Minimal Pairs and Hyperarticulation of Singleton and Geminate Consonants as Enhancement of Lexical/Pragmatic Contrasts

Shin-ichiro Sano

Keio University
shinichirosano@gmail.com

NELS 48 @ University of Iceland
October 28, 2017

Synopsis

- Message Oriented Phonology (MOP, e.g., Aylett and Turk 2004; Hume and Bromberg 2005; Bell et al. 2009; Jaeger 2010; Cohen-Priva 2012; Shaw et al. 2014; Hall et al. 2016)
  - Informativity $\Rightarrow$ linguistic behavior
- This project
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates
- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftarrow$ informativity (Shannon's entropy)
- Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts

Synopsis

- Message Oriented Phonology (MOP, e.g., Aylett and Turk 2004; Hume and Bromberg 2005; Bell et al. 2009; Jaeger 2010; Cohen-Priva 2012; Shaw et al. 2014; Hall et al. 2016)
  - Informativity $\Rightarrow$ linguistic behavior
- This project
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates
- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftarrow$ informativity (Shannon's entropy)
- Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts

Synopsis

- Message Oriented Phonology (MOP, e.g., Aylett and Turk 2004; Hume and Bromberg 2005; Bell et al. 2009; Jaeger 2010; Cohen-Priva 2012; Shaw et al. 2014; Hall et al. 2016)
  - Informativity $\Rightarrow$ linguistic behavior
- This project
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates
- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftarrow$ informativity (Shannon's entropy)
- Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts

Roadmap

1. Background
   - Research topic
   - Message Oriented Phonology
   - Research objective
2. Method
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis
3. Results
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity
4. Discussion
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change

Roadmap

Research topic

Length contrast and singleton & geminate

(Arisaka 1940; Hashimoto 1950; Hattori 1960; Koizumi 1978; Vance 1987, 2008; Kawagoe 2015; Kawahara 2015)

- In Japanese, length or duration – important role lexically and pragmatically
- Modern Japanese has a variety of length contrasts.

Vowel length

short		*long*	
/obasan/	'aunt'	/obaasan/	'grandmother'
/biru/	'building'	/biiru/	'beer'

Research topic

Length contrast and singleton & geminate

(Arisaka 1940; Hashimoto 1950; Hattori 1960; Koizumi 1978; Vance 1987, 2008; Kawagoe 2015; Kawahara 2015)

- In Japanese, length or duration – important role lexically and pragmatically
- Modern Japanese has a variety of length contrasts.

Vowel length

short			*long*	
/obasan/	'aunt'	vs.	/obaasan/	'grandmother'
/biru/	'building'	vs.	/biiru/	'beer'

Research topic

- Modern Japanese has a variety of length contrasts.

  Consonant length

short		*long*
/kat̲a/ 'frame'	vs.	/katt̲a/ 'bought'
/hat̲o/ 'dove'	vs.	/hatt̲o/ 'hat'

- Short consonant (e.g. /p, t, k, b, d, g/) – singleton
  Long consonant (e.g. /pp, tt, kk, bb, dd, gg/) – geminate or *sokuon*

- Geminates – twice or three times as long as singletons (differs according to place and voicing)

- Pragmatic effect – emphatic lengthening
  /sugoi/ 'great' ⇒ *vowel* /sug̲oo̲i/ *consonant* /sug̲g̲oi/
  (cf. Podesva 2004, hyperarticulation of /t/ release in English)

Research topic

- Modern Japanese has a variety of length contrasts.

