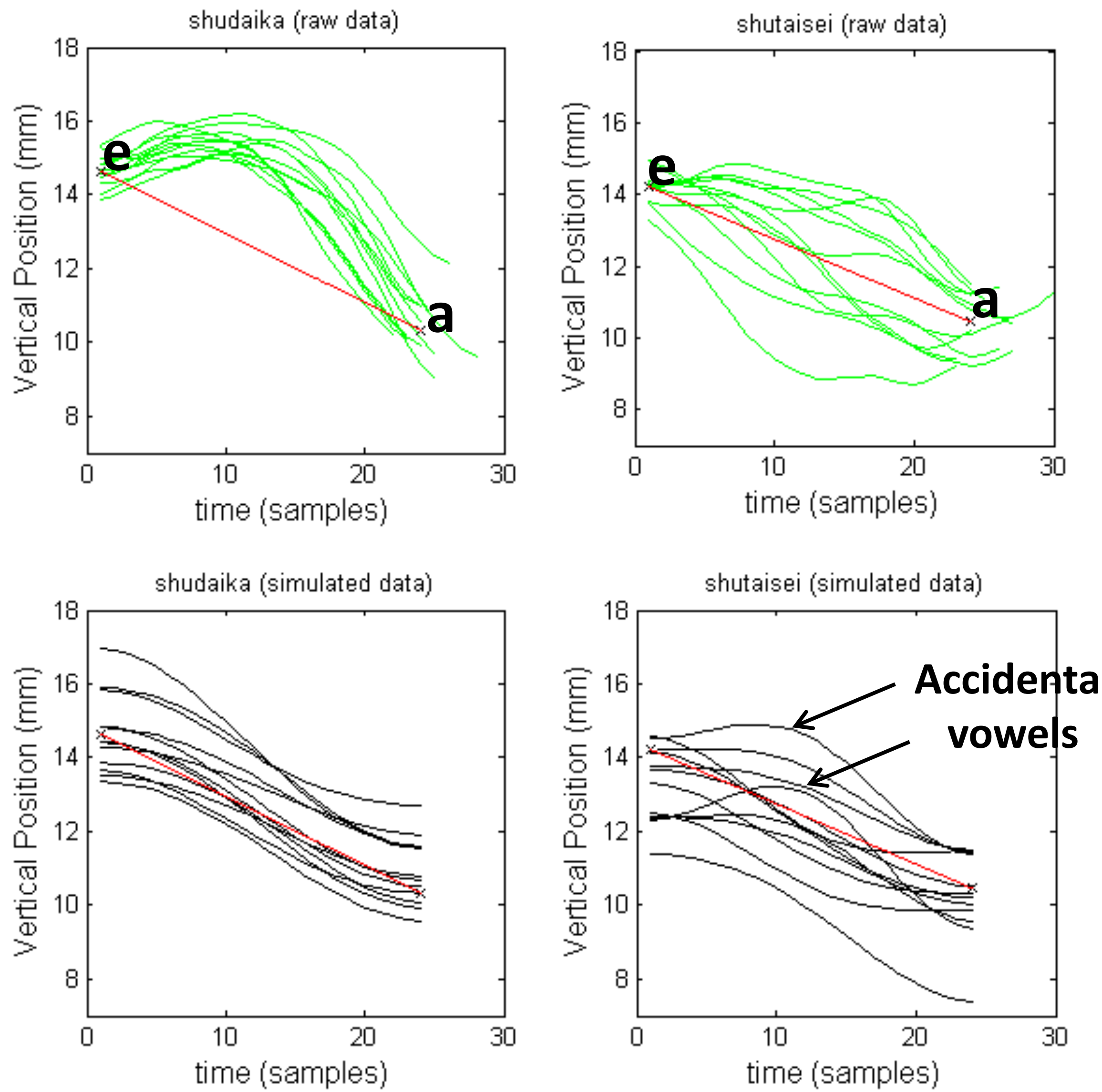
# Targetless simulations

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

Targetlessness evaluated againts the backdrop of realistic variability.

**Simulated "targetless" trajectories (black lines)**

When simulated with natural quantities of variability, the targetless trajectory can sometimes look like a vowel.