  Consonant length

	short		*long*
/kata/	'frame'	vs.	/katta/ 'bought'
/hato/	'dove'	vs.	/hatto/ 'hat'

- Short consonant (e.g. /p, t, k, b, d, g/) – singleton
  Long consonant (e.g. /pp, tt, kk, bb, dd, gg/) – geminate or *sokuon*

- Geminates – twice or three times as long as singletons
  (differs according to place and voicing)

- Pragmatic effect – emphatic lengthening
  /sugoi/ 'great' $\Rightarrow$ *vowel* /sugooi/ *consonant* /suggoi/
  (cf. Podesva 2004, hyperarticulation of /t/ release in English)

Research topic

Previous studies on singleton/geminate

(Han 1962, 1994; Homma 1981; Beckman 1982; Hirata & Whiton 2005; Kawahara 2006; Idemaru and Guion 2008; Ridouane 2010; Sano 2016, in press)

- Differences between singletons and geminates
- Identification of cues/factors affecting the choices
  - Phonetic studies
    duration, (duration of) preceding/following C/V, intensity, F0, F1
  - Phonological studies
    lexical strata (native, Sino-Japanese, mimetics, loanwords), geminates in inflection & compound formation, geminate devoicing & OCP (Kawahara & Sano 2013, 2017; Sano 2017)

- Difference in constriction duration – primary acoustic correlate of the singleton-geminate contrast. $\Rightarrow$ useful test of MOP's assumption

Research topic

Previous studies on singleton/geminate

(Han 1962, 1994; Homma 1981; Beckman 1982; Hirata & Whiton 2005; Kawahara 2006; Idemaru and Guion 2008; Ridouane 2010; Sano 2016, in press)

- Differences between singletons and geminates
- Identification of cues/factors affecting the choices
  - Phonetic studies
    duration, (duration of) preceding/following C/V, intensity, F0, F1
  - Phonological studies
    lexical strata (native, Sino-Japanese, mimetics, loanwords), geminates in inflection & compound formation, geminate devoicing & OCP (Kawahara & Sano 2013, 2017; Sano 2017)
- Difference in constriction duration – primary acoustic correlate of the singleton-geminate contrast. $\Rightarrow$ useful test of MOP's assumption

Roadmap

Message Oriented Phonology

Mathematical theories (e.g. Information Theory, Bayesian Inference)

- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

MOP

- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

- Posterior probability of a message given a phonological form (signal) in context is a *multiplicative function* of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)
- Predictability – high ⇒ Signal specificity – low, and vice versa

Message Oriented Phonology

Mathematical theories (e.g. Information Theory, Bayesian Inference)

- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

MOP

- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

- Posterior probability of a message given a phonological form (signal) in context is a *multiplicative function* of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)
- Predictability – high ⇒ Signal specificity – low, and vice versa

Message Oriented Phonology

Mathematical theories (e.g. Information Theory, Bayesian Inference)

- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

MOP

- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

- Posterior probability of a message given a phonological form (signal) in context is a *multiplicative function* of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)
- Predictability – high ⇒ Signal specificity – low, and vice versa

Message Oriented Phonology

Application of *informativity* in linguistics

(e.g. Aylett & Turk 2004; Hume & Bromberg 2005; Bell et al. 2009; Jäger 2010; Cohen-Priva 2012; Wedel et al. 2013a, b; Shaw et al. 2014; Kawahara 2016; Nelson & Wedel 2017)

- Contrastive hyperarticulation of VOT in English (Nelson & Wedel 2017)
- Functional load in diachronic change (Wedel et al. 2013a, b)
- Predictability of vowels in English (Aylett & Turk 2004)
- Predictability of content words in English (Bell et al. 2009)
- Informativity & trunation in Chinese compounds (Shaw et al. 2014)
- Quality of epenthetic V in English & French (Hume & Bromberg 2005)
- Informativity & syntactic patterns (Jaeger 2010)
- Informativity & geminate devoicing in Japanese (Kawahara 2016)

- Supporting evidence that informativity plays an important role in linguistic behavior.

Message Oriented Phonology

Application of *informativity* in linguistics

(e.g. Aylett & Turk 2004; Hume & Bromberg 2005; Bell et al. 2009; Jäger 2010; Cohen-Priva 2012; Wedel et al. 2013a, b; Shaw et al. 2014; Kawahara 2016; Nelson & Wedel 2017)

- Contrastive hyperarticulation of VOT in English (Nelson & Wedel 2017)
- Functional load in diachronic change (Wedel et al. 2013a, b)
- Predictability of vowels in English (Aylett & Turk 2004)
- Predictability of content words in English (Bell et al. 2009)
- Informativity & trunation in Chinese compounds (Shaw et al. 2014)
- Quality of epenthetic V in English & French (Hume & Bromberg 2005)
- Informativity & syntactic patterns (Jaeger 2010)
- Informativity & geminate devoicing in Japanese (Kawahara 2016)

- Supporting evidence that informativity plays an important role in linguistic behavior.

Roadmap

Research objective

This study

- applies equation (p.9), and examines Japanese gemination patterns.
  ⇑ Japanese gemination exhibits some information-related biased patterns.

- message – geminacy
- signal – phonological form (singleton/geminate)
- context – whether segment is minimally contrastive or not
- predictability – probability of singleton/geminate (prior expectation) (negative value of Shannon's entropy)
  (how likely singletons/geminates are to be realized by the current signal)
- signal specificity – duration or SG ratio
  (how clearly the speech signal conveys geminacy in message transmission)
  1. SG ratio: mean duration of geminates/mean duration of singletons
  2. Shannon's entropy $H(x)$: $H(x) = -\sum P(x)*log_2(P(x))$
     ⇒ operationalized in the analysis

Research objective

Hypothesis

- the durations of minimally contrasting singletons & geminates are hyperarticulated (singletons – shorter, geminates – longer) to provide more information to distinguish their host words from their minimal pair counterparts (competitors).

Roadmap

Corpus

CSJ-RDB (NINJAL 2012)

(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - *phonetic/phonological information*
    e.g. segment, mora, word, phrase, accent, intonation,
    time (onset/end) →duration
  - *morphological information*
    e.g. grammatical category, inflectional form

⇓

Detailed data retrieval (specify the target)

- Organization: APS (formal) / SPS (casual)
  →study the difference in register

Corpus

CSJ-RDB (NINJAL 2012)

(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - *phonetic/phonological information*
    e.g. segment, mora, word, phrase, accent, intonation,
    time (onset/end) $\longrightarrow$duration
  - *morphological information*
    e.g. grammatical category, inflectional form

$\Downarrow$

Detailed data retrieval (specify the target)

- Organization: APS (formal) / SPS (casual)
  $\longrightarrow$study the difference in register

Corpus

CSJ-RDB (NINJAL 2012)

(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - *phonetic/phonological information*
    e.g. segment, mora, word, phrase, accent, intonation,
    time (onset/end) $\longrightarrow$duration
  - *morphological information*
    e.g. grammatical category, inflectional form

$\Downarrow$

Detailed data retrieval (specify the target)

- Organization: APS (formal) / SPS (casual)
  $\longrightarrow$study the difference in register

Roadmap

Data retrieval

- I retrieved the target data from the CSJ-RDB.
- 12 speech samples in the CSJ-RDB
- Target segments: singletons & geminates
  (& word, phrase that contained the target)
- Using MySQL (implementing the programming language SQL)
  http://www.navicat.com
- Search formula in SQL
- Employing phonetic/phonological and morphological information

Data retrieval

Filtering

- Tokens were excluded from the dataset if targeted segments were a part of filled pauses or word fragments (a morpheme marked with (D)).
- Tokens were included if targeted segments possessed a non-standard pronunciation (a morpheme marked with (W)).

Annotation

- Duration was calculated based on the annotation in the CSJ-RDB (end time – onset time)
- Labels regarding minimal pair categories.