# Token-by-token evaluation

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

Training data = **voiced tokens** & **noisy null**

Test data = **voiceless tokens**

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

where...

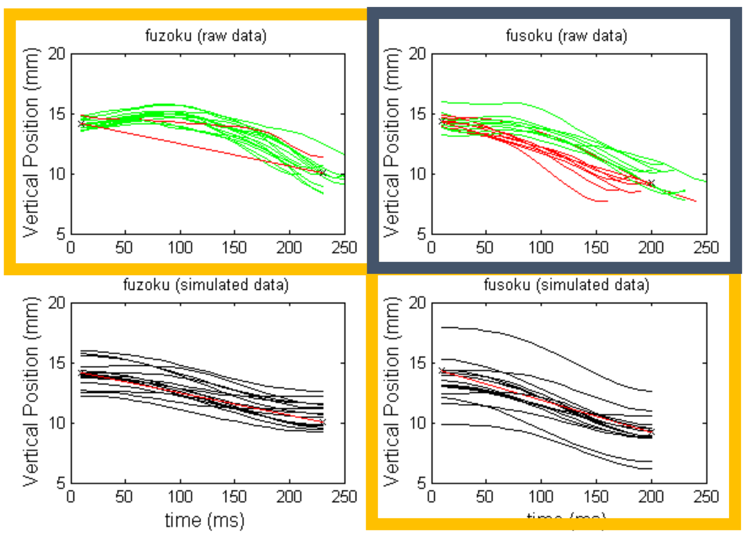
*T = {target present, targetless}*

$c_1$= *1st DCT Coefficient*

$c_2$= *2nd DCT Coefficient*

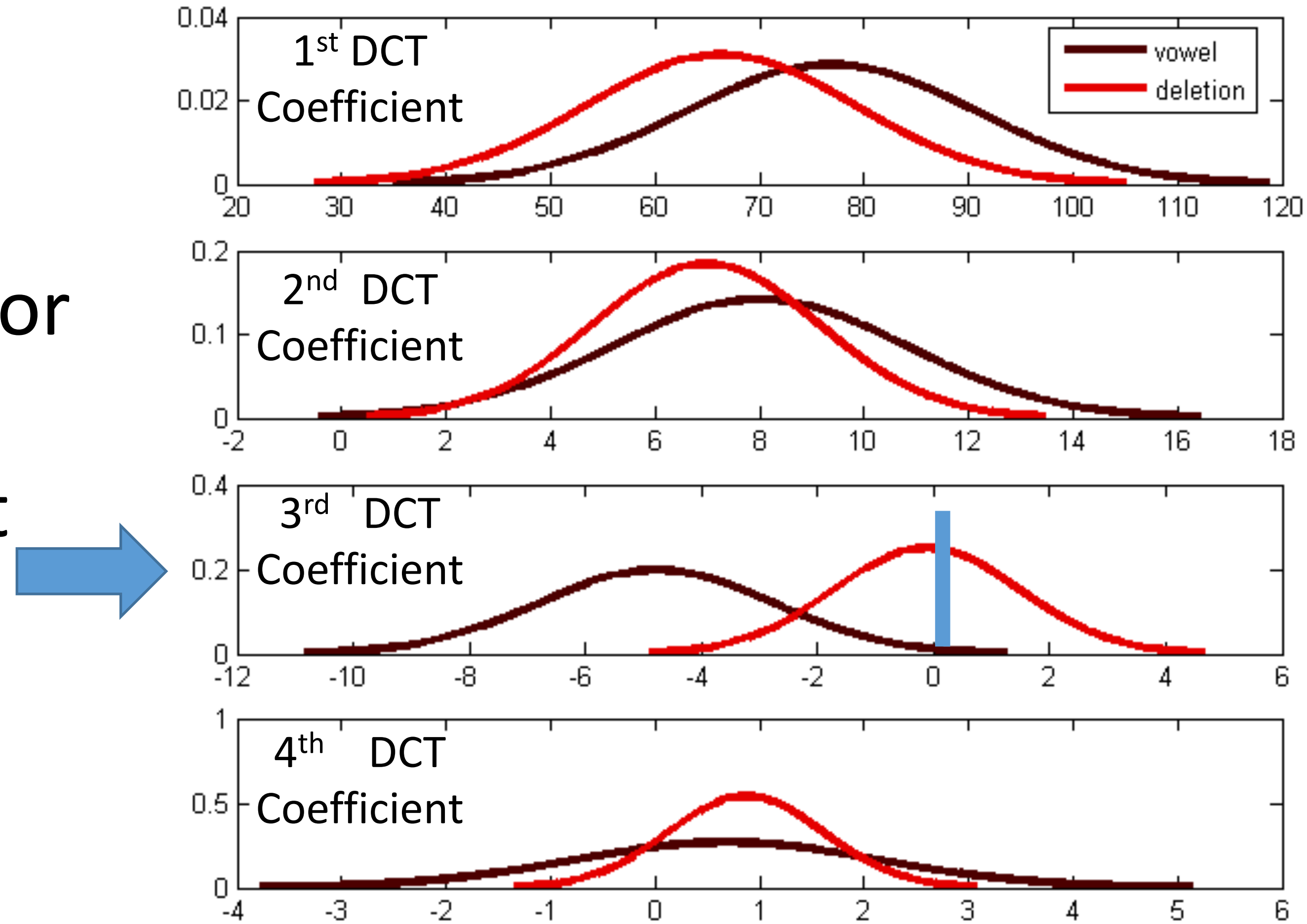
$c_3$= *3rd DCT Coefficient*