Data retrieval

- Manually annotated in an item-by-item manner, with reference to the Japanese dictionary.
- If a member of a pair is a proper noun, jargon, an archaic form, a dialectal form, or one that differs from its counterpart in accent and/or grammatical category, the pair is not regarded as a minimal pair.
- Three categories
  1. lexically contrastive minimal pairs
  2. pragmatically contrastive minimal pairs
  3. absence of minimal pairs

*pragmatically contrastive: gemination due to emphasis (e.g. /sugoi/ vs. /suggoi/ 'great') or allophonic pairs due to style/register (e.g. /mina/ vs. /minna/ 'everyone,' /fakusu/ vs. /fakkusu/ 'fax')

Data retrieval

- Manually annotated in an item-by-item manner, with reference to the Japanese dictionary.
- If a member of a pair is a proper noun, jargon, an archaic form, a dialectal form, or one that differs from its counterpart in accent and/or grammatical category, the pair is not regarded as a minimal pair.
- **Three categories**
  1. lexically contrastive minimal pairs
  2. pragmatically contrastive minimal pairs
  3. absence of minimal pairs

*pragmatically contrastive: gemination due to emphasis (e.g. /sugoi/ vs. /suggoi/ 'great') or allophonic pairs due to style/register (e.g. /mina/ vs. /minna/ 'everyone,' /fakusu/ vs. /fakkusu/ 'fax')

Roadmap

Dataset

- Exhaustive search & filtering $\Rightarrow$ 12,583 tokens

Table 1. The frequency distribution of singleton/geminate in the CSJ-RDB

	frequency	ratio
singleton	10,717	0.852
geminate	1,866	0.148

Table 2. The duration of singleton/geminate in the CSJ-RDB

	mean duration
singleton	34.4 msec
geminate	88.1 msec
SG ratio	2.56

(SG ratio: mean duration of geminates/mean duration of singletons)

Dataset

Table 3. The frequency distribution of singleton/geminate by the presence/absence of minimal pairs

	lexical	pragmatic	absence
singleton	162	148	10,407
geminate	161	31	1,674

Each token – subjected to the analysis

Roadmap

Statistical analysis

Distributional skews were tested by

- Linear mixed-effects (hierarchical generalized linear) model
  (Barr 2013; Bar et al. 2013)
- lmer in R (R development Core Team 1993-2017)
- Variables:
  - Dependent variable: duration of singleton/geminate
  - Fixed effect: presence/absence of minimal pairs (lexically or pragmatically contrastive)
  - Ramdom effects (grouping variables): speakers and items
    *Random slope and random intercepts were included in the model to have maximal random effects structure.
- Post-hoc test: multiple comparisons with Tukey's method

Statistical analysis

Distributional skews were tested by

- Linear mixed-effects (hierarchical generalized linear) model
  (Barr 2013; Bar et al. 2013)
- lmer in R (R development Core Team 1993-2017)
- Variables:
  - Dependent variable: duration of singleton/geminate
  - Fixed effect: presence/absence of minimal pairs (lexically or pragmatically contrastive)
  - Ramdom effects (grouping variables): speakers and items
    *Random slope and random intercepts were included in the model to have maximal random effects structure.
- Post-hoc test: multiple comparisons with Tukey's method

Roadmap

Items to be examined

Hypothesis

- Contrastive hyperarticulation: phonetic implementation is exaggerated to enhance the contrast between singletons & geminates

1. If singletons contrast in a minimal pair $\Rightarrow$ is their duration shorter than that do not?
2. If geminates contrast in a minimal pair $\Rightarrow$ is their duration longer than that do not?
3. How about pragmatic contrast (similar to lexical contrast or absence)?

Roadmap

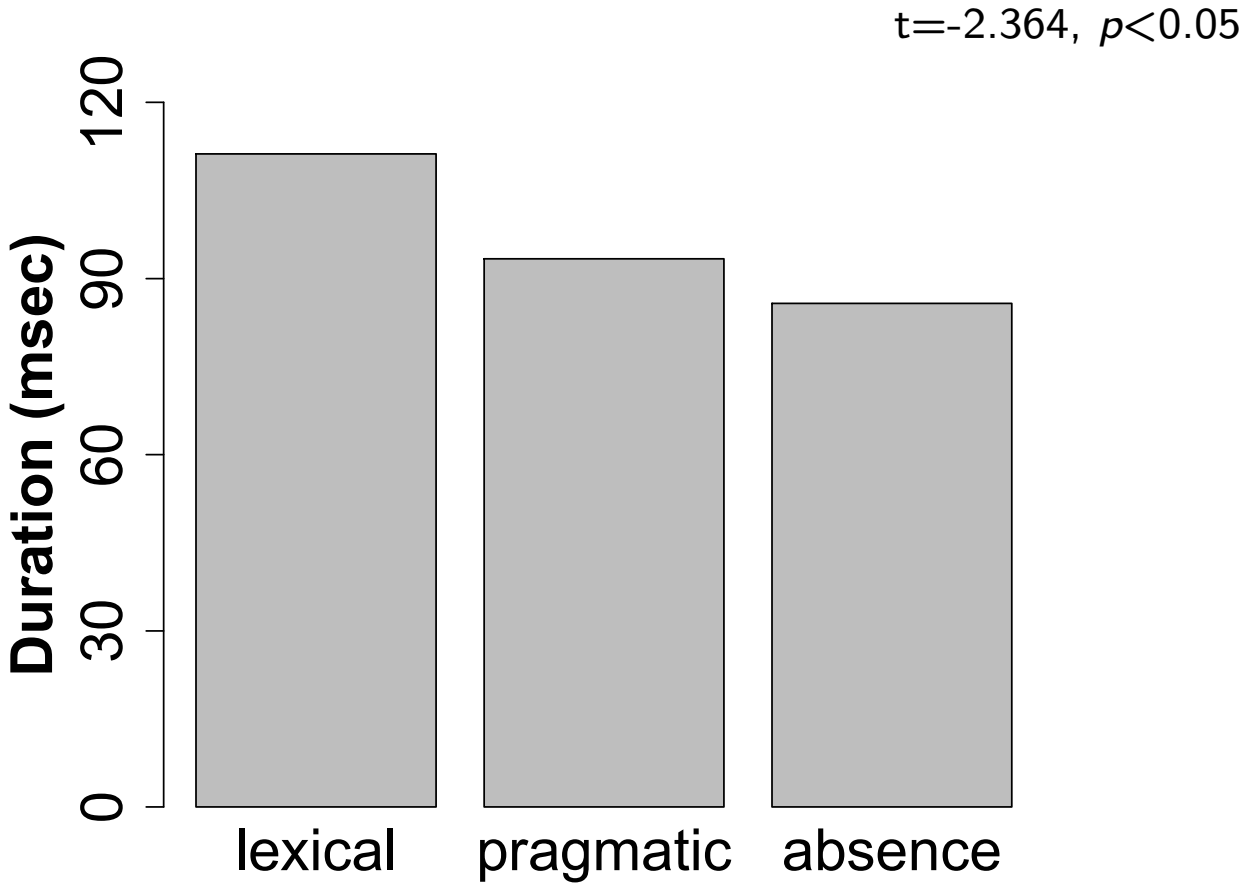
Duration of singletons

Figure 1. The duration of singletons in three categories

- Duration: lexical, pragmatic < absence (p<0.01)

Duration of geminates

Figure 2. The duration of geminates in three categories



- Duration: lexical, pragmatic < absence (p<0.05)

Summary of the results

Observed pattern

- singleton: lexical, pragmatic $<$ absence
- geminate: lexical, pragmatic $>$ absence

1. $\surd$ If singletons contrast in a minimal pair $\Rightarrow$ their duration is shorter than that do not.
2. $\surd$ If geminates contrast in a minimal pair $\Rightarrow$ their duration is longer than that do not.
3. Pragmatic contrast – similar to lexical contrast
(minimal pair => singletons – shorter, geminates longer)