$c_4$= *4th DCT Coefficient*



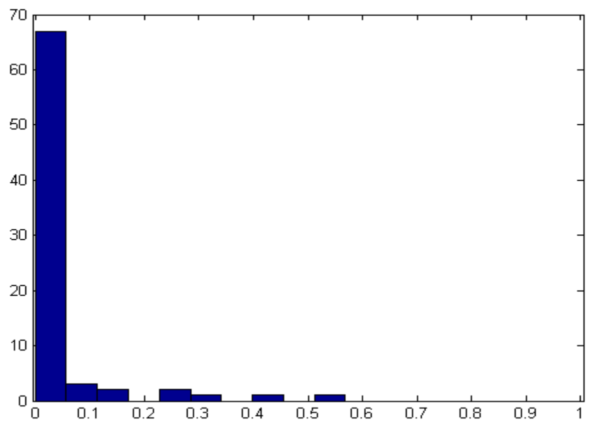
# Average classification parameters

Greatest separation for $3^{rd}$ DCT coefficient

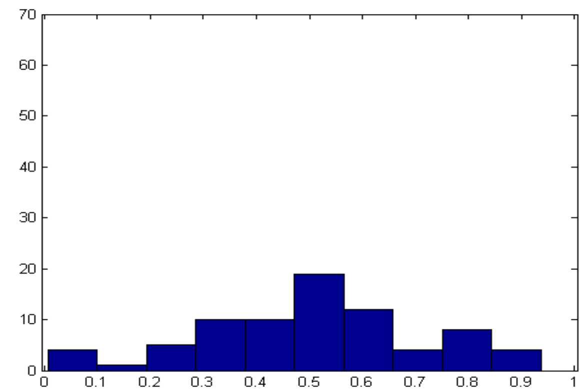


# Hypotheses expressed as (posterior) probability distributions

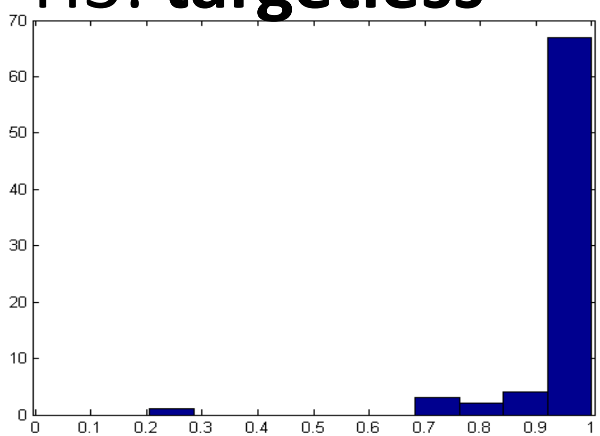
## H1: **full target**




## H2: **reduced target**




## H3: **targetless**




## H4: **variably targetless**