Summary of the results

Observed pattern

- singleton: lexical, pragmatic $<$ absence
- geminate: lexical, pragmatic $>$ absence

$$\Downarrow$$

1. $\surd$ If singletons contrast in a minimal pair $\Rightarrow$ their duration is shorter than that do not.
2. $\surd$ If geminates contrast in a minimal pair $\Rightarrow$ their duration is longer than that do not.
3. Pragmatic contrast – similar to lexical contrast
(minimal pair => singletons – shorter, geminates longer)

Roadmap

Patterns in phonetic implementation and informativity

- How the results (phonetic implementation) relates to informativity?
- SG ratio: degree of contrast, distance between singleton & geminate
- **Shannon's entropy** *H(x)* represents informativity
- Entropy values were calculated for three categories.

e.g. Lexical contrast

The conditional probabilities of singleton and geminate are:

$P(singleton) = 162/\ (162+161) = 0.502$

$P(singleton) = 161/\ (162+161) = 0.498$

The information content of each segment is:

$Inf(singleton) = -log_2(0.502) = 0.996$

$Inf(geminate) = -log_2(0.498) = 1.004$

The entropy of a singleton-geminate contrast (token frequency: 162 and 161) in lexical contrast is:

$H(singleton\ geminate) = -\sum P(x)*log_2(P(x)) = 0.502*0.498+0.996*1.004 = 1.00$

Patterns in phonetic implementation and informativity

- How the results (phonetic implementation) relates to informativity?
- SG ratio: degree of contrast, distance between singleton & geminate
- **Shannon's entropy** *H(x)* represents informativity
- Entropy values were calculated for three categories.
  e.g. Lexical contrast

  The conditional probabilities of singleton and geminate are:
  $P(singleton) = 162/\ (162+161) = 0.502$
  $P(singleton) = 161/\ (162+161) = 0.498$

  The information content of each segment is:
  $Inf(singleton) = -log_2(0.502) = 0.996$
  $Inf(geminate) = -log_2(0.498) = 1.004$

  The entropy of a singleton-geminate contrast (token frequency: 162 and 161) in lexical contrast is:

  $H(singleton\ geminate) = -\sum P(x)*log_2(P(x)) = 0.502*0.498+0.996*1.004 = 1.00$

Patterns in phonetic implementation and informativity

Table 4. SG ratio and Shannon's entropy in three categories

	lexical	pragmatic	absence
SG ratio	4.45	3.62	2.48
Shannon's entropy	1	0.66	0.58

- SG ratio: lexical > pragmatic > absence
- Shannon's entropy: lexical > pragmatic > absence

- SG ratio follows informativity represented by Shannon's entropy

Patterns in phonetic implementation and informativity

Table 4. SG ratio and Shannon's entropy in three categories

	lexical	pragmatic	absence
SG ratio	4.45	3.62	2.48
Shannon's entropy	1	0.66	0.58

- SG ratio: lexical > pragmatic > absence
- Shannon's entropy: lexical > pragmatic > absence

- SG ratio follows informativity represented by Shannon's entropy

Roadmap

Contrastive hyperarticulation

- Hyporthesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).
- Presence of a minimal pair competitor
  $\Rightarrow$ predictability with which a target segment is identified lower
  $\Rightarrow$ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  $\Rightarrow$ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

predictability: low

signal specificity: high

posterior probability: constant

Contrastive hyperarticulation

- Hyporthesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).
- Presence of a minimal pair competitor
  $\Rightarrow$ predictability with which a target segment is identified lower
  $\Rightarrow$ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  $\Rightarrow$ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

predictability: low

signal specificity: constant

posterior probability: high

Contrastive hyperarticulation

- Hyporthesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).
- Presence of a minimal pair competitor
  $\Rightarrow$ predictability with which a target segment is identified lower
  $\Rightarrow$ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  $\Rightarrow$ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

$\Uparrow$

low

$\Uparrow$ constant $\Uparrow$ high

Contrastive hyperarticulation

- Hyporthesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).
- Presence of a minimal pair competitor
  $\Rightarrow$ predictability with which a target segment is identified lower
  $\Rightarrow$ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  $\Rightarrow$ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

$$P(message \mid signal, context) = P(message \mid context) * P(signal \mid message)$$

posterior probability — predictability — signal specificity

$\Uparrow$ low

$\Uparrow$ constant

$\Uparrow$ high

Contrastive hyperarticulation

- Degree of durational contrast between singletons & geminates (or hyperarticulation) reflected in the SG ratio follows from the informativity of singleton-geminate contrasts represented by Shannons entropy.
- Hyperarticulation of durational contrast is observed both in lexical minimal pairs and pragmatic minimal pairs.
  $\Rightarrow$ role of non-lexical/non-phonemic information in linguistic behavior.

Roadmap

Synchronic pattern – diachronic change

Functional load hypothesis

- Phonological contrasts that carry high functional loads (more minimal pairs) are less likely to neutralize (Martinet 1952; Hockett 1967; Surendran & Niyogi 2006; Wedel et al. 2013a,b).[1]
- Wedel et al. (2013a, b): demonstrate using a large database (9 languages/dialects) that the number of lexical minimal pairs distinguished by a phoneme opposition was a strong predictor of merger probability. (high functional load $\Rightarrow$ contrast – maintained)
- Prediction: frequent enhancement of a phonetic cue to a lexical category will become reflected in its long-term phonetic representation.

[1] Prior to 2013, the functional load hypothesis has been tested by many different ways, which produced mixed results (see Hockett 1967, and Surendran and Niyogi 2006).

Synchronic pattern – diachronic change

This project – supporting evidence

- The results predict that synchronically informative contrasts are hyperarticulated, and thus their phonetic implementation is perceptually salient.
  $\Rightarrow$ This results in diachronic stability, whereby informative contrasts tend to be preserved.
- Thus, the hyperarticulation of individual sounds induced by lexical (pragmatic) competition can influence long-term change in the system of phonemic contrasts (cf. Baese & Goldrick 2009; Peramunage et al. 2011).

Thank you!

This project was supported by JSPS KAKENHI Grant #16K16831, #26284059, and #16H03426.

References

1. Arisaka, Hideyo. 1940. *On'inron* [Phonology]. Tokyo: Sanseido.
2. Aylett, Matthew & 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.
3. Barr, Dale J. 2013. Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology* 4, p.328. doi: 10.3389/fpsyg.2013.00328
4. Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255-278.
5. Bayes, Thomas. 1763. An essay toward solving a problem in the Doctrine of Chances. *Philosophical Transactions* 53. 370-418.
6. Beckman, Mary. 1982. Segmental duration and the 'mora' in Japanese. *Phonetica* 39, 113-135.
7. 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, 91-111.

References

8. Cohen-Priva, Uriel. 2012. Deriving linguistic generalizations from information utility. *Doctoral dissertation, Stanford University*.
9. Hall, Kathleen Currie, Elizabeth Hume, Florian Jäger, & Andrew Wedel. 2016. *The message shapes phonology*. Ms.
10. Han, Mieko. 1962. The feature of duration in Japanese. *Onsei no Kenkyuu* [Study of sounds, The Phonetic Society of Japan] 10, 65-80.
11. Han, Mieko. 1994. Acoustic manifestations of mora timing in Japanese. *The Journal of the Acoustical Society of America* 96, 73-82.
12. Hashimoto, Shinkichi. 1950. *Kokugo on'in no kenkyuu* [Studies on Japanese phonology]. Tokyo: Iwanami.
13. Hattori, Shiro. 1960. *Gengogaku no hoo hoo* [Methodology in linguistics]. Tokyo: Iwanami.
14. Hirata, Yukari & Jacob Whiton. 2005. Effects of speaking rate on the singleton/geminate distinction in Japanese. *The Journal of the Acoustical Society of America* 118, 1647-1660.
15. Hockett, Charles F. 1967. The quantification of functional load. *Word* 23, 301-320.
16. Homma, Yayoi. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9, 273-281.

References

17. Hume, Elizabeth & Ilana Bromberg. 2005. Predicting epenthesis: An information-theoretic account. *Talk presented at 7th Annual Meeting of the French Network of Phonology.*
18. Idemaru, Kaori & Susan Guion. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association* 38(2), 167-186.
19. Jaeger, Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61, 23-62.
20. Kawagoe, Itsue. 2015. The phonology of *sokuon*, or geminate obstruents. In: Haruo Kubozono (Ed.), *Handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton, 79-119.
21. Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82(3), 536-574.
22. Kawahara, Shigeto. 2007 Sonorancy and geminacy. *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III:* 145-186

References

23. Kawahara, Shigeto. 2016. Japanese geminate devoicing once again: Insights from Information Theory. to appear in *Proceedings of FAJL 8 (MITWPL)*. Cambridge: MIT press.
24. Kawahara, Shigeto & Shin-ichiro Sano. 2013. A corpus-based study of geminate devoicing in Japanese: Internal factors. *Language Sciences* 40, 300-307.
25. Kawahara, Shigeto & Shin-ichiro Sano. 2017. /p/-driven geminate devoicing in Japanese: Corpus and experimental evidence. *Journal of Japanese Linguistics* 32, 57-78.
26. Koizumi, Tamotsu. 1978. *Nihongo no seishohoo* [Orthography of Japanese]. Tokyo: Taishukan.
27. Laplace, Pierre Simon. 1812. Theorie Analytique des Probabilites. Courcier, Paris. Reprinted as Oeuvres Completes de Laplace 7, 1878-1912. Paris: Gauthier-Villars.
28. Martinet, André. 1952. Function, structure, and sound change. *Word* 8, 1-32.
29. Nelson, Noah Richard & Andrew Wedel. 2017. The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics* 64. 51-70.

References

30. Payne, Elinor. 2005. Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association* 35(2), 153-181.
31. Ridouane, Rachid. 2010. Geminate at the junction of phonetics and phonology. In: Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Valleé (Eds.), *Laboratory Phonology* 10, 61-90. Berlin: Mouton de Gruyter.
32. Sano, Shin-ichiro. 2016. Durational contrast in gemination and informativity. *Paper presented at Predictability Symposium: The role of predictability in shaping human language sound patterns*, at Western Sydney University, Sydney Australia.
33. Sano, Shin-ichiro. 2017. A corpus-based study of phonological variation: Domain of OCP, and morphological boundary. *Proceedings of the 34th West Coast Conference on Formal Linguistics*, 439-446. Somerville, MA: Cascadilla Procedings Project.
34. Sano, Shin-ichiro. in press. A corpus-based study of singletons and geminates in Japanese: Segmental properties and contextual factors. *Japanese/Korean Linguistics* 24, 439-446. Stanford, CA: CSLI Publications.
35. Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 379-423.

References

36 Shaw, Jason, Chong Han & Yuan Ma. 2014. Surviving truncation: Informativity at the interface of morphology and phonology. *Morphology* 24, 407-432.

37 Surendran, Dinoj & Partha Niyogi. 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In: Ole Nedergaard Thomsen (Ed.), *Competing models of linguistic change: Evolution and beyond*, 43-58. Amsterdam and Philadelphia: John Benjamins.

38 Vance, Timothy. 1987. *An introduction to Japanese phonology*. Albany: State University of New York.

39 Vance, Timothy. 2008. *The sounds of Japanese*. Cambridge: Canbridge University Press.

40 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(3). 395-417.

41 Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013b. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179-186.